



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique Et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique

جامعة وهران 2 محمد بن أحمد
Université d'Oran 2 Mohamed Ben Ahmed

معهد الصيانة والأمن الصناعي
Institut de Maintenance et de Sécurité Industrielle « IMSI »

Département de Maintenance en Electromécanique

MÉMOIRE

Pour l'obtention du diplôme de MASTER

Filière : Génie industriel

Spécialité: Maintenance-Fiabilité-Qualité « MFQ » 'Master 2'

Thème

**Application des techniques d'intelligence artificielle
pour une maintenance prédictive**

Présenté et soutenu publiquement par :

BAYMOUT Thinhinan

GHOUL Ibtissem

Devant Le Jury composé de :

Nom et prénom	Grade	Etablissement	Qualité
ADJLOUA Abdelaziz	MCB	IMSI-Université D'Oran 2	Président
ROUAN-SERIK Mehdi	MCB	IMSI-Université D'Oran 2	Encadreur
CHENNOUFI Mohamed	MCA	IMSI-Université D'Oran 2	Examinateur

Année universitaire : 2021/2022

Table des matières

Table des matières.....	I- II- III.
Liste des figures.....	IV.
Liste des tableaux.....	V.
Remerciements.....	VI.
Résumé.....	VII.
Abstract	I.
ملخص.....	IX.
Introduction générale.....	X.

Chapitre 1 : Maintenance prédictive1.

1.1- Introduction.....	2.
1.2- Concepts de la maintenance.....	3.
1.2-1- Définition de la maintenance.....	3.
1.2-2- Définitions.....	3-4.
1.2-3- Rôle de la maintenance.....	4-5.
1.2-4- Les clés d'une bonne maintenance.....	5-6.
1.2-5- Politiques de la maintenance.....	6.
1.2-5-1- Maintenance corrective.....	6-7.
1.2-5-1-1- Maintenance curative.....	7.
1.2-5-1-2- Maintenance palliative.....	7.
1.2-5-2- Maintenance préventive.....	7-8.
1.2-5-2-1- Maintenance préventive systématique.....	8.
1.2-5-2-2- Maintenance préventive conditionnelle.....	9.
1.2-5-2-3- Maintenance préventive prévisionnelle	9-10.
1.2-5-3- Maintenance proactive.....	10.
1.2-5-4- Maintenance prédictive.....	11.
1.2-6- Objectifs de la maintenance : 'les 5 buts'	11.
1.2-7- La maintenance prédictive.....	12.
1.3-1- Définition.....	12-19.
1.3-2- Les trois phases pratiques de la maintenance prédictive.....	12-13.

1.3-2-1- La détection du défaut qui se développe.....	13.
1.3-2-2- L'établissement d'un diagnostic.....	14.
1.3-2-3- La notion de fiabilité d'un système.....	14.
1.3-3- Relation entre la maintenance et la fiabilité.....	14-16.
1.3-4- Divers outils de la maintenance prédictive.....	16.
1.3-5- Avantage de la maintenance prédictive par rapport à la préventive.....	17.
1.3-6- Avantage de la maintenance prédictive, au service de la productivité de l'entreprise.....	17-18.
1.3-7- Inconvénients.....	19.
1.3-8- Intérêt économique de la maintenance prédictive.....	19.
1.4- Conclusion	20.

Chapitre 2 : Techniques d'intelligence artificielle et l'intelligence artificielle dans la

Maintenance.....	21.
2.1- Introduction.....	22.
2.2- Les avantages, inconvénients et les limites de l'intelligence artificielle.....	23-24.
2.2-1- Les avantages.....	23.
2.2-2- Les inconvénients.....	23-24.
2.2-3- Les limites de l'IA.....	24.
2.2-3-1- Les limites matériels.....	24.
2.2-3-2- Les limites émotionnelles.....	24.
2.2-3-3- Les limites cognitives.....	24.
2.3- Outils et algorithmes mise en œuvre pour l'intelligence artificielle.....	25-60.
2.3-1- Techniques de l'apprentissage.....	25-37.
2.3-2- L'apprentissage automatique (Machine Learning).....	25-37.
2.3-2-1- Comment est-ce que cela fonctionne ?.....	26-27.
2.3-2-2- Types d'apprentissage.....	28-37.
2.3-2-2-1- L'apprentissage supervisé.....	29-34.
2.3-2-2-2- L'apprentissage non supervisé.....	34-35.
2.3-2-2-3- L'apprentissage par renforcement.....	35-37.
2.3-2-3- Le choix d'un type d'apprentissage automatique.....	37.
2.3-3- De l'apprentissage automatique à l'apprentissage en profondeur.....	38-45.
2.3-3-1- L'apprentissage en profondeur 'Deep Learning'.....	38-40.
2.3-3-1-1- Définition d'apprentissage en profondeur.....	39.
2.3-3-1-2- Les architectures d'apprentissage en profondeur.....	40.
2.3-3-2- Les réseaux de neurones artificiels.....	40-43.

2.3-3-2-1- Définition.....	40.
2.3-3-2-2- Architecture d'un réseau de neurones artificiels.....	41.
2.3-3-2-3- Principe de fonctionnement d'un réseau de neurones artificiels.....	41-42.
2.3-3-2-4- Les types de réseau de neurones.....	42-43.
2.3-3-3- Domaines d'applications du Deep Learning.....	43-44.
2.3-3-4- Exemple de Deep Learning.....	44-45.
2.3-3-5- Comparaison entre l'apprentissage automatique et l'apprentissage en profondeur.....	45.
2.3-4- Les domaines d'application de l'apprentissage automatique.....	46-48.
2.3-4-1- La santé.....	46.
2.3-4-2- Le secteur bancaire et financier.....	46.
2.3-4-3- Le transport.....	46.
2.3-4-4- Le commerce.....	47.
2.3-4-5- L'industrie.....	47.
2.3-4-6- L'agriculture intelligente.....	47.
2.3-4-7- Les médias.....	47-48.
2.3-4-8- Le droit.....	48.
2.3-5- Exemple dans le domaine de la santé utilisant le Machine Learning.....	48-58.
2.3-5-1- Les outils de développements et Platform utilisés.....	48-58.
2.3-5-1-1- Google 'Colaboratory'.....	48-49.
2.3-5-1-2- Bibliothèques utilisées.....	49-51.
2.3-5-1-3- Prétraitement de la collection de données.....	51-58.
2.4- Intelligence artificielles dans la maintenance.....	59-61.
2.4-1- L'évolution industrielle et l'industrie 4.0.....	59-60.
2.4-2- Analyse et validation du modèle.....	60.
2.4-3- Les freins à la mise en œuvre d'applications de maintenance à base d'IA.....	60-61.
2.5- Conclusion.....	61.
Chapitre 3 : Classification de la maintenance prédictive utilisant Machine Learning.....	62.
3.1- Introduction.....	63.
3.2- Étapes de construction du modèle.....	64.
3.3- Conclusion.....	71.
Conclusion générale.....	72.
Bibliographie.....	73.

Liste des figures

Figure 1.1 : le contenu de la fonction maintenance.....	5.
Figure 1.2 : Organigramme de politiques de maintenance.....	6.
Figure 1.3 : courbe représentant l'arrêt de production dû à une défaillance. Cette dernière, non prévue, se traduit par un arrêt brusque de la production.....	7.
Figure 1.4 : montre le principe de la maintenance préventive systématique. On remarque que le taux de défaillance a tendance à être plus élevé immédiatement après une réparation.....	9.
Figure 1.5 : Objectifs de la maintenance.....	11.
Figure 1.6 : Représentation graphique des amplitudes vibratoire ainsi que leur évolution dans le temps.....	13.
Figure 1.7 : Suivi de l'évolution du défaut par la surveillance prédictive.....	14.
Figure 1.8 : courbe en baignoire.....	15.
Figure 1.9 : l'impact de la maintenance sur la fiabilité des équipements.....	16.
Figure 2.1 : Les sous-branches de Data Science.....	25.
Figure 2.2 : Les approches et les algorithmes de l'apprentissage automatique.....	28.
Figure 2.3 : Apprentissage supervisé.....	29.
Figure 2.4 : Support Vector Machine (SVM) tirée de Navlani (2018).....	30.
Figure 2.5 : Principe des forêts aléatoires adaptées de Kirasich, Smith, & Sadler (2018).....	33.
Figure 2.6 : Neurone biologique et neurone artificiel.....	34.
Figure 2.7 : Apprentissage non supervisé.....	35.
Figure 2.8 : Le choix de l'algorithme d'apprentissage selon certains facteurs.....	37.
Figure 2.9 : Les sous-branches de l'intelligence artificielle.....	39.
Figure 2.10 : Architecture de base d'un réseau de neurones artificiel.....	41.
Figure 2.11 : Principe de fonctionnement d'un réseau de neurones artificiel.....	42.
Figure 2.12 : le processus d'auto-apprentissage pour l'identification d'un chat.....	45.
Figure 2.13 : La révolution industrielle tirée de Rioux (2016).....	59.

Liste des tableaux

Tableau 2.1 : Comparaison entre l'apprentissage automatique et l'apprentissage en profondeur...45.

Remerciements

Avant tout, nous remercions Dieu de nous avoir donné la force pour accomplir ce travail. Ensuite, nous remercions notre encadreur Monsieur Mehdi ROUAN-SERIK pour ses conseils et ses directives. Ainsi tous les enseignants et enseignantes du Département de Maintenance en Electromécanique

Aussi un grand remerciement à toutes les personnes, qui ont contribué de près ou de loin à l'accomplissement de ce travail.

Résumé

Le monde industriel dispose de machine et d'installation de plus en plus performante et complexe, les exigences de haute sécurité, la réduction des couts d'exploitation et la maitrise de la disponibilité des équipements donnent à la maintenance des systèmes, un rôle prépondérant, elle doit permettre de n'intervenir qu'en présence d'éléments défectueux, de minimiser le temps de réparation, et de fournir un diagnostic fiable et facilement interprétable malgré la complexité des équipements.

L'intelligence artificielle est largement utilisée dans la vie quotidienne et dans de nombreux domaines différents tels que l'industrie, l'automatisation et les systèmes experts, la médecine et la biologie, l'éducation... L'intelligence artificielle est ainsi essentielle au bon fonctionnement de la maintenance prédictive.

L'apprentissage automatique l'une des branches les plus importantes et les plus exploitées Grâce à son succès, il a réalisé un saut quantique dans le domaine de l'intelligence artificielle. Par conséquent, il est devenu le domaine le plus important dans le monde de l'informatique.

L'apprentissage automatique prend racine et fait ses preuves dans divers domaines. Les travaux présentés dans ce mémoire se penchent sur l'utilisation de cette branche de l'informatique appliquée à proposition d'un modèle pour classer l'erreur qu'une machine puisse commettre en une sorte d'erreur et prédit une défaillance de la machine.

Mots clés : maintenance prédictive, l'intelligence artificielle, apprentissage automatique...

Abstract

The industrial world has more and more efficient and complex machines and installations, high security requirements, the reduction of operating costs and the control of equipment availability give system maintenance a preponderant role, it must make it possible to intervene only in the presence of defective elements, to minimize the repair time, and to provide a reliable and easily interpretable diagnosis despite the complexity of the equipment.

Artificial intelligence is widely used in daily life and in many different fields such as industry, automation and expert systems, medicine and biology, education... Artificial intelligence is thus essential to the good operation of predictive maintenance.

Machine learning one of the most important and exploited branches Thanks to its success, it has achieved a quantum leap in the field of artificial intelligence. Therefore, it has become the most important field in the computer world.

Machine learning is taking root and proving its worth in various fields. The work presented in this thesis focuses on the use of this branch of computer science applied to the proposal of a model to classify the error that a machine can make into a kind of error and predicts a failure of the machine.

ملخص

يوجد في العالم الصناعي آلات ومنشآت أكثر كفاءة وأكثر تعقيداً، ومتطلبات أمان عالية، وتقليل تكاليف التشغيل والتحكم في توفر المعدات، مما يمنح صيانة النظام دوراً أساسياً، ويجب أن تجعل من الممكن التدخل فقط في وجود العناصر المعيبة، لتقليل وقت الإصلاح، ولتوفير تشخيص موثوق به ويمكن تفسيره بسهولة على الرغم من تعقيد المعدات.

يستخدم الذكاء الاصطناعي على نطاق واسع في الحياة اليومية وفي العديد من المجالات المختلفة مثل الصناعة والأتمتة والأنظمة الخبيرة والطب والبيولوجيا والتعليم ... وبالتالي فإن الذكاء الاصطناعي ضروري للتشغيل الجيد للصيانة التنبؤية. يعد التعلم الآلي من أهم الفروع وأكثرها استغلالاً بفضل نجاحه فقد حقق نقلة نوعية في مجال الذكاء الاصطناعي. لذلك، أصبح أهم مجال في عالم الكمبيوتر.

يتجذر التعلم الآلي ويثبت قيمته في مختلف المجالات. يركز العمل المقدم في هذه الأطروحة على استخدام هذا الفرع من علوم الكمبيوتر المطبق على اقتراح نموذج لتصنيف الخطأ الذي يمكن أن ترتكبه الآلة إلى نوع من الخطأ والتنبؤ بفشل الجهاز.

Introduction générale

L'intelligence artificielle apporte des outils modernes et bouleverse les processus existants dans de très nombreux domaines comme la santé, l'hôtellerie, l'agriculture et encore l'industrie. Dans notre travail nous focalisons sur l'industrie car ce secteur connaît depuis quelques années une véritable révolution numérique pour répondre aux défis économiques, sociaux ...

Ce mémoire est organisé en trois chapitres en plus de l'introduction et de la conclusion générale... Dans le premier chapitre on présente les concepts nécessaires à la compréhension de la maintenance prédictive. Dans un premier temps, les concepts de maintenance sont exposés. Par la suite, les politiques de maintenance sont présentés par rapport à l'intérêt économique et technique poursuivis par la maintenance des équipements industriels.

Le deuxième chapitre énonce et analyse les techniques d'intelligence artificielle et l'intelligence artificielle dans la maintenance car nous avons présenté de manière globale les principaux axes de l'apprentissage automatique.

Le troisième chapitre comporte une présentation d'un modèle proposé pour classer l'erreur qu'une machine puisse commettre en une sorte d'erreur et prédit une défaillance de la machine.

CHAPITRE 1

MAINTENANCE PREDICTIVE

1.1-Introduction

La maintenance s'est sans cesse améliorée à mesure que les procédés et équipements de production ont évolué, et c'est donc dans l'ordre naturel des choses qu'elle continue de le faire à l'avenir. Pour l'industrie du futur, l'objectif est de réduire les temps d'arrêt à leur plus simple expression, de tendre vers le « zéro ». Un but qui n'est pas irréalisable, car les progrès technologiques au service de la maintenance sont considérables, à condition notamment d'exploiter les données à bon escient et d'organiser la maintenance de manière optimale, car c'est la qualité de leur recueil, de leur analyse et de leur interprétation qui permettent de planifier les interventions et de faire en sorte que la production demeure la plus fluide possible, elles sont donc primordiales quant à la politique de maintenance d'une entreprise. Ces politiques de maintenance basées sur des interventions planifiées, ne réduisent cependant pas toujours les pannes des machines. Elles peuvent engendrer plus de main-d'œuvre, elles n'éliminent pas les pannes catastrophiques et causent des opérations d'entretien inutiles. Les méthodes expertes (maintenance conditionnelle) visent à surveiller la santé des machines sur la base de mesures d'indicateurs significatifs du comportement d'un mécanisme sans que son fonctionnement soit perturbé. La quasi-totalité des défaillances mécaniques peuvent être détectées grâce aux informations découlant d'indicateurs spécifiques.

Disposer d'une usine avec zéro panne est l'idéal de tout industriel. On peut l'atteindre, par la stratégie de maintenance privilégiée, dite prédictive. Allant plus loin que celles curative et même préventive, la maintenance prédictive repose sur une surveillance permanente des machines rendue possible par les nouvelles technologies : des capteurs d'analyse vibratoire permettant la collecte de données ,l'image fiable des défauts générés, des instruments de mesure acoustique utilisant les ultrasons , capables de détecter les défauts mécaniques, électriques ou d'étanchéité, des capteurs relevant de la thermographie (surveillance des niveaux de température) . Aujourd'hui, le concept de diagnostic de machines comprend la détection automatique et la classification des défauts, alors que le pronostic est un concept de maintenance prédictive qui repose principalement sur l'estimation du temps de fonctionnement restant avant qu'une défaillance se produise sur une machine. La prédiction des défaillances promet de réduire considérablement les coûts

1.2-Concepts de la maintenance

1.2-1 Définition de la maintenance

D'après la norme française NF EN 13306 X 60-319, la maintenance peut-être définie par: "L'ensemble de toutes les actions techniques, administratives et de management durant le cycle de vie d'un bien, destinées à le maintenir ou à le rétablir dans un état dans lequel il peut accomplir la fonction requise".

1.2-2- Définitions

➤ **Fonction requise (norme NF EN 13306)**

Fonction ou ensemble de fonctions d'un bien considérées comme nécessaires pour fournir un service donné.

➤ **Bien (norme NF EN 13306)**

Tout élément, composant, mécanisme, sous-système, unité fonctionnelle, équipement ou système qui peut être considéré individuellement.

➤ **Le management de la maintenance (norme NF EN 13306)**

Toutes les activités des instances de direction qui déterminent les objectifs, la stratégie elles responsabilités concernant la maintenance et qui les mettent en application par des moyens tels que la planification, la maîtrise et le contrôle de la maintenance, l'amélioration des méthodes dans l'entreprise, y compris dans les aspects économiques.

➤ **Panne (norme NF EN 13306)**

État d'un bien inapte à accomplir une fonction requise, excluant l'inaptitude due à la maintenance préventive ou à d'autres actions programmées ou à un manque de ressources extérieures. Elle se subdivise en :

✚ **Panne masquée (norme NF EN 13306)**

Situation dans laquelle une panne existe dans une partie du bien, mais ne peut être détectée en raison d'une panne du bien lui-même ou à cause d'une autre panne de cette même partie ou d'une panne d'une autre partie.

✚ **Panne latente (norme NF EN 13306)**

Panne existante, mais qui n'a pas encore été détectée.

✚ **Panne partielle (norme NF EN 13306)**

Panne caractérisée par le fait que le bien ne peut accomplir que quelques-unes mais pas toutes les fonctions requises.

➤ **La politique de maintenance (norme NF EN 13306)**

La stratégie de maintenance Méthode de management utilisée en vue d'atteindre les objectifs de maintenance.

➤ **Défaut**

Ecart entre une caractéristique d'une entité et la caractéristique voulue, cet écart dépassant des limites d'acceptabilité.

➤ **Défaillance**

Altération ou cessation de l'aptitude d'un bien à accomplir une fonction requise.

Altération → défaillance partielle.

Cessation → défaillance complète.

1.2-3- Rôle de la Maintenance

Dans les entreprises, la maintenance est considérée comme fonction stratégique et est définie comme étant un ensemble des actions techniques, administratives et de management destinées à maintenir un bien industriel en bon état de marche durant un cycle de vie ou à le rétablir dans un état dans lequel il peut accomplir la fonction requise. Présentent la fonction maintenance est comme un ensemble d'activités regroupées en deux sous-ensembles: les activités à dominante technique et les activités à dominante gestion.

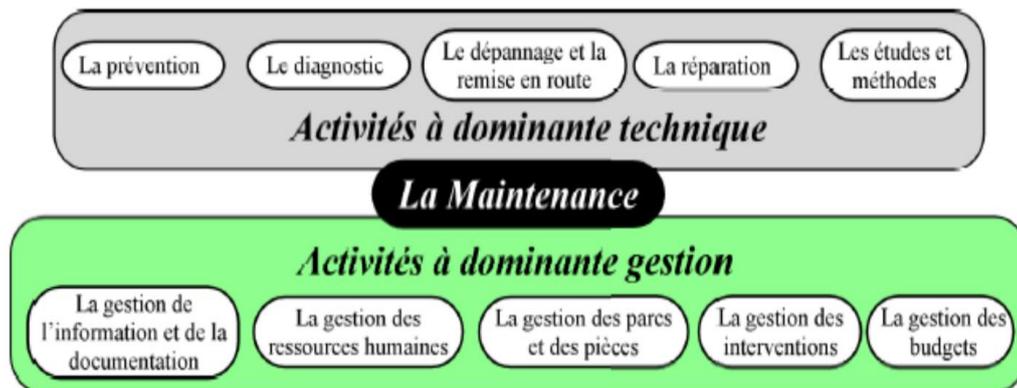


Figure1.1 : le contenu de la fonction maintenance

Dans la définition de la maintenance, nous trouvons deux mots-clés : maintenir et rétablir. Le premier fait référence à une action préventive. Le deuxième fait référence à l'aspect correctif. Nous présentons dans les paragraphes qui suivent les définitions de chaque type de maintenance.

1.2-4- Les clés d'une bonne maintenance

a) Sur les machines

- Répertorier et numéroter l'ensemble des machines et collecter les informations disponibles sur chacune d'elles.
- Etablir un dossier technique pour chaque machine.
- Déterminer le niveau de criticité de chaque équipement (influence d'une panne sur le reste de la production).

b) Avec les opérateurs

- Engager une action de formation pour expliquer ce qu'est la maintenance.
- Confier aux opérateurs des tâches simples : graissage, nettoyage, observations visuelles sur l'état des machines.
- Les inciter à signaler toute anomalie ou incident.

c) Avec le service de maintenance

- Formaliser l'ensemble des procédures : gestion des pièces de rechange, fiches de recueil des observations, etc.
- Etablir un programme prévisionnel avec un suivi des actions entreprises.
- Concentrer les équipes internes sur les actions les plus critiques et sous-traiter les tâches trop spécifiques ou à faible valeur ajoutée.

1.2-5- Politiques de maintenance

La figure présente l'organigramme de politiques de maintenance :

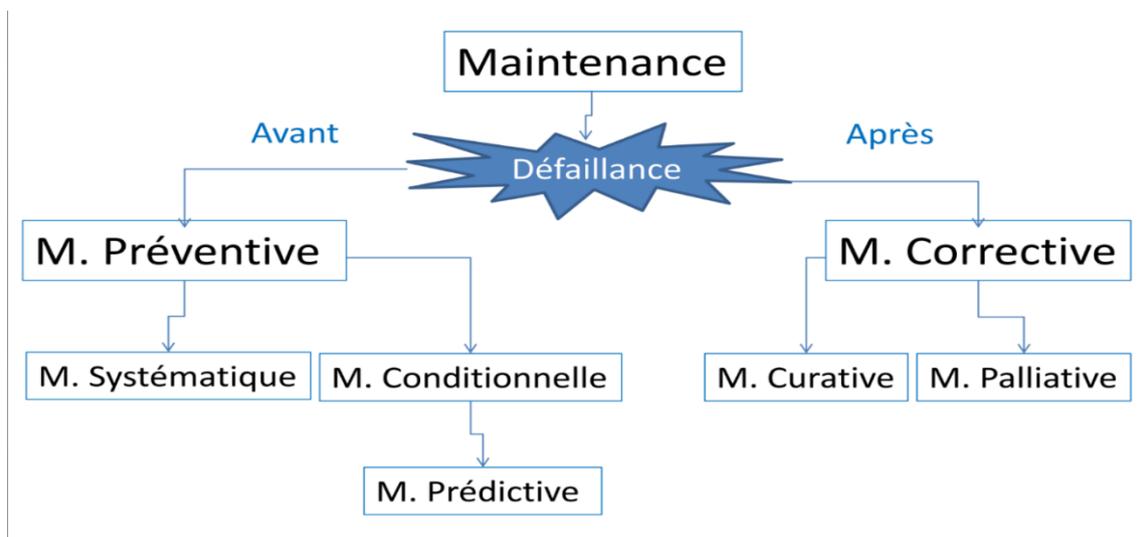


Figure 1.2 : Organigramme de politiques de maintenance

1.2-5-1- Maintenance corrective

La maintenance corrective est exécutée après une panne ou une défaillance. Elle est destinée à remettre un bien dans un état dans lequel il peut accomplir une fonction requise. Par ailleurs, elle permet de repérer et de corriger la source de la défaillance, lorsque l'avarie n'est ni progressive ni graduelle, pour éviter les incidents répétitifs. La maintenance corrective, comprend deux types d'intervention :

- Les interventions palliatives qui remettent le système en état de fonctionnement Provisoire, le dépannage.
- Les interventions curatives permettant de réparer le système d'une manière définitive

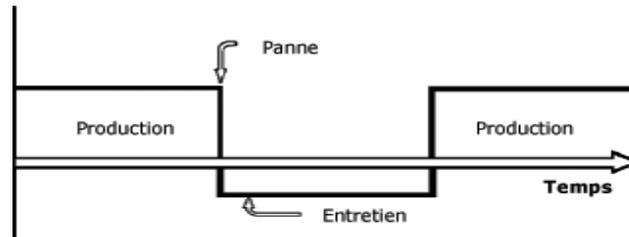


Figure 1.3 : courbe représentant l'arrêt de production dû à une défaillance. Cette dernière, non prévue, se traduit par un arrêt brusque de la production

1.2-5-1-1- Maintenance curative

Ce type de maintenance permet de remettre définitivement en état le système après l'apparition d'une défaillance. Cette remise en état du système est une réparation durable. Les équipements réparés doivent assurer les fonctions pour lesquelles ils ont été conçus. Une réparation est une opération définitive de la maintenance curative qui peut être décidée soit immédiatement à la suite d'une défaillance, soit après un dépannage, ce type de maintenance, provoque donc une indisponibilité du système.

1.2-5-1-2- Maintenance palliative

La maintenance palliative revêt un caractère temporaire, provisoire. Elle est principalement constituée d'opérations qui devront toutefois être suivies d'opérations curatives (réparations). Le dépannage est une opération de maintenance palliative qui est destinée à remettre le système en état provisoire de fonctionnement de manière à ce qu'il puisse assurer une partie des fonctions requises. Les opérations de dépannage sont souvent de courte durée et peuvent être nombreuses, et parce qu'elles ont lieu souvent ; elles sont également très coûteuses.

1.2-5-2- Maintenance préventive

La maintenance préventive est destinée à réduire la probabilité de panne du système, à suivre l'évolution de l'état d'un organe, de manière à prévoir une intervention dans un délai raisonnable et l'achat de la pièce de remplacement nécessaire (donc on n'a pas besoin de la

tenir en stock, si le délai normal le permet). . La maintenance préventive systématique ainsi que la maintenance corrective sont les formes traditionnelles de la maintenance. La première constitue encore aujourd’hui une activité fondamentale alors que la seconde est de plus en plus remplacée par une maintenance préventive conditionnelle qui, comme la maintenance prédictive, témoigne des nouvelles tendances en matière de maintenance. Toutes deux s’appuient sur les progrès informatiques et mettent-en œuvre des techniques de diagnostic sophistiquées.

La maintenance préventive comprend :

- Maintenance préventive systématique, (Contrôles, remplacements systématiques)
- Maintenance préventive conditionnelle, ou contrôles non destructifs. (Prédictive)
- Maintenance préventive prévisionnelle. (Proactive)

1.2-5-2-1- Maintenance préventive systématique

La maintenance préventive systématique est soumise à un programme et à un échéancier fixe dans le but de réduire la probabilité de défaillance ou de dégradation d’une machine de fabrication. Il s’agit d’une maintenance proactive plutôt que réactive, « pensée en amont de l’exploitation ». De plus, elle ne présente pas de réelles garanties quant à la fiabilité des machines. En revanche, la surveillance en état de marche s’avère parfois moins dommageable que des arrêts ponctuels qui permettent d’examiner le degré d’usure. La maintenance systématique a donc été rapidement écartée sauf dans les cas où la sécurité était en jeu, par exemple dans les industries chimiques et pétrolières. Les entreprises lui préfèrent de plus en plus la maintenance conditionnelle. Il faut toutefois rappeler que ces deux types de maintenance sont complémentaires, certains travaux comme le nettoyage, la lubrification, le resserrage, etc. devant être effectués de façon systématique.

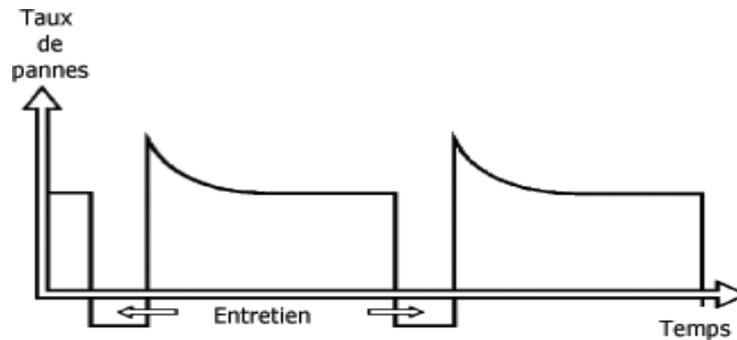


Figure 1.4 : montre le principe de la maintenance préventive systématique. On remarque que le taux de défaillance a tendance à être plus élevé immédiatement après une réparation.

1.2-5-2-2- Maintenance préventive conditionnelle

L'électronique et l'informatique ont été des facteurs déterminants dans le développement de la maintenance préventive conditionnelle, cette dernière s'appuie sur des techniques de diagnostic non destructives qui fournissent de l'information sur l'état de dégradation d'une machine, pour cela nous faisons souvent recours aux analyses prédictives ou prévisionnelles permettant d'estimer la tendance évolutive du dysfonctionnement éventuel. La maintenance préventive conditionnelle sert à optimiser la maintenance en « déclenchant les révisions au dernier moment et en limitant les travaux systématiques, donc les coûts. » Outre, si la maintenance préventive conditionnelle permet une réduction significative des coûts liés à une maintenance corrective ou palliative, elle entraîne par ailleurs un investissement élevé en instrumentation de surveillance, d'analyse et des coûts de formation pour le personnel de maintenance qui en est responsable. Ce type de maintenance est donc plus souvent réalisé par un service externe spécialisé ou par les fournisseurs qui possèdent les laboratoires et l'appareillage nécessaires à ces analyses et procédés de contrôle.

1.2-5-2-3- Maintenance préventive prévisionnelle

La maintenance prévisionnelle est définie, par la norme AFNOR - 2002, comme la maintenance conditionnelle exécutée en suivant les prévisions extrapolées de l'analyse et de

l'évaluation de paramètres significatifs de la dégradation de l'entité aux instants futurs. Elle peut prendre en compte un âge du matériel qui n'est pas forcément calendaire mais par exemple le temps de fonctionnement mesuré depuis la dernière inspection. Cependant, la maintenance systématique est coûteuse car l'usure des pièces dépend beaucoup des conditions de fonctionnement. Elle risque donc d'intervenir trop tôt ou trop tard .La maintenance prévisionnelle est également appelée maintenance proactive.

1.2-5-3- Maintenance proactive

La maintenance proactive est un terme pour désigner le renforcement des maintenances préventive et prédictive. Ainsi, dans cette forme de maintenance, la maintenance prédictive est utilisée pour repérer les causes des problèmes survenant sur les machines ou les processus. Cette modalité de maintenance fournit à l'expert un moyen pour créer une réduction efficace du temps total de pannes des équipements. En fait, dans ce type de maintenance, les équipements fonctionnent presque sans subir d'arrêt non planifié ; le temps moyen entre les défaillances MTBF (Mean Time Between Failures) des équipements est considérablement allongé.

L'objectif majeur d'une maintenance proactive est l'analyse de l'évolution « surveillée» des paramètres significatifs de la dégradation du bien, permettant de retarder et de planifier les interventions. La surveillance de ces paramètres est réalisée dans le cadre d'une automatisation industrielle communicante intégrant les nouvelles technologies et concepts. Le développement des réseaux de communication associés permet d'accéder à un grand nombre de données temps réel existantes et donc d'en établir des paramètres de suivi.

Les systèmes de maintenance proactive reposent parfois sur la technologie M2M (Machine To Machine) et permettent notamment de mieux gérer les consommables.

1.2-5-4- Maintenance prédictive

La maintenance prédictive est intimement liée à la maintenance préventive conditionnelle. Elle couvre les mêmes types de travaux. Cependant, contrairement à la maintenance conditionnelle qui a recours à un capteur ou à la mesure de l'usure, la maintenance prédictive s'appuie sur l'évolution d'un symptôme, d'une dégradation. Par exemple, on tentera d'évaluer, selon le niveau des vibrations, le moment où le seuil critique sera atteint. Les travaux seront alors programmés dans le temps, ni trop tôt ni trop tard, de façon à éviter d'atteindre ce seuil critique.

1.2-6- Objectifs de la maintenance: 'les 5 buts'

- Augmenter la durée de vie des machines
- Diminuer les temps d'intervention
- Augmenter la disponibilité des machines
- Diminuer les coûts d'exploitation
- Supprimer les causes d'accidents.

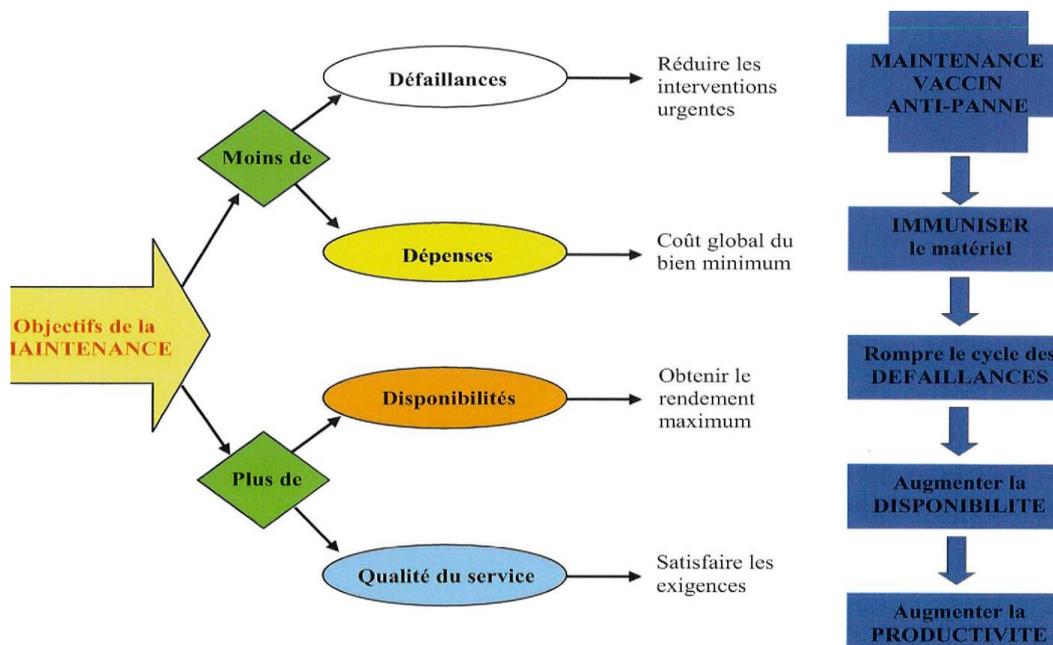


Figure 1.5 : Objectifs de la maintenance

1.2-7- La maintenance et l'entretien

Depuis longtemps on est confondue entre la maintenance et l'entretien bien que c'est deux notions différentes, L'entretien est l'ensemble des interventions pour réparation, dépannage après l'apparition des pannes de bien ou après défaillance de façon systématique pour prendre soin d'équipement, L'entretien est une action statique dans le sens qu'elle ne cherche pas les causes de défaillance, moins de propositions des solutions...etc. D'autre part la maintenance c'est en contraire dynamique, Elle cherche toujours de l'amélioration par trouver des solutions, Maitriser les équipements bien comme il faut, Planifier toutes les interventions selon les contraintes existées (le management de maintenance).

Maintenance

- Amélioration continue.
- Même niveau avec la production.
- le cout doit être maitrisé pour le minimiser.
- Coopération avec les autres fonctions ...etc.

Entretien

- L'absence de l'amélioration.
- Fonction dépendant de la production.
- Le cout est un élément non perturbant.
- Partition par rapport aux autres fonctions...etc.

1.3- Maintenance prédictive

1.3-1- Définition

La maintenance prédictive est définie selon la norme NF EN 13306 X 60-319 : les start-up elles entreprises spécialisées comme une «maintenance conditionnelle exécutée en suivant les prévisions extrapolées de l'analyse et de l'évaluation de paramètres significatifs de la dégradation du bien ». C'est une méthode proactive de maintenance, cela signifie que l'on peut anticiper et déceler la panne avant que celle-ci ne se produise, elle permet d'anticiper la panne d'une machine grâce à la surveillance continue et automatique de celle-ci. (En grande partie grâce aux données recueillies par des capteurs) et en analysant (en utilisant des outils de statistique et de probabilité mathématiques) de telles données pour déterminer la condition de santé et la nécessité probable d'une intervention de maintenance

pour un composant particulier ou système. La maintenance prédictive permet ainsi d'éliminer les procédures inutiles que l'on peut retrouver dans la maintenance préventive.

1.3-2- Les trois phases pratiques de la maintenance prédictive

La maintenance prédictive à le pouvoir de détecter rapidement les anomalies, sans même démonter la machine, le pouvoir de prévenir la panne et en prévoir la réparation le plus tard possible (figure 1.6), en fonction des impératifs de production, ce pouvoir fait de cette dernière une technique efficace permettant une augmentation perceptible de la productivité. En d'autres termes, l'intervention n'aura lieu que si nécessaire en établissant au préalable un diagnostic avant de programmer la réparation. La pratique de la maintenance prédictive comporte 3 phases :

- La détection du défaut qui se développe → analyse du signal
- L'établissement d'un diagnostic → diagnostique machine
- La notion de fiabilité d'un système → courbe baignoire

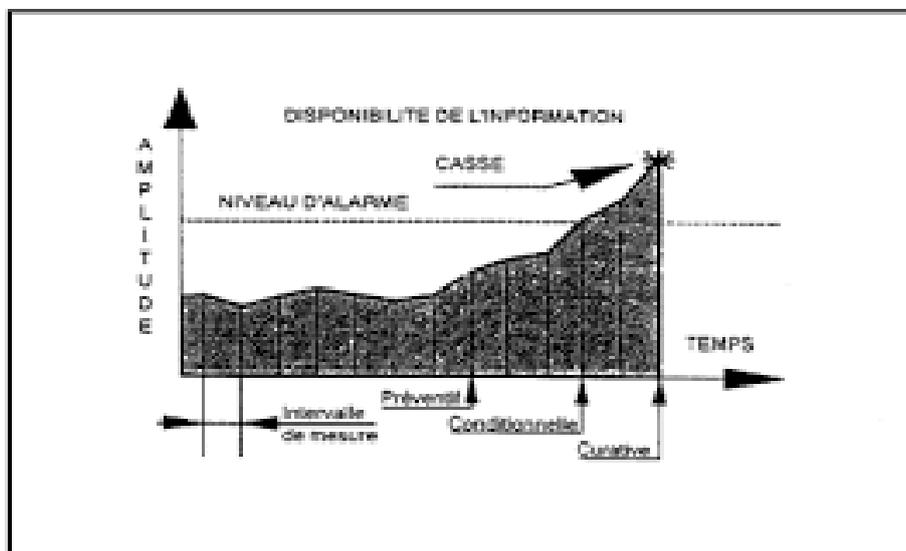


Figure 1.6 : Représentation graphique des amplitudes vibratoire ainsi que leur évolution dans le temps

1.3-2-1- La détection du défaut qui se développe

A la mise en route de chaque équipement, les principales caractéristiques de base des appareils sont enregistrées notamment la signature vibratoire (paramètres intéressants dans le cas de machines tournantes) et de divers paramètres de fonctionnement (température, usure, performances...). Ces caractéristiques ou signatures, serviront de référence pour suivre, par comparaison, l'évolution d'éventuels défauts ultérieurs.

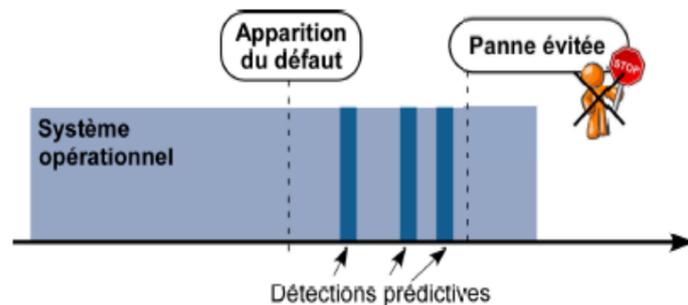


Figure 1.7 : *Suivi de l'évolution du défaut par la surveillance prédictive*

1.3-2-2- L'établissement d'un diagnostic

Dès qu'une anomalie est détectée par les outils caractéristiques, au sens d'analyse paramétrique, un diagnostic concernant l'origine et la gravité du défaut constaté sera établi.

1.3-2-3- La notion de fiabilité d'un système

La fiabilité d'un système s'exprime par la probabilité que ce dispositif accomplisse une fonction requise dans des conditions d'utilisation et pour une période de temps déterminée. La norme AFNOR définit la défaillance comme une altération ou une cessation d'un bien à accomplir une fonction requise. La durée de vie d'un système est une mesure de la quantité de service rendu, sa défaillance puisse survenir à n'importe quel moment, et il est souvent difficile de caractériser cette dernière. Ce qui suit qu'un système est considéré défaillant ou hors d'usage s'il n'est pas en mesure de réaliser la fonction pour laquelle il a été conçu. La défaillance d'un équipement peut être caractérisée par un taux appelé taux de

panne. Ce taux est aussi appelé taux de défaillance, taux de hasard ou taux de mortalité. Il représente également la vitesse d'arrivée de la panne.

Avec :

$$\lambda(t) = \frac{N(t) - N(t + \Delta t)}{N(t)\Delta t}$$

$\lambda(t)$: taux de panne

$N(t)$: Nombre de composants ayant survécu jusqu'à l'instant t .

$N(t + \Delta t)$: Nombre de composants ayant survécu jusqu'à l'instant $t + \Delta t$.

Si nous représentons le taux de panne en fonction du temps, nous obtenons une courbe appelée « en baignoire » qui est divisée en 3 parties (figure 1.8) : La première (1) est appelée période de mortalité infantile ou taux de panne ce taux de défaillance est élevé puis a tendance à décroître. Cela correspond à la période de rodage des systèmes mécaniques ou au « déverminage ». La deuxième (2) partie n'est autre que la durée de vie utile : c'est la zone où le taux de panne est constant. La dernière partie (3) est appelée le vieillissement ou l'usure : en atteignant cet âge, le composant commence à vieillir et le taux de défaillance augmente en fonction du temps, de façon significative cette évolution est connue sous le nom de « courbe en baignoire » comme le montre la Figure 1.8.

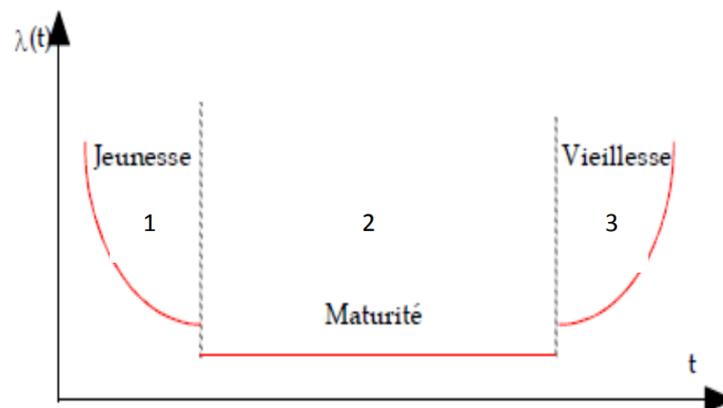


Figure 1.8 : courbe en baignoire

Il importe de rappeler que la fiabilité est une fonction décroissante de l'usage fait de l'équipement. Elle est reliée au taux de panne $\lambda(t)$ par la relation suivante :

$$R(t) = \exp\left(-\int_0^t \lambda(x) dx\right)$$

(t) : est la durée de la mission considérée.

1.3-3- Relation entre la maintenance et la fiabilité

La figure 1.9 présente la contribution des différents types de maintenance en ce qui concerne la fonction de fiabilité ($R(t)$) et la durée de vie utile de l'équipement.

On va sans dire qu'une réduction du taux de panne $\lambda(t)$ entraîne une amélioration de la fonction de fiabilité $R(t)$. C'est dans cette optique que la maintenance améliorative (prédictive) a été instaurée. La maintenance préventive, avec toutes ses variantes, va en revanche tenter de ramener le taux de panne à son niveau le plus bas en remplaçant la composante usée sans améliorer les caractéristiques intrinsèques de l'équipement.

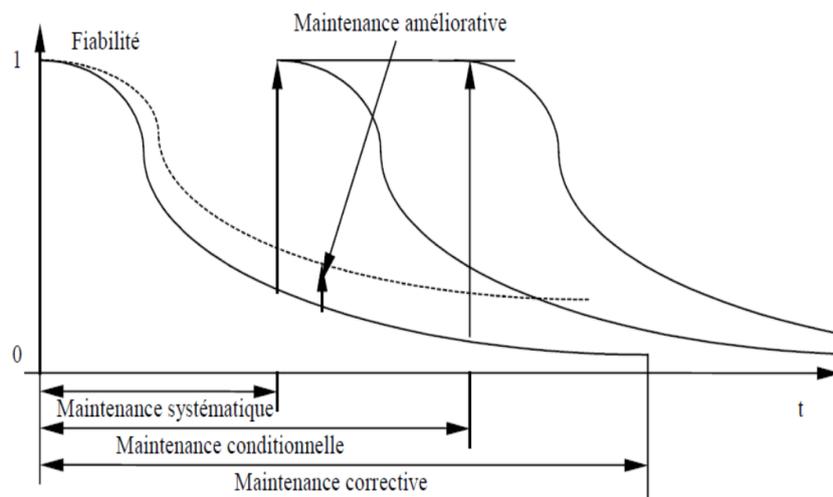


Figure 1.9 : *l'impact de la maintenance sur la fiabilité des équipements*

1.3-4- Divers outils de la maintenance prédictive

a) La GMAO (gestion de la maintenance assistée par ordinateur), les logiciels de supervision

b) L'IIOT (L'Internet Industriel des Objets) qui apporte des solutions en faisant remonter les informations des capteurs vers des plateformes de type Cloud. Ils permettent également de simplifier la mise en place d'un réseau au sein duquel les moyens de production sont interconnectés et peuvent même communiquer entre eux. .

c) Picomto (la solution de digitalisation d'instructions de travail) elle développe également un interrupteur connecté intelligent Check2 Start pour améliorer l'efficacité de votre parc machine.

d) Disposition d'un PC portable, tablette ou même Smartphone connecté à Internet, pour avoir des données précises sur la « santé » des équipements et sur leur rendement, mais aussi recevoir directement une alerte en cas de dysfonctionnement détecté. Le tout, en se trouvant loin du lieu de production et en ayant toute la latitude de régler un certain nombre de paramètres, comme les seuils d'alerte, par exemple. e) Les domaines de l'intelligence artificielle. Aujourd'hui, on n'est capable avec l'aide de capteurs de surveiller les différents composants d'une machine et grâce à une application informatique d'interpréter les données collectées et de prendre des décisions de manière automatique. Telle que, les capteurs pour mesurer :

- La température de l'équipement par imagerie infrarouge.
- La viscosité du lubrifiant.
- Les états de pression pour le flux de fluide et d'air.
- La fréquence de vibration pour détecter la fragilité structurelle, le désalignement, les problèmes de roulement etc.

1.3-5- Avantages de la maintenance prédictive par rapport à la préventive

La maintenance préventive est une méthode de maintenance qui permet d'entretenir des équipements en effectuant des vérifications périodiques sur ceux-ci. Les stratégies de maintenance préventive comprennent :

- Les contrôles de routine.

- Les mesures et ajustements.
- Les changements d'huile et lubrification.
- Les changements de pièces.

Elle intervient pour l'identification et le changement systématique de composants endommagés. Elle doit déterminer le moment idéal pour effectuer cette opération. On ne peut pas savoir à l'avance quelle pièce va s'user. On est donc obligé d'effectuer des inspections périodiques, ces dernières vont nécessiter de la main d'œuvre, l'arrêt de la machine et présentation d'un certain coût. Il peut aussi arriver qu'on change une pièce par précaution alors que celle-ci fonctionne parfaitement bien. Des études ont montré que dans 30% des cas, la maintenance préventive se révèle inutile, et peut même provoquer des problèmes supplémentaires.

a) La maintenance prédictive permet

- de diminuer le nombre d'interruptions des machines pour des opérations de maintenance.
- de diminuer le nombre de pannes. .
- de mieux planifier les interventions. .
- de mieux préparer les équipes d'intervention. .
- de mieux échanger entre les professionnels de maintenance et les équipes de production.
- de mieux anticiper et gérer les besoins de pièces détachées des outils.

b) La maintenance prédictive

C'est passer d'une logique de flux poussé à une logique de flux tiré. Le fournisseur n'intervient que lorsque des signaux émis par une machine reflètent une panne probable à court terme. C'est donc l'état réel de l'actif, et non un calendrier théorique, qui déclenche une intervention. La maintenance prédictive permet d'anticiper les pannes machines tout en minimisant les coûts des opérations de maintenance, le stock de pièces, le temps d'interruption des services, les heures supplémentaires et augmenter la durée de vie utile des machines, la productivité et les profits.

1.3-6- Avantages de la maintenance prédictive, au service de la productivité de l'entreprise

La maintenance prédictive permet d'améliorer plusieurs indicateurs de performance :

- Amélioration du taux de rendement global (TRG) d'un équipement.
- Réduction des coûts de réparation et d'intervention.
- Amélioration et prédictibilité de la qualité du produit fini.
- La maintenance prédictive permet d'améliorer le retour sur investissement. Elle rationalise le processus et permet des gains de coût et de temps.
- Le placement des capteurs revient "10 fois moins chers" qu'une maintenance préventive classique.

1.3-7- Inconvénients

- Complexe à mettre en place et coûteuse

1.3-8- Intérêt économique de la maintenance prédictive

Une installation en fonctionnement même dans des conditions normales entraîne un certain vieillissement du matériel, Pour une exploitation correcte, il est nécessaire de maintenir cette installation en bon état, de trouver le niveau optimum qui maintiendra cette dernière à niveau, sans mettre en danger la sécurité des personnes. Afin que le coût et les incidents diminuent Dans le cas contraire, avec trop de maintenance préventive, le coût augmente. Le but de cette section est de présenter les principaux facteurs économiques qui justifient l'installation d'un système de détection de défauts sur les machines.

- Premièrement, les gains escomptés de l'utilisation d'un système de surveillance sont exposés. Certains de ceux-ci sont difficilement estimables en valeur monétaire mais représentent un avantage certain pour l'entreprise qui les détient.
- Deuxièmement, les différents coûts associés aux arrêts de production non planifiés sont présentés.

1.4- Conclusion

Dans ce domaine, les améliorations sont continuellement apportées pour faire en sorte que l'industrie du futur puisse véritablement réduire le temps d'arrêt de ses moyens de production. La maintenance bénéficie de l'intégration de technologies de pointe, comme les capteurs intelligents, la réalité virtuelle, la télédétection laser pour faire des relevés en 3D d'équipements difficiles d'accès, des capteurs à UV, d'autres à infrarouge ou encore des drones. Ces innovations sont associées aux progrès réalisés en termes de capacité de stockage, de transfert et de traitement d'informations, pour constituer des systèmes qui se veulent plus fiables et plus efficaces. Si ces derniers apportent un avantage certain en ce qui concerne la maintenance corrective, ils sont, pour la plupart, davantage orientés vers l'amélioration de la maintenance prédictive. Ces systèmes sont étudiés pour aider les entreprises à ne plus subir les défaillances engendrant des arrêts, mais à les anticiper grâce à une interprétation efficiente des informations enregistrées par les capteurs. Ces derniers permettent d'alerter les techniciens lorsque les vibrations, les jeux mécaniques, la qualité des fluides (huiles) ou encore les niveaux de températures se rapprochent des seuils critiques, au-delà desquels la machine risque d'être défaillante.

Aujourd'hui, l'entretien prédictif est plus facile à instaurer, plus facile à mettre en place avec les techniciens d'entretien et enfin, plus rentable pour nos machines critiques. Alimenté par des données fiables, alors que les machines fonctionnent au maximum de leur capacité, l'entretien prédictif détecte les pannes potentielles et nous précise ce qui risque de briser suffisamment tôt pour que des travaux soient bien préparés et exécutés efficacement avec les pièces commandées juste à temps. On peut dire que la maintenance prédictive, marquera sans doute le futur de la maintenance industrielle. A l'heure où les usines deviennent de plus en plus intelligentes, la maintenance prédictive jouera un rôle capital.

CHAPITRE 2

TECHNIQUES D'INTELLIGENCE ARTIFICIELLE ET
L'INTELLIGENCE
ARTIFICIELLE DANS LA MAINTENANCE

2.1-Introduction

Avec l'arrivée d'internet, le monde est devenu hyper connecté dans lequel chaque objet manipulé (les voitures, les réfrigérateurs, les vêtements, les réseaux sociaux...) génèrent des millions de données supplémentaires chaque jour, qui s'ajoutent à un océan de données sans fin. Toutes ces données peuvent être utilisées pour garantir des services personnalisés et immédiats attendus. Mais comment peut-on transformer un tel océan de données sans fond en flux régulier d'informations pertinentes pour répondre à ces attentes ? La réponse est l'intelligence artificielle (IA).

L'intelligence artificielle comme étant un ensemble d'algorithmes qui traite un ensemble d'informations ou données, relatives à des tâches, d'une manière similaire ou identique à celle qu'un être humain utilisera pour prendre une décision ou résoudre un problème.

Le point commun entre toutes les définitions d'intelligence artificielle se résume dans l'objectif majeur de l'IA qui présente son ambition d'imiter le processus cognitifs de l'être humain. Ces processus comprennent l'apprentissage (acquisition d'informations et de règles liées à leur utilisation), le raisonnement (application des règles pour parvenir à des conclusions approximatives ou précises)

L'Intelligence artificielle moderne va plus loin car elle regroupe l'ensemble des technologies qui permettent aux machines, du robot à la solution digitale, d'adopter un comportement proche des réactions et capacités humaines. Et pour cela, elle s'appuie sur deux dimensions : le Machine Learning et le Deep Learning.

Ce deuxième chapitre introduit le vocabulaire de l'IA et approfondir en citant leur techniques d'apprentissage, bien précisément l'apprentissage machine (Machine Learning), et parler aussi de l'apprentissage profond (Deep Learning), outils et algorithmes, l'Intelligence Artificielle dans la maintenance.

2.2- Les avantages, inconvénients et les limites de l'intelligence artificielle

L'intelligence artificielle présente de nombreux avantages, concernant essentiellement le monde du travail, ses hautes performances et ses perspectives économiques intéressantes. Mais cela n'empêche pas l'existence de certains inconvénients et limites.

2.2-1- Les avantages

L'intelligence artificielle présente plusieurs avantages qui touchent à plusieurs domaines desquels on peut citer:

- L'IA pourrait remplacer l'homme dans son travail quotidien et pourrait réaliser des tâches pénibles ou dangereuses (comme ménage, courses, cuisine, jardinage...) sans présenter aucune contrainte physique (besoin de repos, nourriture ...etc.)
- Établir des calculs plus rapides et efficaces sur un ordinateur, grâce aux algorithmes de l'IA, et limiter les erreurs par rapport aux calculs réalisés par l'être humain.
- L'IA facilite les déplacements, il existe des véhicules pouvant se déplacer seuls à l'aide de caméras et de capteurs répartis sur ceux-ci ce qu'on appelle des véhicules autonomes.
- L'IA présente des avantages dans la médecine, tels que le suivi des patients à distance, les prothèses intelligentes, ainsi que les traitements personnalisés.
- Les machines, ont pas besoins de pauses et de rafraîchissements fréquents par rapport aux humains. Elles sont programmées pour de longues durées et peuvent fonctionner sans arrêt sans être fatiguées, tout ce dont elles ont besoin est une maintenance de temps en temps.
- L'IA présente aussi des avantages dans les jeux, elle permet aux joueurs de combattre des adversaires plus forts et plus expérimentés et donc de ne plus se lasser d'un jeu trop rapidement.

2.2-2- Les inconvénients

Comme toute discipline scientifique, l'intelligence artificielle présente un ensemble d'inconvénients :

- L'inconvénient le plus envisageable est la présence d'une erreur dans la programmation d'un robot, ce qui pourrait nuire à son bon fonctionnement. Les machines (ordinateurs, robots, véhicules intelligents...), ne savent pas déceler les erreurs de programmation. Les conséquences d'une telle erreur pourraient être catastrophiques à grande échelle, mais le risque que cela se produise reste très faible.

- L'augmentation de taux de chômage, car les entreprises préfèrent remplacer l'homme par des robots mécanisés dotés de l'IA qui ne se fatiguent pas et qui n'ont besoin que d'une maintenance de temps en temps, ce qui entraîne la suppression de postes.
- Le prix des recherches pour l'IA est très élevé, si l'on veut fabriquer des robots capables d'être autonomes dans la vie quotidienne, cela coûterait extrêmement cher.

En résumé, l'IA ne présente pas beaucoup d'inconvénients face aux nombreux avantages mais il faut savoir rester toujours vigilant.

2.2-3- Les limites de l'IA

Le domaine de l'intelligence artificielle n'est pas infini, il est soumis à des limites. On compte trois limites principales :

2.2-3-1- Les limites matérielles

Elle fonctionne à l'aide d'un réseau de transistors qui communiquent entre eux (presque aussi vite que la lumière, soit 3.00×10^8 m/s) tandis que le cerveau humain communique moins rapidement, mais contient beaucoup plus de neurones par rapport aux nombres de transistors que possède l'IA.

2.2-3-2- Les limites émotionnelles

L'homme est capable d'avoir des sentiments, de ressentir des choses, cela est très compliqué à reproduire. Les sentiments sont dus à la nature et non à une programmation.

2.2-3-3- Les limites cognitives

Les machines sont de plus en plus efficaces pour traiter des données, mais ce traitement comporte des points faibles. En effet, les ordinateurs fonctionnent sous système binaire, c'est-à-dire avec des 1 et des 0, appelés bytes, c'est en quelque sorte leur langage. A la différence du cerveau humain qui réfléchit en sollicitant son réseau, d'une complexité sans pareille et très étendue de neurones.

2.3- Outils et algorithmes mise en œuvre pour l'intelligence artificielle

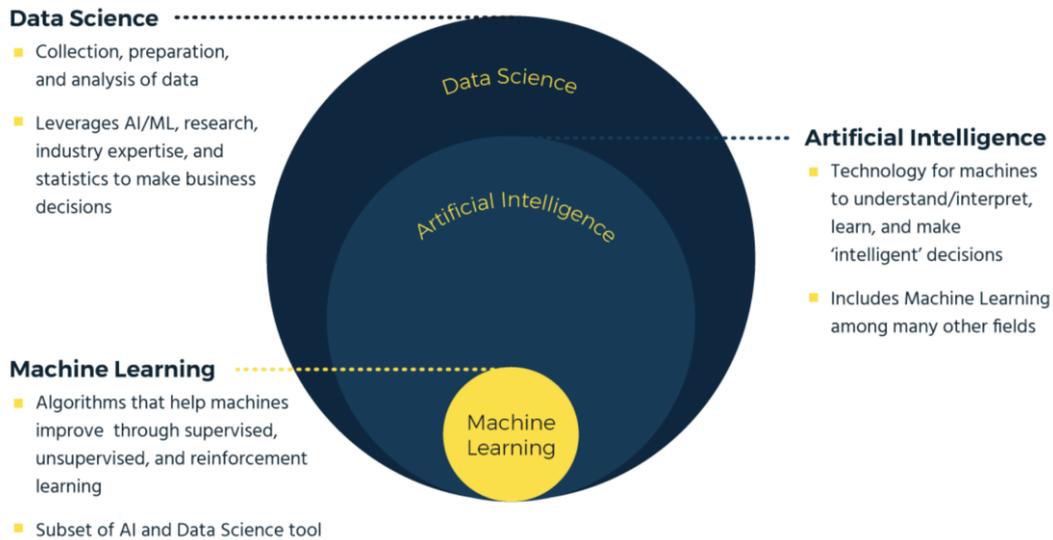


Figure 2.1 : Les sous-branches de Data Science

2.3-1- Techniques de l'apprentissage

Les scientifiques et afin de créer une intelligence artificielle, utilisent un ensemble de techniques différentes regroupées sous une même discipline qu'on appelle l'apprentissage machine (ou machine Learning 'en anglais'). Donc dire intelligence artificielle revient automatiquement à dire "Machine Learning", ou l'apprentissage automatique.

2.3-2- L'apprentissage automatique (Machine Learning)

La définition de l'apprentissage automatique selon Wikipédia (septembre 2020) est : « L'apprentissage automatique (en anglais machine Learning) littéralement « apprentissage machine » ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d' « apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être

explicitement programmés pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes. »

L'apprentissage automatique fait référence au développement, à l'analyse et à l'implémentation de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques.

★ Phase d'apprentissage

L'objectif de la Machine Learning est de laisser la machine construire son système de raisonnement sans avoir à imposer un programme au préalable. Pour **cette phase d'apprentissage**, la machine se base sur plusieurs exemples afin de comprendre la logique du modèle qu'elle doit intégrer. Pour se former, le Machine Learning va s'intéresser aux masses de données à analyser afin d'en déterminer l'algorithme de transformation. A partir des exemples étudiés, la Machine Learning commence ainsi sa phase d'apprentissage en toute autonomie.

★ Phase de prédiction

Après avoir intégré le raisonnement et l'algorithme du problème en question, la Machine Learning doit être en capacité de **déterminer la finalité d'une situation donnée**. Plus l'apprentissage de la Machine Learning est complet, plus les prédictions obtenues par cet outil seront précises.

★ Les données

Comme le suggère la définition proposée par Wikipédia, les algorithmes de l'apprentissage automatique sont basés sur des données.

Deux grandes familles de jeux de données peuvent être utilisées :

- les données étiquetées : chaque observation est fournie avec une étiquette
- les données non-étiquetées : comme le nom l'indique, aucune étiquette n'est fournie.

2.3-2-1- Comment est-ce que cela fonctionne ?

Le Machine Learning repose sur l'analyse, via des algorithmes, d'une somme de données issues de différentes sources (capteurs, autres machines...) et rendue plus facile par le développement du Big data ('un ensemble très volumineux de données qu'aucun outil classique de gestion

de base de données ou de gestion de l'information ne peut vraiment travailler (Bremme, 2015)). En appliquant des statistiques et probabilités à ces données, la machine va prendre une décision. Bonne ou mauvaise, cette dernière devra être soit validée, soit corrigée. Enrichissant de fait la base de statistiques, ce qui aidera le programme à faire, lorsque la même situation se représentera, le meilleur choix.

L'apprentissage automatique est considéré comme un sous-domaine de l'intelligence artificielle. Son objectif majeur est de comprendre la structure des données et de les faire intégrer dans des modèles simples que tout le monde peut comprendre et utiliser.

Bien que l'apprentissage automatique soit un domaine de l'informatique, il diffère de ces approches traditionnelles. Les algorithmes dans les approches informatiques sont des ensembles d'instructions programmées explicitement pour calculer et résoudre des problèmes bien définis, tandis que les algorithmes d'apprentissage automatique donnent aux ordinateurs la possibilité de s'entraîner sur un ensemble de données qui lui sont fournies comme entrées et utilisent l'analyse statistique pour produire des valeurs situées dans une plage spécifique comme sorties. Pour cela, 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.

De nos jours, l'apprentissage automatique est omniprésent dans toutes nos interactions avec les banques, nos achats en ligne et notre utilisation des médias sociaux. Les algorithmes d'apprentissage automatique entrent en jeu dans ses cas pour garantir la fluidité, l'efficacité et la sécurité de nos expériences.

Les technologies les plus récentes sont toutes dotées de l'apprentissage automatique et leurs utilisateurs bénéficient des avantages de cette nouvelle discipline. Par exemple, la technologie de reconnaissance faciale permet aux plateformes des réseaux sociaux d'aider les utilisateurs à marquer et partager des photos d'amis. La technologie de reconnaissance optique des caractères est capable de convertir les images du texte en caractères mobiles. Les moteurs de recommandation, qui sont alimentés par l'apprentissage automatique, suggèrent à l'utilisateur les films ou émissions de télévision à regarder en fonction de ses préférences.

2.3-2-2- Types d'apprentissage

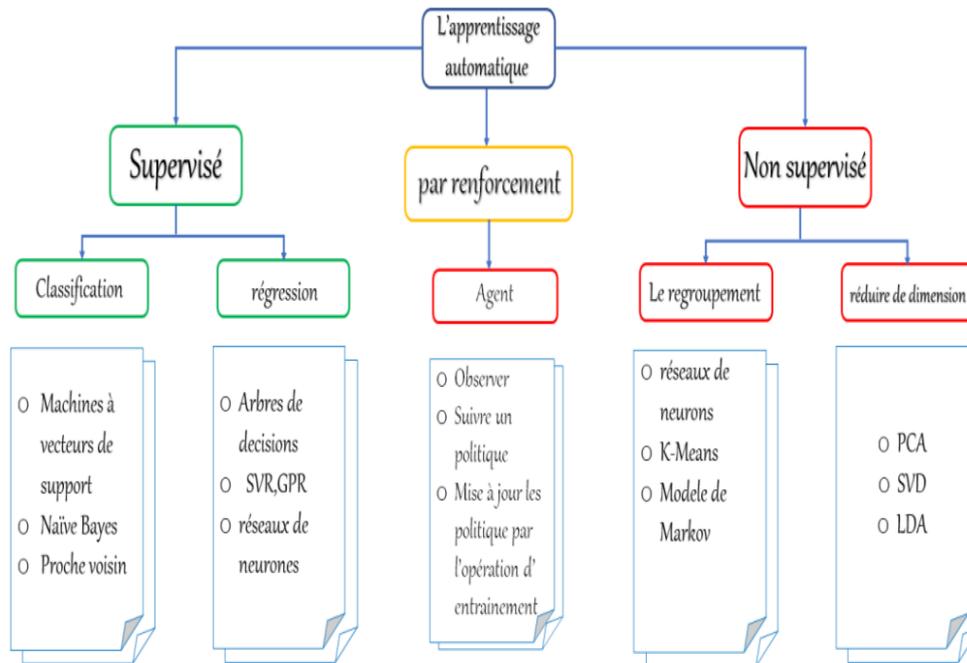


Figure 2.2 : Les approches et les algorithmes de l'apprentissage automatique

Il existe également de différents types d'apprentissage automatique. La définition du type d'apprentissage est basée sur la réponse à ces deux questions suivantes :

- Est-ce que cet apprentissage compte sur la supervision humaine dans son entraînement et apprentissage ?
- Est-ce que ce type cet apprentissage utilise une base de données fournie par l'être humain ?

- ✚ Si la réponse est oui pour les deux questions, nous avons un apprentissage supervisé.
- ✚ Si la réponse est non pour la première question, et oui pour la deuxième question, nous parlons d'un apprentissage non supervisé.
- ✚ Si la réponse est non pour les deux questions, le type de l'apprentissage est l'apprentissage par renforcement.

Dans ce qui suit, nous définissons chacun de ces types : apprentissage supervisé, apprentissage non supervisé, et apprentissage par renforcement.

2.3-2-2-1- L'apprentissage supervisé

Dans l'apprentissage supervisé l'être humain aide l'algorithme pour apprendre, un data scientist sert de guide et il apprend l'algorithme les résultats qu'il doit trouver. Le même cas lorsqu'on apprend un enfant pour identifier les fruits, en les mémorisant dans sa mémoire. Dans l'apprentissage supervisé, l'algorithme apprend grâce à un jeu de données déjà étiqueté et dont le résultat est prédéfini.

- ★ Un expert est employé pour étiqueter correctement des exemples.
- ★ L'apprenant doit alors trouver ou approximer la fonction qui permet d'affecter la bonne étiquette à ces exemples.



Figure 2.3 : *Apprentissage supervisé*

Les algorithmes de l'apprentissage automatique supervisé sont les plus couramment utilisés, il y a deux types d'apprentissage supervisé :

- **La classification :** la classification consiste à trouver le lien entre une variable d'entrée (X) et une variable de sortie discrète (Y), en suivant une loi multinomiale.
- **Régression :** la régression consiste à prédire une valeur continue pour la variable de sortie.

Les algorithmes les plus célèbres utilisés dans cette approche sont les suivants:

- **SVM (Machines à vecteurs de support) :**
Est un apprentissage automatique très puissant et polyvalent modèle, capable d'effectuer la classification linéaire ou non linéaire, la régression, et même détection des valeurs aberrantes. C'est l'un des modèles les plus populaires de l'apprentissage automatique

« Machine Learning », et n'importe qui intéressés par cette approche devraient l'avoir dans leur boîte à outils. Les SVM sont particulièrement bien adaptés à la classification d'ensembles de données complexes mais de petite ou moyenne taille. L'algorithme SVM consiste à chercher à la fois l'hyperplan optimal ainsi que de minimiser les erreurs de classification.

○ **Principe et algorithme**

En règle générale, les machines à vecteurs de support (SVM) sont considérées comme une approche de classification, mais elles peuvent être utilisées dans les problèmes de classification et de régression. SVM construit un hyperplan dans un espace multidimensionnel pour séparer au mieux les différentes classes. L'idée centrale du SVM est de trouver un hyperplan marginal maximum (MMH) qui divise au mieux l'ensemble de données en classes et pour pouvoir classer les nouveaux points par la suite. En fait, l'objectif principal est de séparer l'ensemble de données de la meilleure façon possible. La distance entre les deux points les plus proches est connue sous le nom de marge. L'objectif est de sélectionner un hyperplan avec la marge maximale possible entre les vecteurs de support dans l'ensemble de données. Pour chercher, l'hyperplan marginal maximal, le SVM commence par générer des hyperplans qui séparent au mieux les classes. Ensuite, il sélectionne l'hyperplan droit avec la ségrégation maximale des points de données les plus proches, comme indiqué dans la figure ci-dessous.

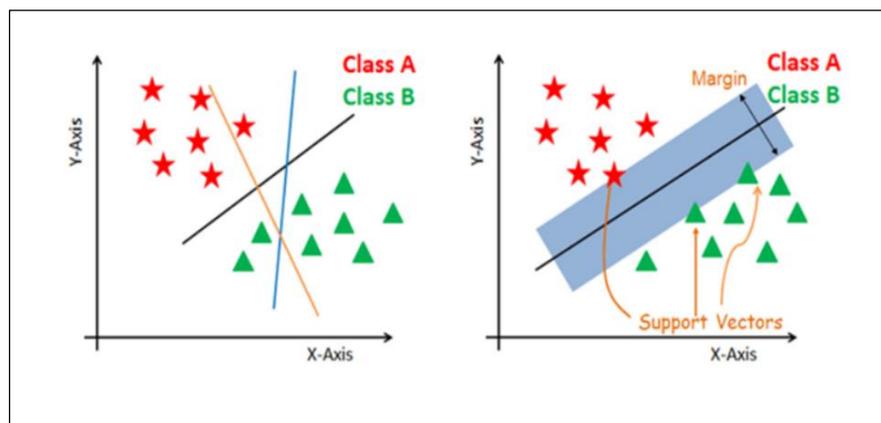


Figure 2.4 : *Support Vector Machine (SVM) tirée de Navlani (2018)*

- Implémentation en Python

Pour implémenter l'algorithme SVM, il y a quatre étapes. La première étape consiste à faire un prétraitement des données. Par la suite, il faut ajuster l'algorithme SVM à l'ensemble d'apprentissages. La troisième étape est celle de la prédiction de la classe de l'ensemble test. Ensuite, on crée une matrice de confusion pour pouvoir évaluer le modèle. Enfin, la dernière étape consiste à visualiser et interpréter le résultat.

Pour réaliser ces étapes, il existe de nombreuses bibliothèques disponibles en langage Python. Il suffit d'appeler les fonctions avec les paramètres adéquats choisis en fonction des besoins.

- **La méthode des k plus proches voisins** : cet algorithme consiste à essayer différentes valeurs de K pour obtenir la séparation la plus satisfaisante.

- Principe et algorithme

L'algorithme des K plus proches est l'un des algorithmes d'apprentissage le plus simple basé sur la technique d'apprentissage supervisé et peut être utilisé pour la régression ainsi que pour la classification, mais il est principalement utilisé pour les problèmes de classification. L'algorithme KNN suppose la similitude entre le nouveau cas/données et les cas disponibles et place le nouveau cas dans la catégorie la plus similaire aux catégories disponibles. Il est également appelé algorithme d'apprentissage paresseux, car il n'apprend pas immédiatement de l'ensemble d'entraînement, au lieu de cela, il stocke l'ensemble de données et au moment de la classification, il effectue une action sur l'ensemble de données. Ainsi, lorsqu'il obtient de nouvelles données, il classe ces données dans une catégorie très similaire aux nouvelles données. Le fonctionnement de l'algorithme KNN peut être expliqué sur la base de quatre grandes étapes. On commence par sélectionner le nombre K des voisins dont on calcule la distance euclidienne. Par la suite, on prend les K voisins les plus proches selon cette distance, et on compte le nombre de points de données dans chaque catégorie. Finalement, on affecte les nouveaux points de données à une catégorie pour laquelle le nombre de voisins est maximum.

Comme on l'a discuté, le principe de l'algorithme du classificateur KNN consiste à trouver K nombre prédéfini d'échantillons d'apprentissage les plus proches de la distance du nouveau point et à prédire son étiquette. La valeur minimale de K est 1. Cela signifie utiliser un seul voisin pour la prédiction. Le maximum est le nombre de points de données dont on dispose. La recherche a montré qu'aucun nombre optimal de voisins ne convient à tous les

types d'ensembles de données. Chaque jeu de données possède ses propres exigences. Dans le cas d'un petit nombre de voisins, le bruit aura une plus grande influence sur le résultat, et un grand nombre de voisins rend le calcul coûteux.

- **Naïve Bayes** : est un classifieur assez intuitif à comprendre. Il se base sur le théorème de Bayes des probabilités conditionnelles, et il suppose que les variables sont indépendantes entre elles. Cela permet de simplifier le calcul des probabilités.
- **Les arbres de décision** : un arbre de décision sert à classer les futures observations, sachant qu'un corpus d'observations est déjà étiqueté.
- **Les Forêts Aléatoires** : cet algorithme fonde sur les arbres de décision, un modèle construit par de multiples arbres de décisions.

- Principe et algorithme

La forêt aléatoire (RF) est un algorithme d'apprentissage supervisé. Elle peut être utilisée à la fois pour la classification et la régression. Les forêts aléatoires créent des arbres de décision sur des échantillons de données sélectionnés au hasard, obtiennent la prédiction de chaque arbre et sélectionnent la meilleure solution au moyen d'un vote. En effet, c'est une méthode d'ensemble (basée sur l'approche divisé pour régner) d'arbres de décision générés sur un ensemble de données divisé de manière aléatoire. Cette collection de classificateurs d'arbres de décision est également connue sous le nom de forêt. Les arbres de décision individuels sont générés à l'aide d'un indicateur de sélection d'attributs tels que le gain d'informations, le rapport de gain et l'indice de Gini pour chaque attribut. Chaque arbre dépend d'un échantillon aléatoire indépendant.

Dans un problème de classification, chaque arbre vote et la classe la plus populaire est choisie comme résultat final. La figure 2.5 illustre le processus. L'algorithme est décrit en quatre étapes. La première consiste à sélectionner les échantillons aléatoires à partir d'un ensemble de données donné. Dans la deuxième étape, on construit un arbre de décision pour chaque échantillon pour avoir ensuite un résultat de prédiction à partir de chaque arbre de décision. Par la suite, on effectue un vote pour chaque résultat prévu. Et enfin, la sélection du résultat de la prédiction finale se fait selon la prédiction possédant le plus de votes. Ce processus est illustré par la figure 2.5 :

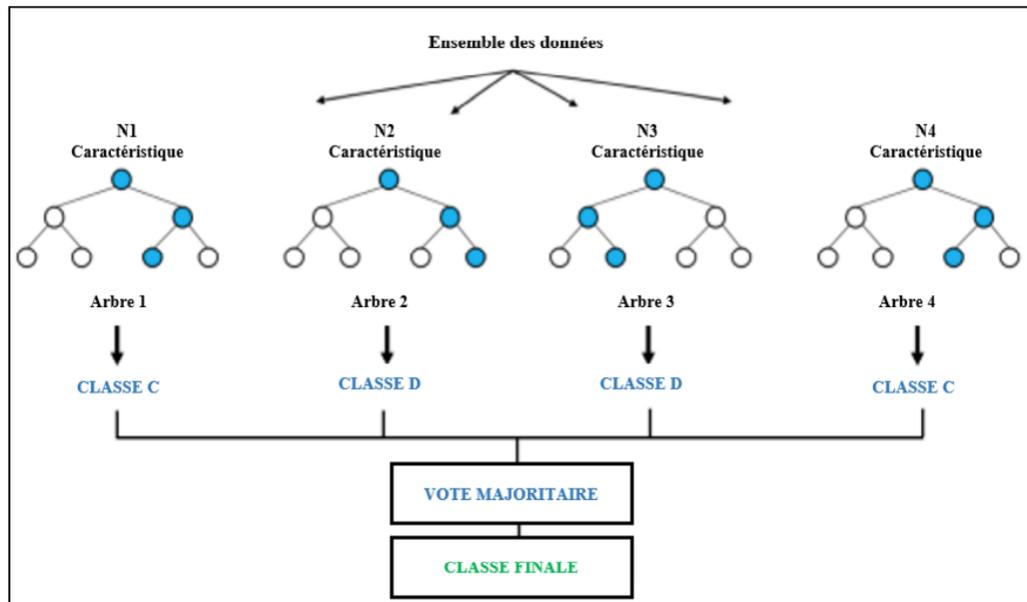


Figure 2.5 : Principe des forêts aléatoires adaptées de Kirasich, Smith, & Sadler (2018)

- Implémentation en Python

Pour implémenter l'algorithme RF, on commence par un prétraitement des données. On ajuste ensuite l'algorithme RF à l'ensemble d'apprentissages. On prédit le résultat de l'ensemble test. Après, on crée la matrice de confusion pour évaluer la performance du modèle utilisé. Enfin, on visualise et on interprète le résultat obtenu. Pour cela, on examinera l'algorithme « RFclassifier » se trouvant dans la bibliothèque « sklearn ». Cependant, pour la RF, on doit choisir le nombre d'arbres à créer et le nombre de variables à utiliser à chaque division d'un nœud. Les paramètres du modèle sont donc :

- `max_depth` : le `max_depth` d'un arbre dans la forêt aléatoire est défini comme le chemin le plus long entre le nœud racine et le nœud feuille.
- `n_estimators` : C'est le nombre d'arbres utilisés.

- **Régression Logistique** : l'algorithme de régression logistique consiste à trouver les meilleurs coefficients pour minimiser l'erreur entre la prédiction faite pour des destinations visitées et la vraie étiquette donnée (bon, mauvais etc.).
- **Les réseaux de neurones** : Les réseaux neuronaux sont un modèle informatique qui partage certaines propriétés avec le cerveau humain, dans lequel de nombreuses unités simples travaillent en parallèle sans centralisation, ils permettent de trouver des patterns complexes

dans les données, il se compose de valeurs d'entrées, poids, fonction de transfert et une valeur de sortie

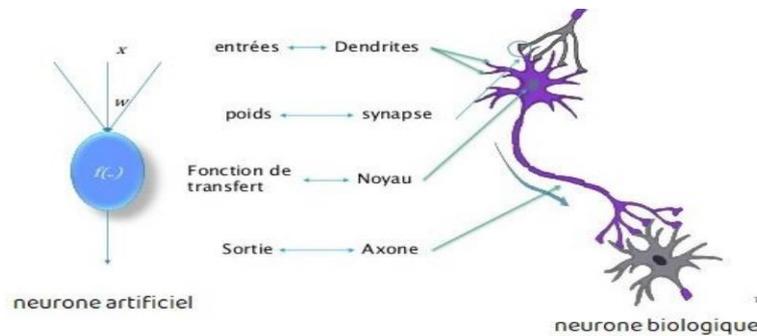


Figure 2.6 : *Neurone biologique et neurone artificiel*

Il existe aussi d'autres algorithmes, tels que l'algorithme de régression linéaire, les algorithmes Génétiques. Certains algorithmes de régression peuvent également être utilisés pour la classification, et la régression, à titre d'exemple l'algorithme de la régression logistique.

2.3-2-2-2- L'apprentissage non supervisé

Avec l'apprentissage non supervisé la machine n'a pas besoin de l'aide pour apprendre. L'apprentissage non supervisé est une approche plus indépendante, dans laquelle un ordinateur apprend à identifier des processus et des schémas complexes sans aucun guide, Il implique une formation basée sur des données sans étiquette, qui ne contiennent aucun résultat spécifique.

- Aucun expert n'est disponible.
- L'algorithme doit découvrir par lui-même la structure des données.
- Le clustering est un algorithme d'apprentissage non supervisés.



Figure 2.7 : *Apprentissage non supervisé*

Il y a deux types d'apprentissage non supervisé :

- **Regroupement(Clustering)** : c'est une méthode d'analyse statistique utilisée pour organiser des données brutes en silos homogènes, à l'intérieur de chaque grappe, les données sont regroupées selon une caractéristique commune.
- **Réduction de la dimension** : l'objectif est de simplifier les données sans perdre trop d'informations, à titre d'exemple, fusionner plusieurs caractéristiques en un seul caractère.

Les algorithmes les plus célèbres utilisés dans cette approche sont :

- **K-Moyenne** : est un algorithme de Regroupement (Clustering) il regroupe dans les même Cluster (Groupes) les données similaires (qui se ressemblent). Il utilise un raffinement itératif pour produire un résultat final.
- **Analyse de classification hiérarchique (HCA)** : la mise dans un cluster hiérarchique est similaire à la mise dans un cluster normal, sauf que dans ce cas nous souhaitons mettre en place une hiérarchie des clusters. Cela peut s'avérer très important surtout quand nous désirons une flexibilité par rapport au nombre de clusters voulu.
- **PCA (Analyse des composants principaux)** : l'algorithme PCA consiste à transformer des variables liées entre elles, vers de nouvelles variables séparées les uns des autres. Ces nouvelles variables sont nommées les composants principales, elles permettent au praticien de réduire le nombre de variables et de rendre l'information moins redondante.
- **Apriori** : l'algorithme Apriori s'utilise dans une base de données transactionnelle pour extraire des ensembles d'éléments fréquents, puis générer des règles d'association.

2.3-2-2-3- L'apprentissage par renforcement :

Avec l'apprentissage par renforcement la machine n'a pas besoin de l'aide de l'être humain, ni en termes de supervision, ni en termes de fourniture de données.

L'apprentissage par renforcement est une branche très différente. Le système d'apprentissage, appelé un agent dans ce contexte (Voir figure 4), peut observer l'environnement, sélectionner et effectuer des actions, et enfin obtenir des récompenses ou des pénalités (des récompenses négatives). La machine peut apprendre toute seule la meilleure stratégie à suivre, appelée une politique, pour obtenir plusieurs récompenses au fil du temps. Une politique définit l'action que l'agent devrait choisir lorsqu'il est dans une situation donnée.

- L'algorithme apprend un comportement étant donné une observation.
- L'action de l'algorithme sur l'environnement produit une valeur de retour qui guide l'algorithme d'apprentissage.

★ Pourquoi l'AR ?

- Il est très utile dans le cadre de problèmes où :
 - des stratégies comportementales efficaces sont inconnues a priori ou sont difficilement automatisables.
 - Lorsqu'il y a de l'incertain dans la manière dont l'environnement évolue.
- L'AR se distingue des autres approches d'apprentissage par plusieurs aspects :
 - L'apprentissage se fait sans supervision
 - Il repose sur le principe d'essai/erreur

★ Limitations du cadre de l'AR

- Quand les espaces d'états et d'actions sont très grands
- Mémoire nécessaire pour représenter une table Temps nécessaire pour pouvoir remplir efficacement cette table

★ Domaine d'applications

- Les jeux
 - Backgammon
 - Échecs

- Applications industrielles
 - Routage de paquets
 - Contrôle et maintenance de machines
 - Contrôle d'une flotte d'ascenseurs.....

2.3-2-3- Le choix d'un type d'apprentissage automatique

Avec la présence de différents types de classifieurs pour l'apprentissage automatique, l'opération de choix d'un type est une question typique « Quel algorithme dois-je utiliser ? ». La réponse à cette question varie les facteurs suivants :

- ✓ La taille, la qualité et la nature des données.
- ✓ Le temps de calcul disponible.
- ✓ L'urgence de la tâche.
- ✓ Le but d'utilisation de ces données.

La figure suivante, fournir des indications sur les algorithmes à essayer en premier selon les facteurs mentionnés ci-dessus.

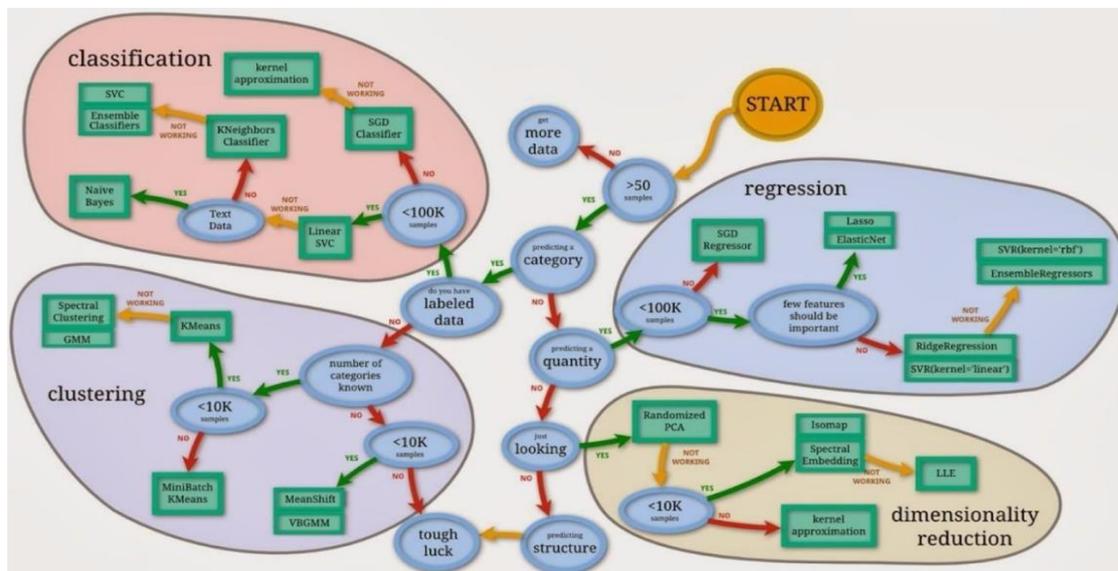


Figure 2.8 : Le choix de l'algorithme d'apprentissage selon certains facteurs

2.3-3- De l'apprentissage automatique à l'apprentissage en profondeur

Nous ne pouvons pas différer sur l'efficacité de l'apprentissage automatique dans la résolution d'une variété de problèmes, mais il y a beaucoup de problèmes que ce type fait face, dont les plus importants sont le temps, la vitesse et l'efficacité, si un algorithme d'apprentissage automatique renvoie une prédiction inexacte, alors un ingénieur doit intervenir et faire des ajustements, dans ce cas, il y a un manque d'efficacité qui sera une perte de temps avec une prévision lente.

Les algorithmes simples d'apprentissage automatique décrits dans ce chapitre fonctionnent très bien sur une grande variété de problèmes importants. Cependant, ils n'ont pas réussi à résoudre les problèmes centraux de l'intelligence artificielle, tels que la reconnaissance de la parole ou la reconnaissance d'objets. Le développement de l'apprentissage en profondeur a été motivé en partie par l'échec de algorithmes traditionnels, lorsque de travaille avec des données de grande dimension, pour bien généraliser sur de telles tâches l'intelligence artificielle.

2.3-3-1- L'apprentissage en profondeur « Deep Learning »

- ★ Le Deep learning, pour la reconnaissance

Le Deep learning est une notion beaucoup plus récente, apparue il y a une dizaine d'années. Cette technologie intervient pour sa part dans la reconnaissance par les machines du langage naturel et d'images. Elle fonctionne par biomimétisme : le Deep learning reproduit en effet le mécanisme des réseaux de neurones : il s'agit d'imiter le fonctionnement du cerveau humain !

Il avance ainsi grâce à une analyse par couches, chacune d'entre elles aidant la machine à comprendre l'élément (image ou suite de mots) qui se présente face à elle et éliminant, couche après couche, les réponses non-pertinentes. Le Deep learning est donc extrêmement gourmand en données, ce qui explique qu'il s'est développé en même temps que le phénomène du Big Data, que l'Internet des objets et l'industrie 4.0.

L'apprentissage en profondeur « Deep Learning » a été un défi à définir pour beaucoup spécialiste dans le domaine car il a changé de forme lentement au cours de la dernière décennie. Une définition utile précise que l'apprentissage en profondeur est un réseau neuronal avec plus de deux couches. Le problème avec cette définition est qu'elle fait écho à l'existence de ce domaine depuis les années 80 du siècle dernier, formant ainsi une grande contradiction puisque beaucoup de gens

pensent que ce domaine est relativement nouveau, pour réfuter cette contradiction, il faut distinguer le moment où le domaine est apparu et celui où il a été cadré et exploité.

2.3-3-1-1- Définition d'apprentissage en profondeur

Le terme « apprentissage profond » a été introduit dans le domaine de l'apprentissage automatique par « Rina Dechter » en 1986, et dans les réseaux de neurones artificiels par « Igor Aizenberg » et ses collègues en 2000, dans le contexte des neurones à seuil booléen (Schmidhuber, 2015), l'apprentissage profond désigne une technique d'apprentissage d'une machine, c'est une sous-branche de l'intelligence artificielle qui vise à construire automatiquement des connaissances à partir de grandes quantités d'information. Les caractéristiques essentielles du traitement ne seront plus identifiées par un traitement humain dans un algorithme préalable, mais directement par l'algorithme d'apprentissage profond.

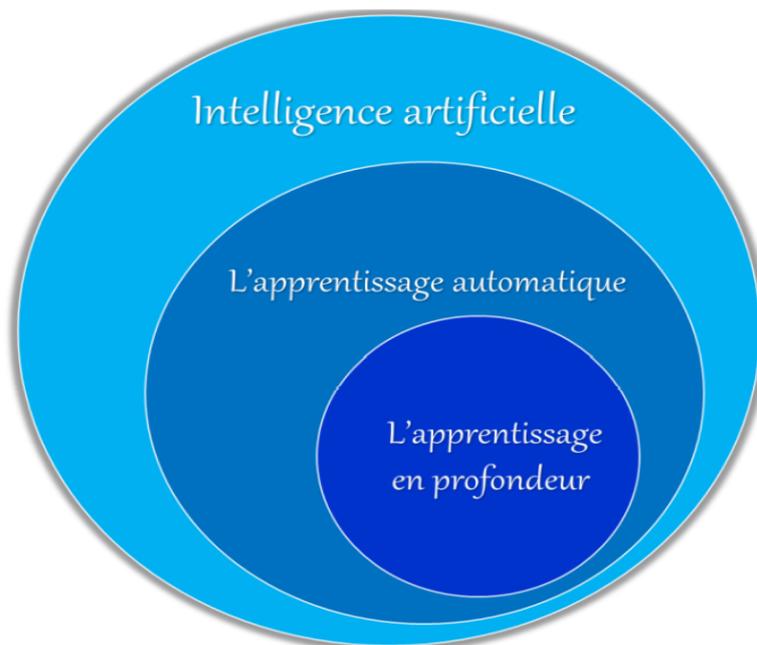


Figure 2.9 : *Les sous-branches de l'intelligence artificielle*

L'apprentissage en profondeur permet donc implicitement de répondre à des questions du type « que peut-on déduire de ces données ? » et décrire des caractéristiques parfois cachées ou des

relations entre des données souvent impossibles à identifier pour l'homme. L'apprentissage profond est un réseau neuronal avec un grand nombre de paramètres et de couches.

★ **Certaines de ses caractéristiques:**

- Plus de neurones.
- Des moyens plus complexes de connecter les couches neurones dans les réseaux neuronaux.
- Puissance de calcul.
- Extraction automatique des fonctionnalités.

★ **L'apprentissage profond s'applique dans divers domaines, tel que :**

- L'intelligence artificielle en général.
- La reconnaissance visuelle et la comparaison de forme.
- La robotique.
- La santé et la bio-informatique.
- La sécurité.

2.3-3-1-2- Les architectures d'apprentissage en profondeur

Les trois grandes architectures de réseaux profonds:

- ✓ Réseaux de neurones pré-entraînés non supervisés.
- ✓ Réseaux neuronaux convolutionnels (CNN).
- ✓ Réseaux neuronaux récurrents (RNN).

Habituellement, de nombreuses catégories considèrent les réseaux de neurones supervisés comme appartenant au domaine d'apprentissage automatique plutôt que d'apprentissage en profondeur.

2.3-3-2- Les réseaux de neurones artificiels

2.3-3-2-1- Définition

Un réseau de neurones artificiels, ou (Artificial Neural Network en anglais), est un système informatique matériel et / ou logiciel s'inspirant du fonctionnement du cerveau humain pour apprendre. Il s'agit d'une variété de technologie Deep Learning, qui fait elle-même partie de la sous-

catégorie d'intelligence artificielle et du Machine Learning. Ce genre de réseau est défini par un ensemble de couches de neurones qui sont fortement interconnectées entre elles.

2.3-3-2-2- Architecture d'un réseau de neurones artificiels

En général un réseau de neurones est constitué d'un ensemble de couches successives dont chacune prend ses entrées sur les sorties de la précédente c.-à-d. que cet ensemble est entièrement connectée. Chaque couche est un ensemble de neurones n'ayant pas de connexion entre eux et qui reçoivent des informations numériques en provenance de neurones voisins. L'ensemble de couches est composé d'une couche d'entrée qui lit les valeurs d'entrées, une couche de sortie qui fournit les résultats du système et entre ces deux se cache une à plusieurs couches dites cachées qui participent au transfert. Comme le montre la figure suivante :

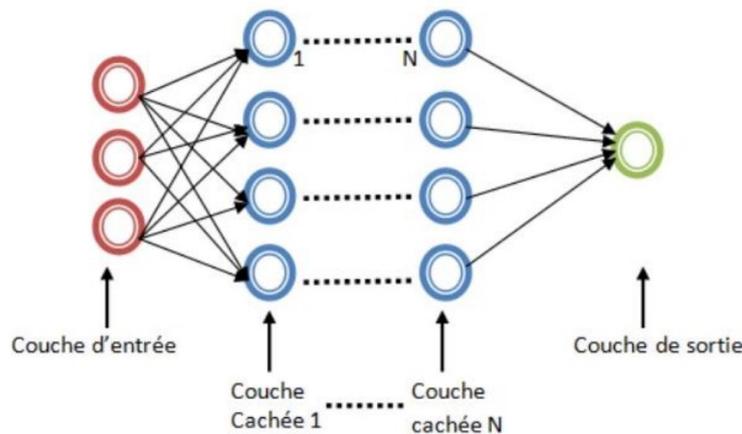


Figure 2.10 : Architecture de base d'un réseau de neurones artificiel

2.3-3-2-3-Principe de fonctionnement d'un réseau de neurones artificiel

Définissons maintenant le principe de fonctionnement d'un réseau de neurone en représentant les grandes étapes réalisées lors l'utilisation de ce réseau :

- On attribue à chaque neurone constituant le réseau un coefficient qu'on appelle poids.

- Chaque donnée d'entrée (input) va être multipliée par ce coefficient et on va appliquer une fonction d'addition à l'ensemble des résultats. Cette fonction est appelée fonction de combinaison.
- Puis on va appliquer une fonction d'activation, en comparant la somme obtenue à un seuil déterminé. Si la somme est en dessous du seuil (souvent dans le cas où la somme est négative), le neurone ne s'active pas, ce qui signifie que la donnée n'est pas intéressante. Si cette somme est au-dessus du seuil (souvent dans le cas où la somme est positive), alors le neurone s'active et envoie l'information au neurone de la couche suivante (couches cachées), jusqu'à ce que la donnée ultime atteigne le dernier neurone. On aura alors un résultat final (output).

Ce principe est illustré dans la figure suivante :

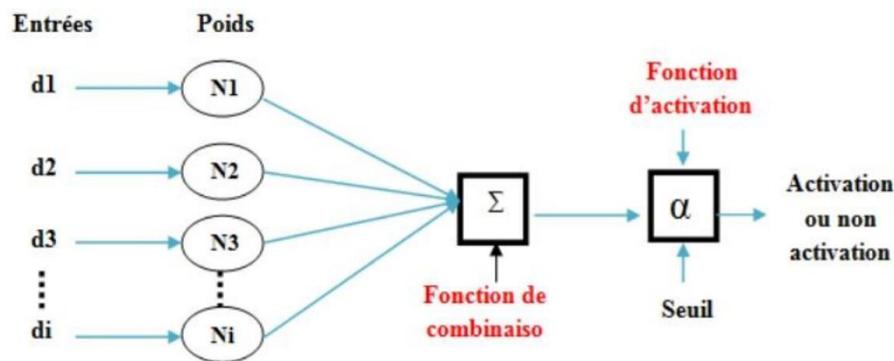


Figure 2.11: Principe de fonctionnement d'un réseau de neurones artificiel

2.3-3-2-4- Les types de réseaux de neurones

Il existe différents types de réseaux de neurones, et ils sont classés en fonction du nombre de nœuds cachés du modèle ou encore du nombre d'entrées et de sorties de chaque nœud.

La propagation des informations entre les différents neurones peut varier et dépend du type de réseaux de neurones.

- Les réseaux de neurones dit "feed-forward" (à propagation avant): c'est la variante la plus simple, l'information ne se déplace que dans une seule direction, elle traverse directement

l'entrée aux nœuds de traitement (couches cachées) puis aux sorties, avec absence de cycle ou de boucle dans le réseau.

- Les réseaux de neurones récurrents: Ce mode d'apprentissage est un peu plus complexes, ils comportent au moins un cycle dans leurs structures, ils sauvegardent les résultats produits par les nœuds de traitement et nourrissent le modèle à l'aide de ces résultats. Parmi ses applications, on trouve: la reconnaissance automatique de formes, la traduction automatique de la parole...etc.
- Les réseaux de neurones convolutifs : leur fonctionnement est inspiré par un processus biologique qui est le cortex visuel des animaux, les données à traiter traversent une multitude de couches (de la couche d'entrée vers la couche de sortie). Ce type de réseaux s'applique dans différents domaines: la reconnaissance faciale, la reconnaissance d'images et vidéos, le traitement du langage naturel ...etc.

2.3-3-3- Domaines d'applications du Deep Learning

Le Deep Learning touche un grand nombre de domaine d'applications, dont on peut citer :

- **La reconnaissance faciale**

Un algorithme de Deep Learning apprend à détecter sur une photo les caractéristiques du visage tels que: les yeux, la bouche, le nez...etc. Et cela en fournissant à l'algorithme un ensemble d'images du visage qui vont être utilisées pour l'entraînement, et à force de les entraîner, il sera capable de détecter un visage sur une image.

- **La détection d'objets**

Les algorithmes de détection d'objets sont capables maintenant d'identifier au pixel près un élément ou une personne sur une image qui contient beaucoup d'éléments (images complexes).

- **La Traduction automatique**

La traduction automatique a fait d'immenses progrès grâce à l'utilisation des réseaux de neurones. Cependant, ce type d'IA a besoin de gros volumes de contenus traduits au préalable par des humains. Elle s'appuie sur l'apprentissage supervisé par lequel la machine fait une supposition puis reçoit la réponse correcte d'un humain, ce qui lui permet ensuite d'ajuster son traitement en conséquence. Cette méthode est efficace pour les langues très répandues, comme l'anglais ou le français, pour lesquelles il existe de nombreux documents parallèles.

- **La reconnaissance d'images**

La reconnaissance d'images en fonction de ce qu'elles représentent (par exemple, des voitures, des êtres humains) est une tâche difficile pour un ordinateur, il doit d'abord effectuer une analyse afin d'extraire les caractéristiques de chaque image. Avec le Deep Learning l'information d'entrée (ici une image) est analysée couche après couche, le résultat de l'analyse d'une couche est transmis à la couche suivante. Les informations ainsi collectées sont cartographiées de façon à constituer un algorithme flexible, et donc l'ordinateur est capable, à l'aide de plusieurs opérations de déterminer si une image appartient à la catégorie voiture ou bien être humain.

2.3-3-4- Exemple du Deep Learning :

Ils ont étudié un exemple concret de la reconnaissance d'images, à l'aide des réseaux de neurones, et qui consiste à reconnaître les photos d'un chat afin d'illustrer cette notion.

Tout d'abord, le réseau de neurones prend en entrée un ensemble d'images de chats différents mélangés avec des objets qui ne sont pas des chats, puis il va effectuer une analyse de chaque image afin de s'entraîner pour pratiquer le Deep Learning. Ces images sont ensuite converties en données et transférées sur le réseau. Ensuite les neurones artificiels attribuent un poids aux différents éléments. Enfin la couche finale de neurones va rassembler les différentes informations pour déduire le résultat s'il s'agit d'une image d'un chat ou non.

Le réseau de neurones va ensuite effectuer une comparaison entre cette réponse et les bonnes réponses proposées par les humains. Si les réponses sont pareilles, le réseau garde cette réussite en mémoire et s'en servira plus tard pour reconnaître les chats. Dans le cas contraire, le réseau ajuste le poids placé sur les différents neurones afin de corriger son erreur.

Ce processus représenté dans la Figure 10 est répété des milliers de fois jusqu'à ce que le réseau arrive à reconnaître un chat sur une photo dans toutes les circonstances.

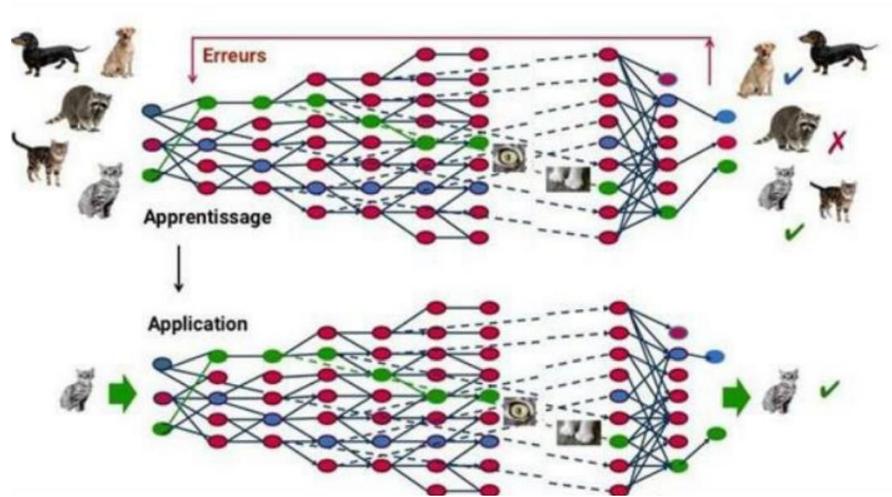


Figure 2.12 : le processus d'auto-apprentissage pour l'identification d'un chat

2.3-3-5- Comparaison entre l'apprentissage automatique et l'apprentissage en Profondeur

Tableau 2.1 : Comparaison entre l'apprentissage automatique et l'apprentissage en profondeur

	L'apprentissage automatique	L'apprentissage en profondeur
Dépendances des données	Une performance excellente avec des petites bases de données	Une performance excellente avec les BIG-DATA ¹
Dépendances du matériel	Travail sur une machine faible	Besoin d'une machine fort avec a un CPU ² fort aussi
Les caractéristiques des données	Besoin de comprendre les caractéristiques des données	Ne besoin pas de comprendre les caractéristiques des données
Le temps d'exécution	Quelque minute à des heures	Des semaine
Interprétabilité	Certains algorithmes sont faciles (Arber de disions)et d'autres sont impossibles (SVM, XGBoost ³)	Difficile à impossible

2.3-4- Les domaines d'applications de l'apprentissage automatique

L'intelligence artificielle, définie comme intelligence présentée par les machines, a de nombreuses applications dans la société d'aujourd'hui. Elle a été utilisée pour développer et faire progresser de nombreux domaines et industries, y compris la santé, le secteur financier et bancaire, le transport, le commerce, l'industrie, l'agriculture intelligente, les médias, le droit...etc.

2.3-4-1- La santé

C'est l'un des secteurs les plus dynamiques en intelligence artificielle. Aujourd'hui un grand nombre de systèmes se basant sur l'intelligence artificielle commence à diagnostiquer des maladies, notamment les cancers, d'une façon aussi bien voire mieux que les spécialistes.

L'intelligence artificielle peut favoriser la détection précoce d'une maladie et aussi pourrait accélérer la mise au point de nouveaux médicaments et réduire le temps écoulé entre la découverte d'une nouvelle molécule et sa mise sur le marché

2.3-4-2- Le secteur bancaire et financier

Un des principaux usages de l'IA est:

- la mise en place des agents conversationnels pour répondre aux questions des clients, sur la base de milliers de conversations analysées et enregistrées.
- Aider les conseillers dans leur décision d'accorder ou non un prêt à une personne et les aider à traiter en priorité les mails urgents adressés par les clients (sinistre, perte ou vol de carte).
- La détection des opérations frauduleuses, de plus en plus nombreuses et complexes à débusquer, est facilitée par le recours à l'IA.

2.3-4-3- Le transport

Les voitures d'aujourd'hui ont des fonctions d'aide à la conduite basées sur l'IA telles que le stationnement automatique. L'intelligence artificielle a été utilisée dans ce domaine afin de pouvoir optimiser les applications de gestion du trafic, ce qui permet de réduire les temps d'attente, la consommation d'énergie et les émissions.

Cette utilisation devrait fournir aussi un transport sûr, efficace et fiable tout en essayant de minimiser l'impact sur l'environnement. À l'avenir, des voitures entièrement autonomes sont envisagées à être développées.

2.3-4-4- Le commerce

L'intelligence artificielle a un impact énorme sur le commerce. Aujourd'hui, les robots effectuent plusieurs tâches d'aide, par exemple la vérification de l'inventaire, la surveillance des planchers qui sont sales et les nettoyer, faire le suivi des étiquettes de prix, le travail à la caisse et beaucoup plus. Toutes ces tâches sont réalisées par des machines afin de faire face aux défis de la main-d'œuvre que les supermarchés rencontrent comme difficultés.

2.3-4-5- L'industrie

Dans le monde de l'industrie aussi, la machine Learning permet d'ajuster le calendrier de maintenance en fonction des données collectées directement sur les équipements, pour éviter au maximum les pannes et le manque à gagner important qu'elles génèrent pour l'entreprise. L'analyse de données permet également d'optimiser les cycles de production complexes de certaines usines.

2.3-4-6- L'agriculture intelligente

Dans le domaine agricole, le déploiement de l'IA se fait de plusieurs manières, D'abord, elle aide les agriculteurs à surveiller leurs champs et ça grâce à l'utilisation des capteurs, ensuite à l'aide des techniques d'apprentissage automatique et le Deep Learning, une analyse très précise est faite sur ces informations captées.

Par exemple, dans le cas de la reconnaissance d'images de plantes, les techniques d'IA aident à détecter certains défauts comme la présence de ravageurs sur les feuilles.

On peut aussi trouver d'autres applications de l'IA dans l'agriculture comme l'augmentation des rendements. Ceux-ci s'appuient sur des paramètres comme le climat, l'état des semences et des sols, les niveaux d'irrigation etc... pour aider les agriculteurs à savoir quoi planter, où et quand, quels plants surveiller et quand récolter.

2.3-4-7- Les Médias

L'analyse de contenu multimédia audiovisuel tel que des films, des programmes télévisés, des vidéos publicitaires ou les contenus générés par les utilisateurs présente un domaine vers lequel certaines applications d'IA se sont orientées. Les solutions impliquent souvent l'utilisation de la vision par ordinateur, qui représente un domaine d'application majeur de l'IA.

Les scénarios de tels cas d'utilisation comprennent l'analyse d'images à l'aide des techniques de reconnaissance d'objets ou de reconnaissance faciale, ou l'analyse de vidéos pour la reconnaissance de scènes. La motivation pour l'utilisation de l'analyse des médias basée sur l'intelligence artificielle peut être, entre autres:

- La facilité de la recherche multimédia.
- La création d'un ensemble de mots-clés descriptifs pour un élément multimédia.
- La reconnaissance automatique de la parole à des fins d'archivage ou à d'autres fins.

2.3-4-8- Le droit

L'intelligence artificielle intervient aussi dans le domaine de droit. Elle fournit un ensemble d'algorithmes qui sont susceptibles d'analyser une jurisprudence afin de déterminer les arguments clés, fournir un conseil juridique sur la qualité d'un contrat ou la probabilité de gagner un contentieux et même d'estimer le montant des indemnités à espérer.

2.3-5- Exemple dans le domaine de la sante utilisant Le Machine Learning

« Prédire le diabète avec ML »

2.3-5-1- Les outils de développement et Platform utilisé

- ★ L'outil de programmation utilisé est «Google Colab ».

2.3-5-1-1- Google "Colaboratory" :

L'outil "Colaboratory" de Google est un outil collaboratif de développement. Basée sur la plateforme Jupyter et écrit dans le langage Python, cette plateforme permet à un groupe de développeurs d'écrire du code pour un même projet grâce à un système de Netbooks et de cellules de code. L'outil est pleinement intégré dans la famille des outils Google. Il est possible de faire appel aux autres services pour utiliser un fichier local ou alors de mettre en ligne un fichier depuis son ordinateur. Le service de stockage Google Drive permet de conserver en ligne ses fichiers afin de les retrouver sur n'importe quel appareil lié à son compte Google. Il est possible de monter son lecteur Google Drive dans l'interface de Colobotary. Pour effectuer cette opération, ouvrez une cellule de code et copiez le code suivant à l'intérieur Colaboratory, souvent raccourci en "Colab", vous permet d'écrire et d'exécuter du code Python dans votre navigateur, avec :

- Aucune configuration requise

- Accès gratuit aux GPU
- Partage facile Colab vous permet d'importer un ensemble de données d'images, d'entraîner un classificateur d'images sur cet ensemble et d'évaluer le modèle, tout cela avec quelques lignes de code. Les notebooks Colab exécutent ce code sur les serveurs cloud de Google. Vous avez donc à votre disposition toute la puissance du matériel Google, y compris les GPU et TPU, quelle que soit la puissance de votre ordinateur. Vous n'avez besoin que d'un navigateur.

2.3-5-1-2- Bibliothèques utilisées

➤ Numpy

NumPy est l'abréviation de « Numerical Python » et c'est un ensemble fondamental pour le calcul scientifique en Python. NumPy fournit à Python une vaste bibliothèque mathématique capable d'effectuer des calculs numériques de manière efficace et efficiente afin de pouvoir travailler avec des tableaux multidimensionnels et des structures de données matricielles, très courante dans les domaines de la science des données et de l'apprentissage automatique.

- Pourquoi utiliser NumPy ?
 - Les tableaux Numpy utilisent moins de mémoire que les listes python normales.
 - Une liste python normale est un groupe de pointeurs pour séparer les objets Python (par exemple les nombres à l'intérieur de la liste).
 - Un tableau numpy est conçu pour être un tableau de valeurs uniformes, sans utiliser d'espace pour les pointeurs de type.
 - Numpy peut également lire les informations plus rapidement et dispose de nombreuses opérations de diffusion pratiques qui peuvent être effectuées à travers les dimensions du tableau.

➤ Pandas

Pandas est un package pour la manipulation et l'analyse de données en Python. Le nom Pandas est dérivé du terme « Panel Data ». Pandas intègre deux structures de données supplémentaires dans Python, à savoir Pandas Series et Pandas DataFrame. Ces structures de données nous permettent de travailler avec des données étiquetées et relationnelles de manière simple et intuitive. Le succès récent des algorithmes d'apprentissage automatique est en partie dû aux énormes quantités de données dont nous disposons pour former nos algorithmes. Cependant, lorsqu'il s'agit de données, la quantité n'est pas la seule chose qui compte, la qualité des données est tout aussi importante. Il arrive souvent que de grands ensembles de données ne soient pas prêts à être intégrés dans les algorithmes d'apprentissage. Le plus souvent, les grands ensembles de données auront souvent des valeurs manquantes, des valeurs aberrantes, des valeurs incorrectes, etc... Avoir des

données avec beaucoup de valeurs manquantes ou mauvaises, par exemple, ne va pas permettre aux algorithmes d'apprentissage automatique de bien fonctionner. Par conséquent, une étape très importante de l'apprentissage automatique consiste à examiner d'abord les données et à s'assurer qu'elles sont bien adaptées à votre algorithme de formation en effectuant une analyse de base des données. C'est là que les pandas entrent en jeu. Les Pandas Series et les DataFrames sont conçus pour une analyse et une manipulation rapides des données, tout en étant flexibles et faciles à utiliser.

- Voici quelques caractéristiques qui font des Pandas un excellent ensemble pour l'analyse des données :
 - Permet l'utilisation d'étiquettes pour les lignes et les colonnes.
 - Peut calculer des statistiques continues sur des données de séries chronologiques.
 - Manipulation aisée des valeurs NaN.
 - Est capable de charger des données de différents formats dans des DataFrames.
 - Peut réunir et fusionner différents ensembles de données.
 - Il s'intègre avec NumPy et Matplotlib. Pour ces raisons, entre autres, les Pandas DataFrames sont devenus l'un des objets Pandas les plus utilisés pour l'analyse des données en Python.

➤ Matplotlib

Ce paquet est la bibliothèque Python la plus populaire actuellement pour produire des tracés et autres visualisations de données en 2D. Comme l'analyse des données nécessite des outils de visualisation, c'est la bibliothèque qui convient le mieux à cette fin. On peut facilement générer des graphiques, des histogrammes, des diagrammes à barres, des diagrammes de dispersion et bien d'autres en utilisant Python et quelques lignes de code. Au lieu de passer du temps à chercher des solutions, nous pouvons nous concentrer sur la génération de diagrammes pour une analyse et une exploration plus rapides des données.

- Parmi toutes les caractéristiques qui en ont fait l'outil le plus utilisé dans la représentation graphique des données, il y en a quelques-unes qui ressortent :
 - Développement progressif et visualisation interactive des données.
 - Un contrôle plus strict des éléments graphiques.
 - Exporter vers de nombreux formats, tels que PNG, PDF, SVG et EPS. Les fonctionnalités de traçage intégrées de Seaborn et de Pandas sont toutes deux construites à l'aide de matplotlib.

➤ Seaborn Seaborn

Est une bibliothèque de visualisation construite à l'aide de matplotlib qui se concentre sur la création de tracés statistiques standard avec un simple appel de fonction d'une ligne. Cela se fait au prix de moins d'options de personnalisation que le pur matplotlib. Cependant, on peut toujours modifier les attributs des tracés créés dans seaborn en ajoutant des appels de fonctions dans matplotlib.pyplot.

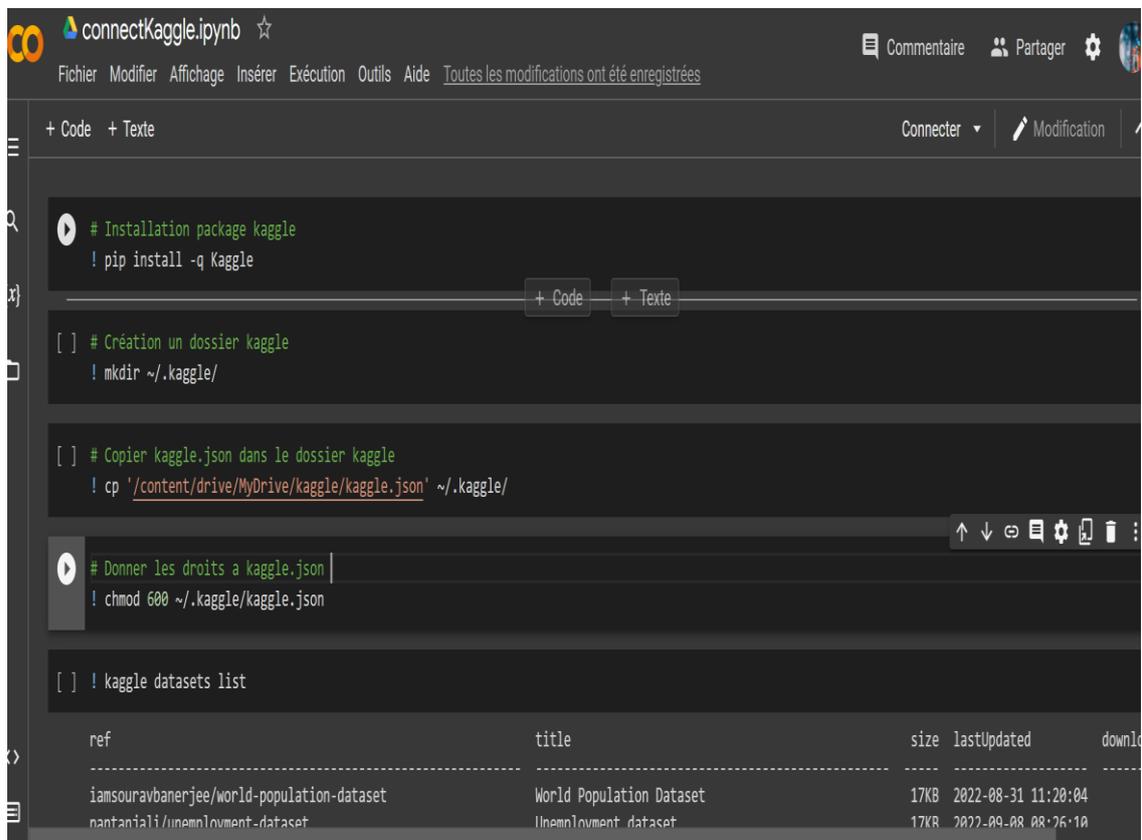
➤ Scikit-Learn

Scikit-Learn est une librairie d'apprentissage automatique construite à partir de NumP, SciPy et Matplotlib. Scikit-Learn offre des outils simple et efficace pour les taches commune en analyse de données, telle que la classification, la régression, le clustering, réduction de la dimensionnalité, et le prétraitement.

2.3-5-1-3- Prétraitement de la collection de données

Afin d'implémenter et évaluer un modèle d'apprentissage automatique pour prédire le diabète nous utiliserons la collection de données proposé par le site Ensemble de données sur le diabète - Indiens Pima, la collection est héberger sur Kaggle via le lien suivant (<https://www.kaggle.com/datasets/nancyalaswad90/review>).

- Selon le rapport des Centers of Disease Control and Prevention, environ un adulte sur sept aux États-Unis souffre de diabète. Mais d'ici quelques années, ce taux peut augmenter. Dans cet esprit, aujourd'hui, On montera comment vous pouvez utiliser l'apprentissage automatique pour prédire le diabète à l'aide de Python.



```
connectKaggle.ipynb
Fichier Modifier Affichage Insérer Exécution Outils Aide Toutes les modifications ont été enregistrées

+ Code + Texte
Connecter Modification

# Installation package kaggle
! pip install -q Kaggle

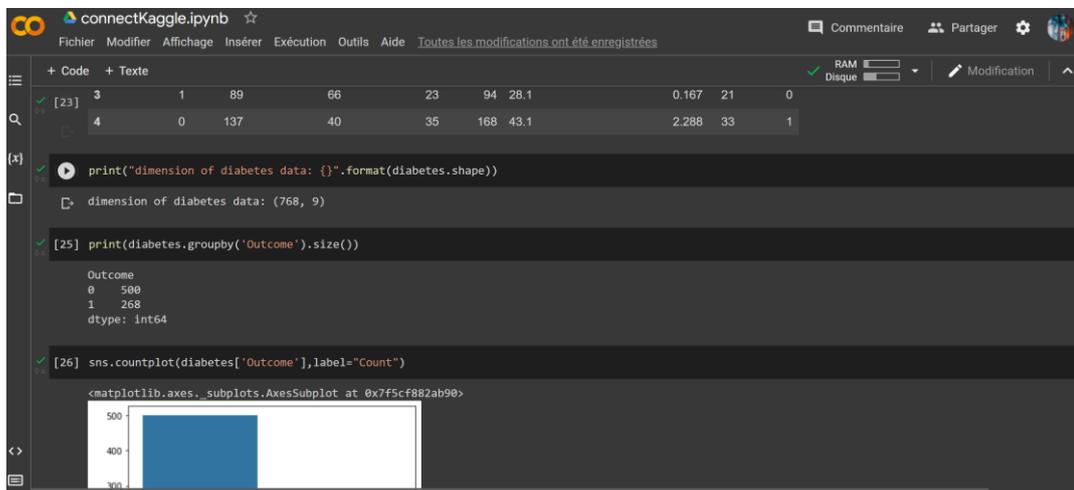
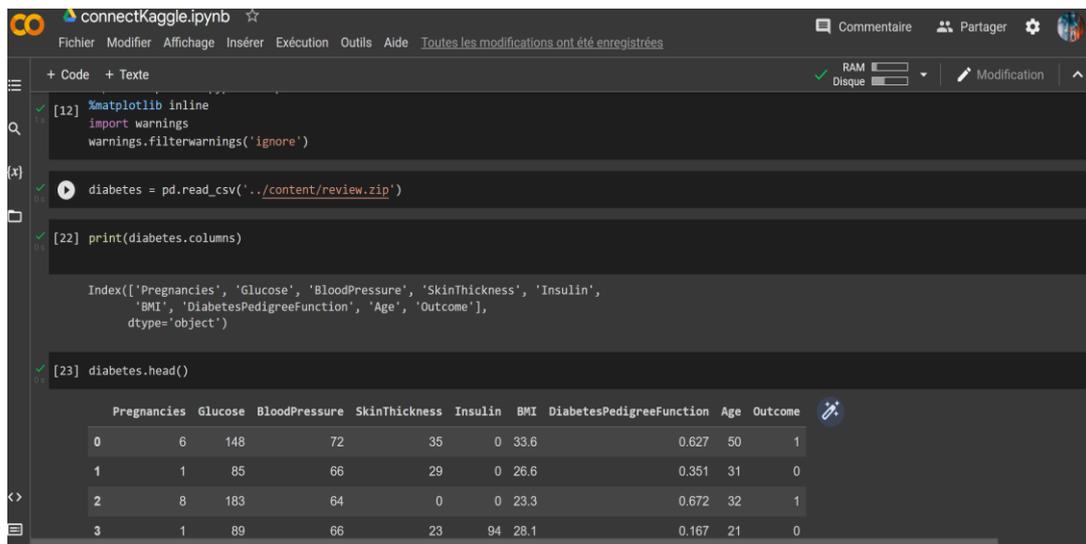
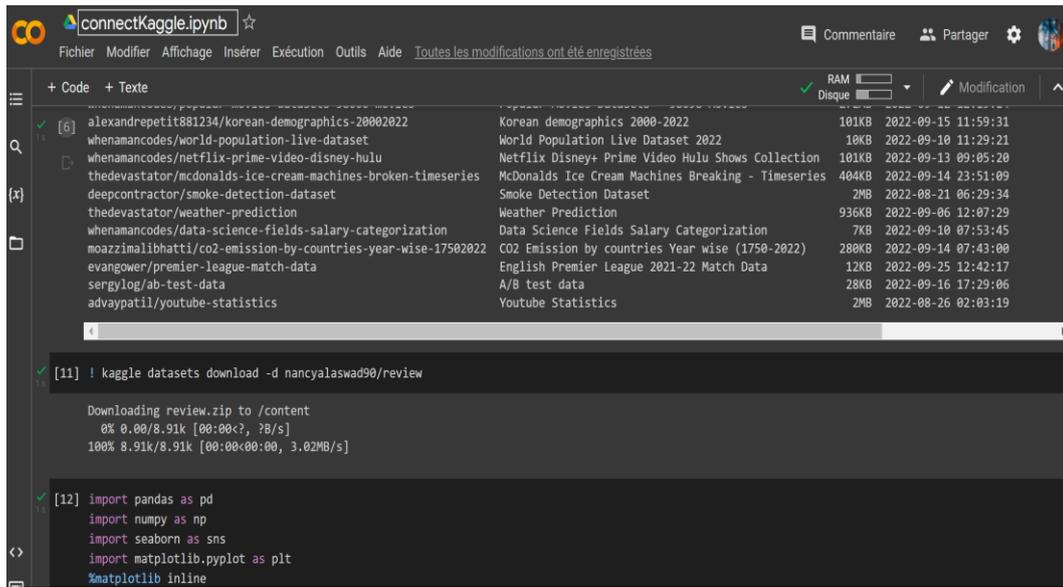
# Création un dossier kaggle
! mkdir ~/.kaggle/

# Copier kaggle.json dans le dossier kaggle
! cp '/content/drive/MyDrive/kaggle/kaggle.json' ~/.kaggle/

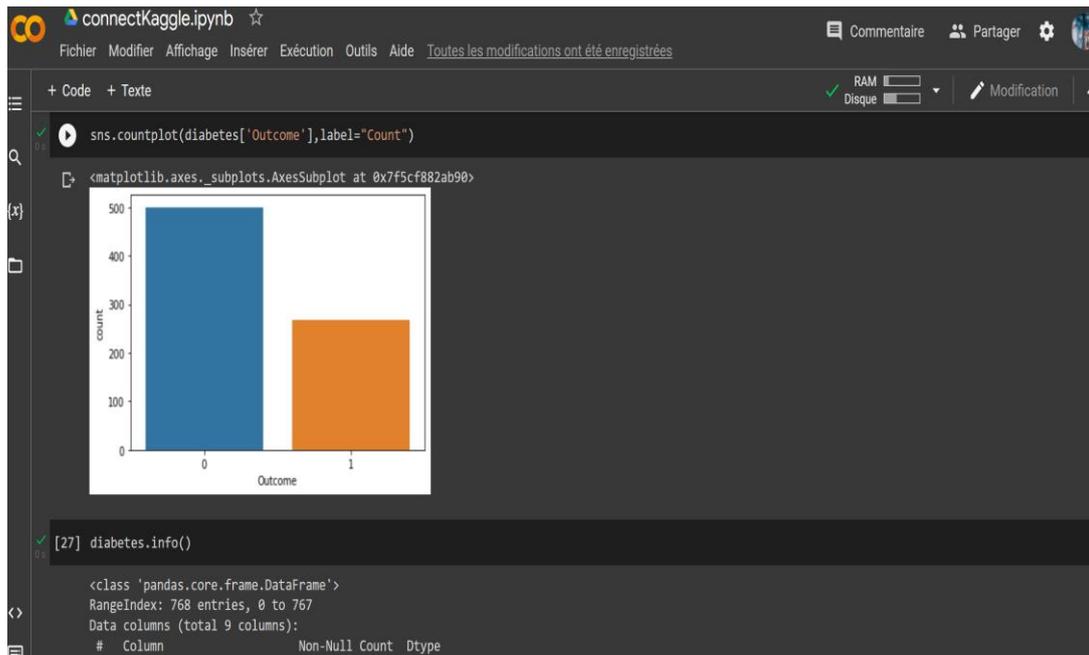
# Donner les droits a kaggle.json
! chmod 600 ~/.kaggle/kaggle.json

! kaggle datasets list
```

ref	title	size	lastUpdated	download
iamsouravbanerjee/world-population-dataset	World Population Dataset	17KB	2022-08-31 11:20:04	
nantanali/unemployment-dataset	Unemployment dataset	17KB	2022-09-08 08:26:10	



- L'ensemble de données sur le diabète se compose de 768 points de données, avec 9 caractéristiques chacun :
- Le "outcome" ou 'résultat' est la caractéristique que nous allons prédire, 0 signifie pas de diabète, 1 signifie diabète. Sur ces 768 points de données, 500 sont étiquetés 0 et 268 est 1



- D'abord on utilisant le premier modele 'K-Les voisins les plus proches' pour prédire le diabète :
- L'algorithme k-Nearest Neighbors est sans doute l'algorithme d'apprentissage automatique le plus simple. La construction du modèle consiste uniquement à stocker l'ensemble de données d'apprentissage. Pour faire une prédiction pour un nouveau proches dans l'ensemble de données d'apprentissage - ses "voisins les plus proches".
- Tout d'abord, examinons si nous pouvons confirmer le lien entre la complexité du modèle et la précision :

```

connectKaggle.ipynb
Fichier Modifier Affichage Insérer Exécution Outils Aide Toutes les modifications ont été enregistrées

+ Code + Texte
RAM Disque Modification

# Column Non-Null Count Dtype
---
0 Pregnancies 768 non-null int64
1 Glucose 768 non-null int64
2 BloodPressure 768 non-null int64
3 SkinThickness 768 non-null int64
4 Insulin 768 non-null int64
5 BMI 768 non-null float64
6 DiabetesPedigreeFunction 768 non-null float64
7 Age 768 non-null int64
8 Outcome 768 non-null int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB

[28] from sklearn.model_selection import train_test_split

[29] X_train, X_test, y_train, y_test = train_test_split(diabetes.loc[:, diabetes.columns != 'Outcome'], diabetes['Outcome'], stratify=diabetes['Outcome'], random_state=66)

[30] from sklearn.neighbors import KNeighborsClassifier

[31] training_accuracy = []
test_accuracy = []

0 s terminée à 12:36
    
```

```

connectKaggle.ipynb
Fichier Modifier Affichage Insérer Exécution Outils Aide Toutes les modifications ont été enregistrées

+ Code + Texte
RAM Disque Modification

[31] test_accuracy = []

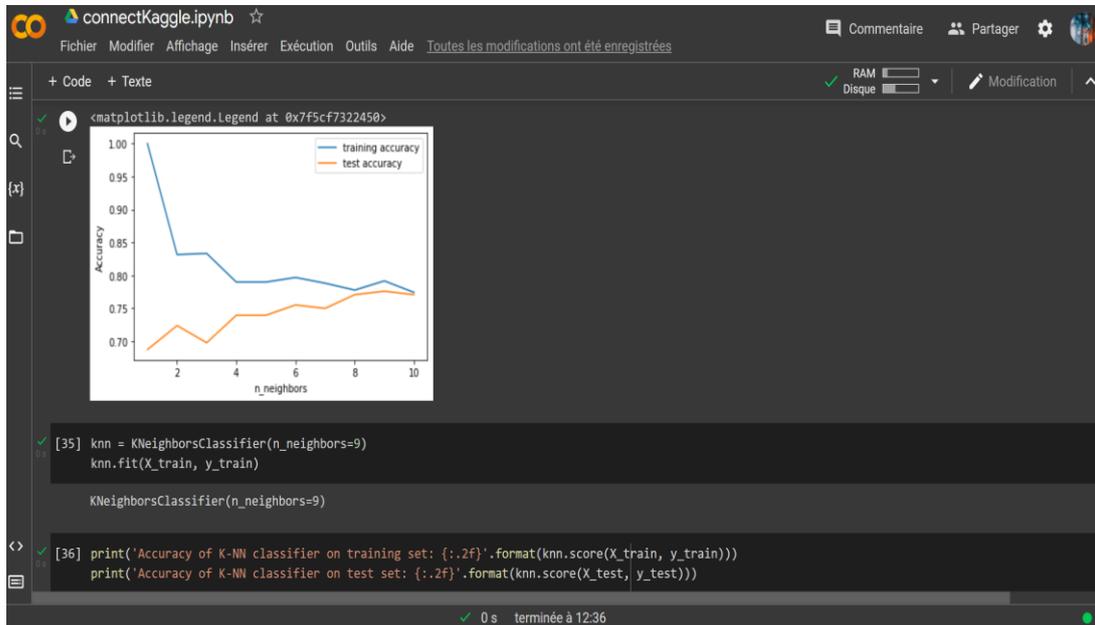
# try n_neighbors from 1 to 10
neighbors_settings = range(1, 11)

[33] for n_neighbors in neighbors_settings:
    knn = KNeighborsClassifier(n_neighbors=n_neighbors)
    knn.fit(X_train, y_train)
    # record training set accuracy
    # eğitim seti doğruluğunu kaydet
    training_accuracy.append(knn.score(X_train, y_train))
    # record test set accuracy
    # kayıt test seti doğruluğu
    test_accuracy.append(knn.score(X_test, y_test))

[34] plt.plot(neighbors_settings, training_accuracy, label="training accuracy")
plt.plot(neighbors_settings, test_accuracy, label="test accuracy")
plt.ylabel("Accuracy")
plt.xlabel("n_neighbors")
plt.legend()

<matplotlib.legend.Legend at 0x7f5cf7322450>
100 — training accuracy
    
```

- Vérifions le score de précision de l'algorithme des k plus proches voisins pour prédire le diabète.



- Ensuite on utilisant le deuxième modèle 'Classificateur d'arbre de décision' ou 'Decision Tree Classifier' pour prédire le diabète

```

[36] Accuracy of K-NN classifier on training set: 0.79
      Accuracy of K-NN classifier on test set: 0.78

[37] from sklearn.tree import DecisionTreeClassifier

[38] tree = DecisionTreeClassifier(random_state=0)
      tree.fit(X_train, y_train)

      DecisionTreeClassifier(random_state=0)

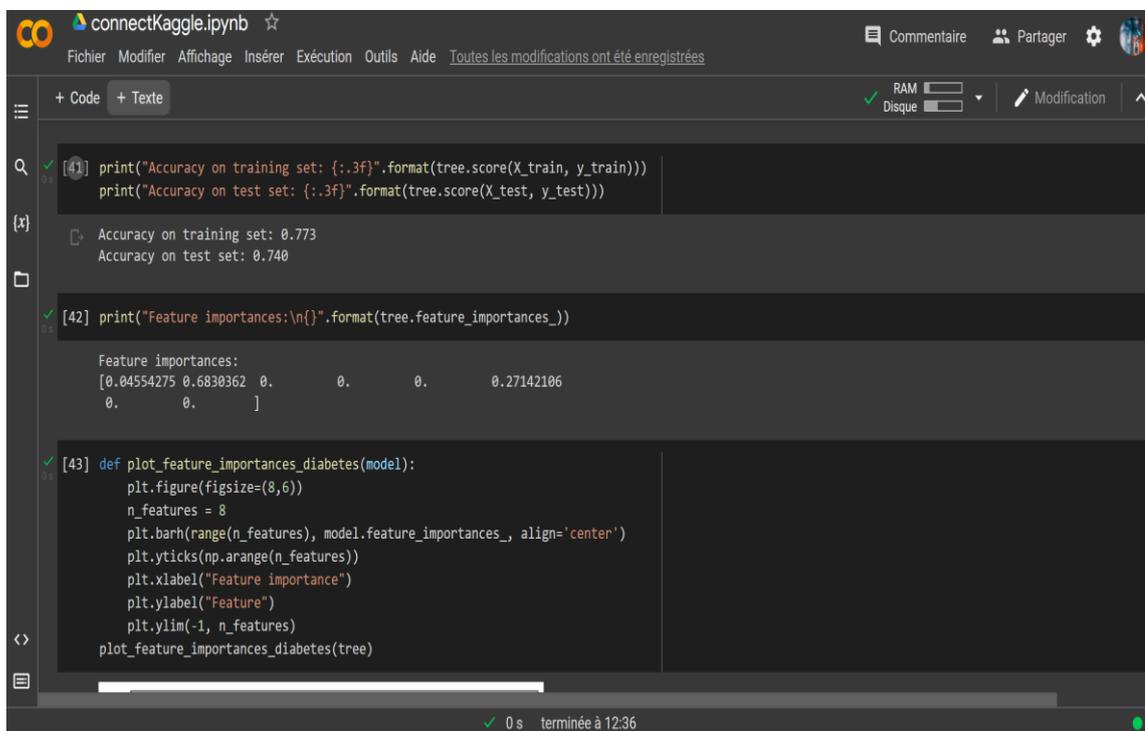
[39] print("Accuracy on training set: {:.3f}".format(tree.score(X_train, y_train)))
      print("Accuracy on test set: {:.3f}".format(tree.score(X_test, y_test)))

      Accuracy on training set: 1.000
      Accuracy on test set: 0.714

[40] tree = DecisionTreeClassifier(max_depth=3, random_state=0)
      tree.fit(X_train, y_train)

      DecisionTreeClassifier(max_depth=3, random_state=0)
    
```

- La précision de l'ensemble d'apprentissage avec Decision Tree Classifier est de 100 %, tandis que la précision de l'ensemble de test est bien pire. Cela indique que l'arbre est sur ajusté et ne se généralise pas bien aux nouvelles données. Par conséquent, nous devons appliquer une pré-taille à l'arbre.
- On le refaire en faisant set `max_depth=3`, limiter la profondeur de l'arbre diminue le surajustement. Cela conduit à une précision moindre sur l'ensemble d'apprentissage, mais à une amélioration sur l'ensemble de test.
- Importance des fonctionnalités dans les arbres de décision 'Feature Importance in Decision Trees' : L'importance des fonctionnalités indique l'importance de chaque fonctionnalité pour la décision prise par un classificateur d'arbre de décision. Il s'agit d'un nombre compris entre 0 et 1 pour chaque caractéristique, où 0 signifie « pas utilisé du tout » et 1 signifie « prédit parfaitement la cible ». L'importance de la fonctionnalité totalise toujours 1 :
- Maintenant, visualisons l'importance des caractéristiques de l'arbre de décision pour prédire le diabète



```
connectKaggle.ipynb
Fichier Modifier Affichage Insérer Exécution Outils Aide Toutes les modifications ont été enregistrées

+ Code + Texte
RAM
Disque
Modification

[41] print("Accuracy on training set: {:.3f}".format(tree.score(X_train, y_train)))
print("Accuracy on test set: {:.3f}".format(tree.score(X_test, y_test)))

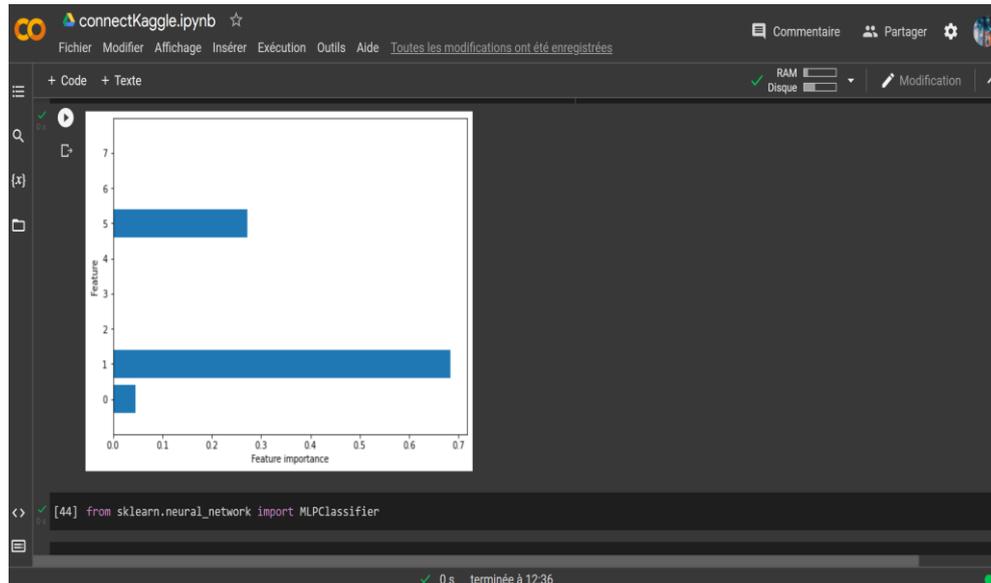
Accuracy on training set: 0.773
Accuracy on test set: 0.740

[42] print("Feature importances:\n{}".format(tree.feature_importances_))

Feature importances:
[0.04554275 0.6830362 0. 0. 0. 0.27142106
 0. 0. ]

[43] def plot_feature_importances_diabetes(model):
    plt.figure(figsize=(8,6))
    n_features = 8
    plt.barh(range(n_features), model.feature_importances_, align='center')
    plt.yticks(np.arange(n_features))
    plt.xlabel("Feature importance")
    plt.ylabel("Feature")
    plt.ylim(-1, n_features)
    plot_feature_importances_diabetes(tree)
```

0 s terminée à 12:36



- La fonction Glucose est donc la plus utilisée pour prédire le diabète.
- Apprentissage en profondeur pour prédire le diabète Deep Learning
- Entraînons un modèle d'apprentissage en profondeur pour prédire le diabète :

La précision des perceptrons multicouches (MLP) n'est pas du tout aussi bon que celle des autres modèles, cela est probablement dû à la mise à l'échelle des données. Les algorithmes d'apprentissage en profondeur s'attendent également à ce que toutes les caractéristiques d'entrée varient de la même manière et, idéalement, à avoir une moyenne de 0 et une variance de 1. Maintenant, je vais redimensionner nos données afin qu'elles remplissent ces conditions pour prédire le diabète avec une bonne précision.

```
connectKaggle.ipynb ☆
Fichier Modifier Affichage Insérer Exécution Outils Aide Toutes les modifications ont été enregistrées
+ Code + Texte
RAM
Disque
Modification

[45] mlp = MLPClassifier(random_state=42)
mlp.fit(X_train, y_train)
MLPClassifier(random_state=42)

[46] print("Accuracy on training set: {:.2f}".format(mlp.score(X_train, y_train)))
print("Accuracy on test set: {:.2f}".format(mlp.score(X_test, y_test)))
Accuracy on training set: 0.73
Accuracy on test set: 0.72

[47] from sklearn.preprocessing import StandardScaler

[48] scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.fit_transform(X_test)

[49] mlp = MLPClassifier(random_state=0)
mlp.fit(X_train_scaled, y_train)
MLPClassifier(random_state=0)
```

- Augmentons maintenant le nombre d'itérations, le paramètre alpha et ajoutons des paramètres plus forts aux poids du modèle :

```

print("Accuracy on training set: {:.3f}".format(
    mlp.score(X_train_scaled, y_train)))
print("Accuracy on test set: {:.3f}".format(mlp.score(X_test_scaled, y_test)))

Accuracy on training set: 0.823
Accuracy on test set: 0.802

[51] mlp = MLPClassifier(max_iter=1000, alpha=1, random_state=0)
mlp.fit(X_train_scaled, y_train)

MLPClassifier(alpha=1, max_iter=1000, random_state=0)

[52] print("Accuracy on training set: {:.3f}".format(
    mlp.score(X_train_scaled, y_train)))
print("Accuracy on test set: {:.3f}".format(mlp.score(X_test_scaled, y_test)))

Accuracy on training set: 0.806
Accuracy on test set: 0.797

[53] plt.figure(figsize=(20, 10))
plt.imshow(mlp.coefs_[0], interpolation='none', cmap='viridis')
plt.yticks(range(0))
    
```

- Le résultat est bon, mais nous ne sommes pas en mesure d'augmenter davantage la précision du test. Par conséquent, notre meilleur modèle à ce jour est le modèle d'apprentissage en profondeur par défaut après mise à l'échelle. Maintenant, je vais tracer une carte thermique des poids de la première couche dans un réseau de neurones appris sur le pour prédire le diabète à l'aide de l'ensemble de données.



2.4- Intelligence artificielle dans la maintenance

2.4-1- L'évolution industrielle et l'industrie 4.0

L'idée de l'industrie 4.0 a été connue pour la première fois en 2011 pour améliorer l'efficacité de l'industrie manufacturière. Avec la complexité croissante, les demandes et la concurrence sur le marché international, les exigences de produits personnalisés sont les principaux défis des entreprises. L'industrie 4.0 peut potentiellement avoir un impact sur les systèmes de fabrication grâce aux progrès rapides de la numérisation et aux technologies telles que l'Internet des objets, les données massives et l'intelligence artificielle y compris l'apprentissage machine. Auparavant, trois transformations modernes ont eu lieu : 1re révolution, caractérisée par la mécanisation, 2e révolution définie par l'électrification, et 3e révolution marquée par l'automatisation et la mondialisation. La figure 11 illustre ces révolutions :

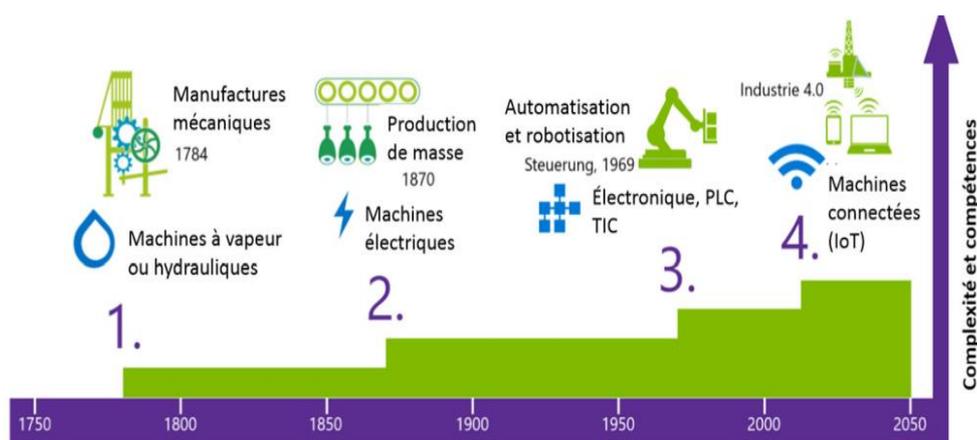


Figure 2.13 : La révolution industrielle tirée de Rioux (2016)

Dans l'industrie, l'évolution vers l'industrie 4.0 peut se faire à plusieurs niveaux, y compris les opérations de maintenance des équipements.

Le coût engendré par les pannes des équipements industriels est faramineux. Concernant le secteur manufacturier, la Société Internationale d'Automatisation (ISA) estime que les industriels perdent chaque année 647 milliards de dollars dans le monde en raison des temps d'arrêt.

Pour relever ce défi, les entreprises adoptent désormais une approche plus proactive dans leur stratégie de maintenance. Plutôt que d'attendre la panne avant une réparation (maintenance réactive) ou le remplacement de pièces selon un calendrier strict (maintenance préventive), ils utilisent des dispositifs et des systèmes intelligents pour prévoir et résoudre les problèmes avant qu'ils ne surviennent (maintenance prédictive).

Au fil des ans, les défauts de la maintenance ponctuelle sont devenus évidents. Une étude de Boeing montre que près de 85% des équipements tombent en panne malgré une maintenance régulière basée sur un calendrier. La maintenance prédictive devient donc une nécessité dans la maintenance d'un outil industriel.

2.4-2- Analyse et validation du modèle

Après avoir obtenu les résultats des modèles algorithmiques, les techniques basées sur les données sont combinées avec des techniques basées sur les connaissances pour prendre de meilleures décisions et stratégies (Sufiyan et al. 2021). Des experts expérimentés dans l'industrie examinent les modèles et les résultats, ce qui conduit à des améliorations des procédures d'exploitation, de maintenance, de surveillance, de test et d'audit pour garantir des actions plus sûres et plus efficaces. Ensuite, une phase de déploiement peut être faite et qui représente la dernière étape du flux de travail pour concevoir un système de classification. Elle représente la mise en exploitation du système.

2.4-3- Les freins à la mise en œuvre d'applications de maintenance à base d'IA

Estimer les pertes financières dues à l'arrêt ou au dysfonctionnement d'une machine repose sur de nombreux paramètres : augmentation des rebuts, prix de la matière première, coût des pièces de rechange, retard de livraison, détérioration de la qualité, allongement des temps d'intervention de maintenance... Les industriels ont besoin d'identifier le ROI attendu sur leurs équipements. C'est pourquoi, donc il faut identifier le périmètre restreint sur lequel il y a un espoir de gain clairement identifié. Ce qui permet de jauger l'efficacité de la solution et son impact sur l'organisation de l'entreprise qui la met en œuvre. Une fois que l'ensemble du processus et des éléments à prendre en considération sont bien maîtrisés, on étend le périmètre de surveillance à d'autres équipements.

Il ne faut pas commencer par ceux qui sont les plus faciles à connecter mais par ceux qui offrent un espoir de gain clairement identifié. Un système de surveillance pour la maintenance prédictive à base d'IA doit être un outil d'aide à la décision. Il faut que le processus de prédiction

soit clairement expliqué à la personne qui sera amenée à prendre des décisions. Les techniques d'IA doivent apporter une information qualifiée complémentaire aux données opérationnelles traditionnelles. Si elles délivrent seulement des informations que tous les techniciens d'un site industriel connaissent déjà, elles ne présentent pas d'intérêt.

2.5- Conclusion

Dans ce chapitre nous avons parlé de manière générale l'intelligence artificielle. Nous avons montré le principe de l'intelligence artificielle et cité quelques définitions afin de pouvoir tirer les grandes lignes de cette nouvelle discipline.

Nous avons aussi parlé de quelques utilisations de cette dernière et de ses domaines d'application. A la fin nous avons récolté les principaux avantages de l'intelligence artificielle, ainsi que les inconvénients et les limites qu'elle présente.

Et aussi nous allons nous intéresser aux outils ou algorithmes indispensables à la conception d'un système intelligent. A la fin, une description de quelques notions sur l'IA dans la maintenance (commençons par montrer l'évolution industrielle et l'apparition de l'industrie 4.0. Ensuite, on met l'accent sur les notions liées à la maintenance prédictive.

CHAPITRE 3

Classification de la maintenance prédictive utilisant Machine Learning

3.1- Introduction

L'apprentissage automatique est une composante importante du domaine de la science des données. Grâce à l'utilisation de méthodes statistiques, plusieurs algorithmes sont formés pour effectuer des classifications ou des prédictions. La maintenance prédictive ou prévisionnelle est le processus qui consiste à utiliser l'analyse des données, l'apprentissage automatique, l'intelligence artificielle et d'autres techniques de gestion des données pour prédire quand des équipements industriels sont susceptibles de tomber en panne.

Avec l'avènement des IIoT, et plus généralement des données au sein des usines, la maintenance prédictive se développe de plus en plus. Les solutions logicielles et leurs algorithmes sont capables d'analyser les données, de repérer des modèles et de prédire le moment où certains matériels sont susceptibles de tomber en panne. Cela est particulièrement utile pour la maintenance régulière des équipements.

En fait, ce chapitre traite un projet de science des données du monde réel avec Python : "Predictive Maintenance Classification" dans lequel un modèle d'apprentissage automatique est proposé, qui classe l'erreur qu'une machine peut commettre en une sorte d'erreur et prédit une défaillance de la machine. La réalisation de cette étude, utilisant plusieurs modèles d'apprentissage automatique (Random Forest, K Nearest Neighbors, Support Vector Machine...) sur une seule classe, ainsi que la mesure de la capacité prédictive des différents algorithmes d'apprentissage automatique qui ont été utilisés pour examiner uniquement les valeurs des scores de précision.

3.2- Étapes de construction du modèle

Les outils utilisés sont le langage Python et la bibliothèque « IPython » qui est une bibliothèque d'apprentissage machine. Elle comporte divers algorithmes de classification, de régression et d'agrégation, et est conçue pour interagir avec les bibliothèques numériques et scientifiques Python comme « NumPy », « SciPy », « matplotlib », « mpl_Finance », « sklearn », « stockstats » ... qui sont aussi utilisés dans le code développé suivant :

The image displays three sequential screenshots of a Jupyter Notebook titled 'maintenanceprédictive.ipynb'. The first screenshot shows the installation of the Kaggle API and the download of a dataset. The second screenshot shows the installation of various Python libraries including yfinance, investpy, stockstats, pandas, numpy, matplotlib, and seaborn. The third screenshot shows the import of specific modules from these libraries, such as mpl_dates, candlestick_ohlc, skew, norm, argrelextrema, get_dummies, KMeans, MiniBatchKMeans, train_test_split, classification_report, confusion_matrix, json, sys, csv, os, and matplotlib inline. It also shows the configuration of pandas display options for max_rows and max_columns.

```
[1] # Installation package kaggle
! pip install -q kaggle

[2] # Creation un dossier kaggle
! mkdir ~/.kaggle/

[3] # Copier kaggle.json dans le dossier kaggle
! cp '/content/drive/MyDrive/kaggle/kaggle.json' ~/.kaggle/

[4] # Donner les droits a kaggle.json
! chmod 600 ~/.kaggle/kaggle.json

[6] ! kaggle datasets list

[7] ! kaggle datasets download -d shivamb/machine-predictive-maintenance-classification

Downloading machine-predictive-maintenance-classification.zip to /content
 0% 0.00/137K [00:00<, 7B/s]
100% 137K/137K [00:00<00:00, 57.1MB/s]
```

```
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:100% !important; }</style>"))

[9] import warnings
warnings.filterwarnings("ignore")

! pip install yfinance
! pip install investpy
! pip install stockstats
! pip install mpl_finance
import pandas as pd
import numpy as np
import datetime as dt
import datetime, pytz
import seaborn as sns
import plotly.graph_objects as go
import matplotlib.pyplot as plt
from matplotlib import pyplot
import yfinance
import investpy
from stockstats import StockDataFrame

import matplotlib.dates as mpl_dates
from mpl_finance import candlestick_ohlc
from scipy.stats import skew, norm
from scipy.signal import argrelextrema
from pandas import get_dummies
import copy
from sklearn.cluster import KMeans
import time
from sklearn.cluster import MiniBatchKMeans
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
import json
import sys
import csv
import os
%matplotlib inline
pd.set_option("display.max_rows",2000)
pd.set_option("display.max_columns",50)
```

Le développement d'un modèle d'apprentissage machine repose sur plusieurs étapes s'appelle Exploratory Data Analysis (EDA) : 'Analyse Exploratoire des Données'. Ainsi, dans leur développement ils ont suivi les étapes suivantes :

Data Collection : ‘La collecte de données’ en combinant toutes les données en une seule avec les mêmes attributs : Les jeux de données réels de maintenance prédictive étant généralement difficiles à obtenir et en particulier difficiles à publier, les données fournies par le référentiel UCI sont un jeu de données synthétique qui reflète au mieux de leur connaissance la maintenance prédictive réelle rencontrée dans l'industrie. Pour cela, un code Python a été développé lit les fichiers (*.csv) et les combine en une seule table de données :

	UDI	Product ID	Type	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Target	Failure Type
0	1	M14860	M	298.1	308.6	1551	42.8	0	0	No Failure
1	2	L47181	L	298.2	308.7	1408	46.3	3	0	No Failure
2	3	L47182	L	298.1	308.5	1498	49.4	5	0	No Failure
3	4	L47183	L	298.2	308.6	1433	39.5	7	0	No Failure
4	5	L47184	L	298.2	308.7	1408	40.0	9	0	No Failure
...
9995	9996	M24855	M	298.8	308.4	1604	29.5	14	0	No Failure
9996	9997	H39410	H	298.9	308.4	1632	31.8	17	0	No Failure
9997	9998	M24857	M	299.0	308.6	1645	33.4	22	0	No Failure
9998	9999	H39412	H	299.0	308.7	1408	48.5	25	0	No Failure
9999	10000	M24859	M	299.0	308.7	1500	40.2	30	0	No Failure

✚ L'ensemble de données se compose de 10 000 points de données stockés sous forme de lignes avec 14 entités en colonnes :

- UDI ‘UID’ : identifiant unique allant de 1 à 10000 ;
- ProductID : composé d'une lettre L, M ou H pour faible (50 % de tous les produits), moyen (30 %) et élevé (20 %) comme variantes de qualité du produit et un numéro de série spécifique à la variante
- Air Temperature ‘Température de l'air’ [K] : générée à l'aide d'un processus de marche aléatoire normalisé ultérieurement à un écart type de 2 K autour de 300 K ;
- Process Temperature ‘Température de procédé’ [K] : générée à l'aide d'un processus de marche aléatoire normalisé à un écart type de 1 K, ajouté à la température de l'air plus 10 K ;
- Rotational Speed ‘Vitesse de rotation’ [tr/min] : calculée à partir d'une puissance de 2860 W, recouverte d'un bruit normalement distribué ;
- Torque ‘Couple’ [Nm] : les valeurs de couple sont normalement réparties autour de 40 Nm avec un écart type de 10 Nm et aucune valeur négative ;
- Tool Wear ‘Usure d'outil’ [min] : Les variantes de qualité H/M/L ajoutent 5/3/2 minutes d'usure d'outil à l'outil utilisé dans le processus ;

Un autre point important concernant l'ensemble de données est qu'il existe 2 variables cibles :

- Target ‘Cible’ : Échec ou pas.
- Failure Type ‘Type de panne’ : Type de panne

Étant donné que les prendre comme n'importe quelle variable dans le modèle de prédiction peut être un problème dans la performance du modèle, nous utilisons les deux comme variables dépendantes.

- ❖ Chaque ligne est une observation (également appelée : échantillon, exemple, instance, enregistrement)

- ❖ Chaque colonne est une caractéristique (également connue sous le nom : prédicteur, attribut, variable indépendante, entrée, régresseur, covariable)

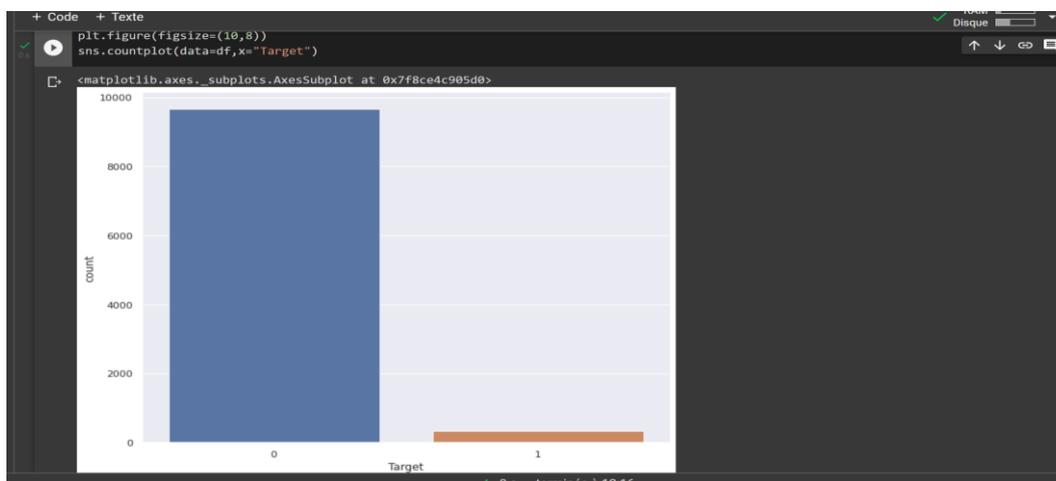
Data Cleaning ‘Nettoyage des données ’ :Lorsqu’il s’agit de données du monde réel, les données sales sont la norme plutôt que l’exception. Nous devons continuellement prédire les valeurs correctes, imputer les valeurs manquantes et trouver des liens entre divers artefacts de données tels que des schémas et des enregistrements. Nous devons cesser de traiter le nettoyage des données comme un exercice au coup par coup (résoudre différents types d’erreurs de manière isolée) et exploiter à la place tous les signaux et toutes les ressources (telles que les contraintes, les statistiques disponibles et les dictionnaires) pour prévoir avec précision les actions correctives.

Car dans ce projet ils créaient une nouvelle variable, une nouvelle colonne, à utiliser dans leur modèle. Lors de la création de cette colonne, ils utilisent les variables "Total Wear" et "Torque". En même temps, il supprime les variables ID et ProductID dans l’ensemble de données de l’ensemble de données :

```
[14] df = df.drop(["UDI", 'Product ID'],axis=1)
[15] df['nf'] = df['Tool wear [min]'] * df['Torque [Nm]']
```

Visualization ‘Visualisation’ :Avec la visualisation interactive, vous pouvez pousser le concept un peu plus loin en utilisant la technologie pour explorer les tableaux et les graphiques pour plus de détails, en modifiant de manière interactive les données que vous voyez et la façon dont elles sont traitées.

- 🌈 visualisation l’état de la variable cible en fonction des variables numériques avec différents graphiques :
- ✓ Dans cette section, nous faisons une visualisation en fonction de la variable cible ‘Target’ dans l’ensemble de données utilisant ‘countplot’. comme se suit :



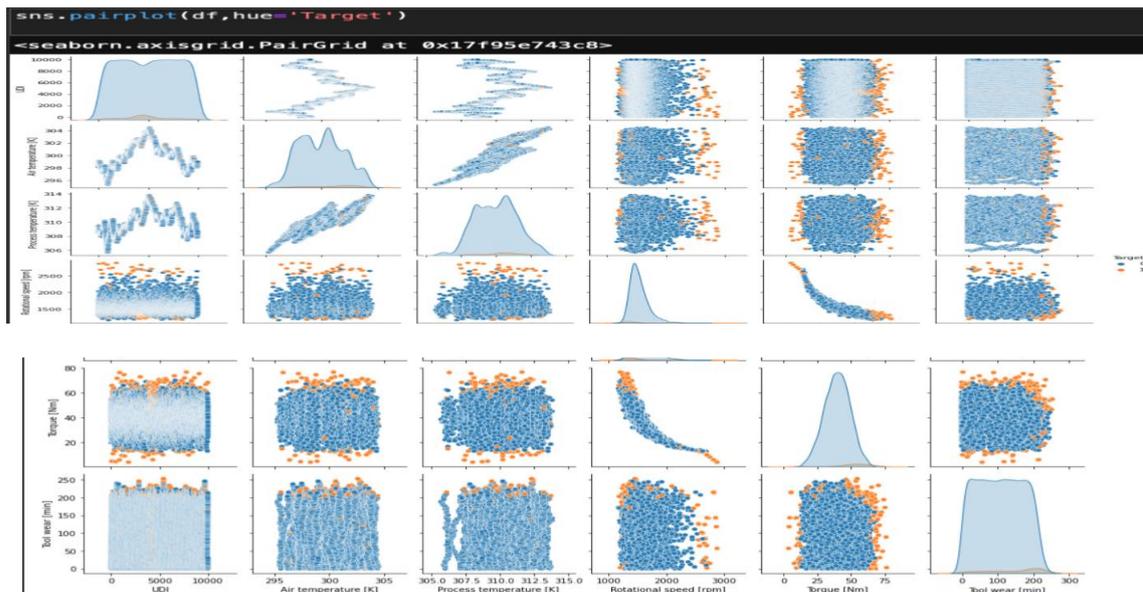
Target : Echec ou pas Echec=1, Pas d’échec=0

- ✓ Maintenant en utilisant la variable cible 'Target', on obtient le nombre d'erreurs selon les types d'erreur utilisant aussi 'countplot' :

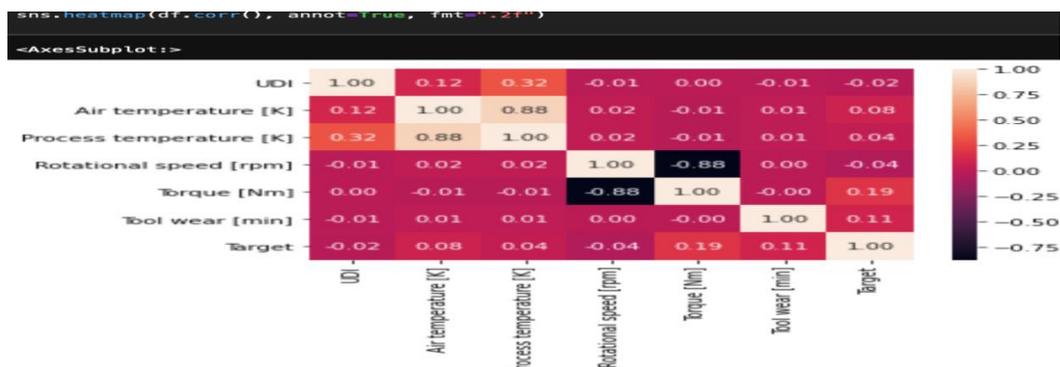


- Heat Dissipation est le maximum

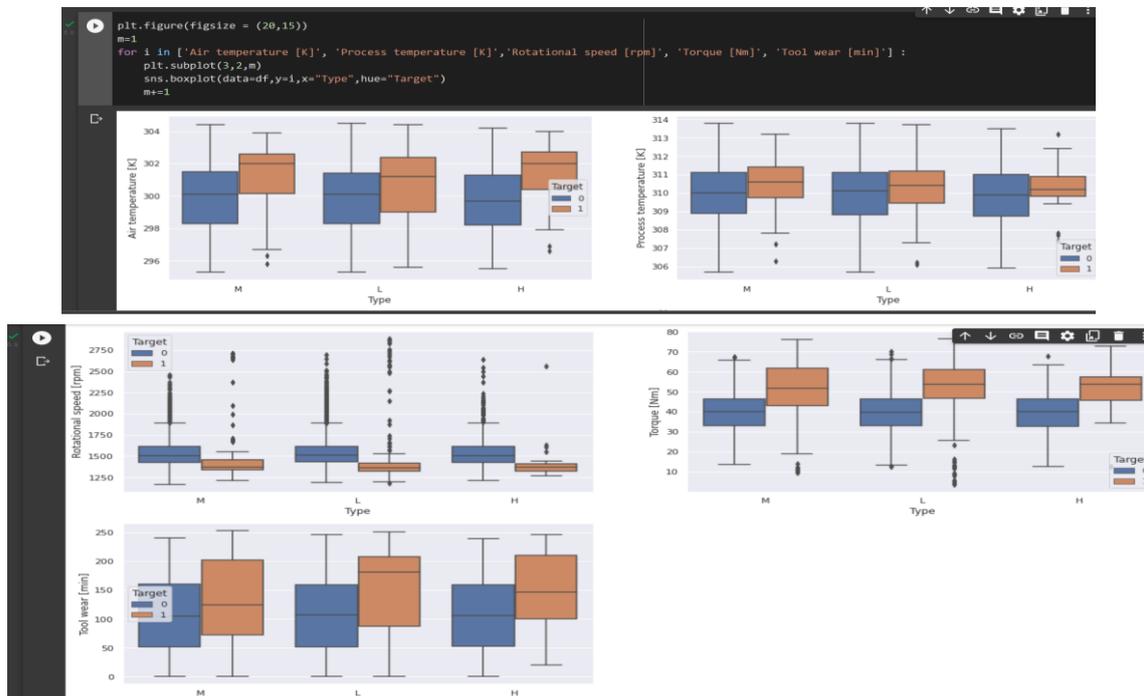
- ✓ Examinent maintenant les distributions des autres variables en fonction de la variable cible à l'aide de pairplot :



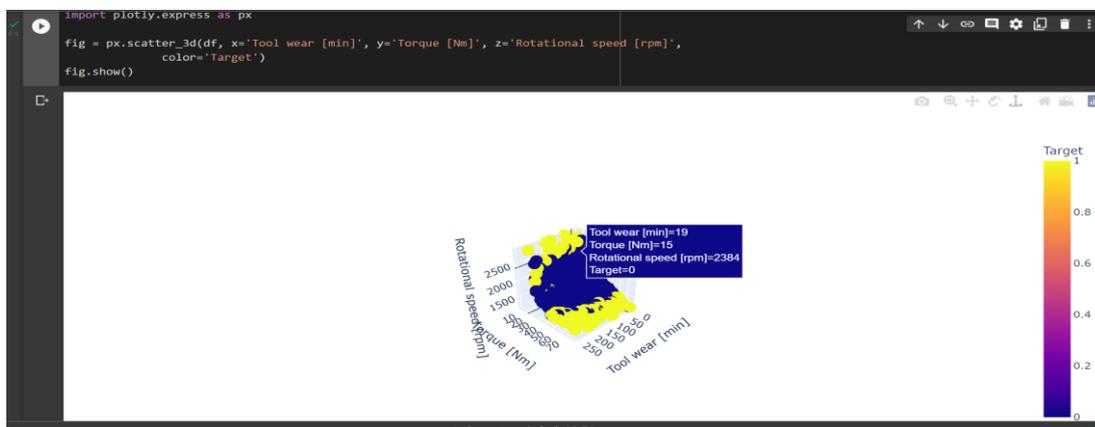
- ✓ Ils utilisent la carte thermique pour voir le niveau de corrélation et les directions des variables avec la variable cible 'Target' et entre elles :



- ✓ Observons maintenant la variabilité des autres variables numériques en fonction de la variable cible avec boxplot :



- ✓ Ils effectuent le processus de visualisation sur l'axe commun en fonction des variables « Torque », « Rotational Power », « Tool Wear ». Pour cela, le graphique 3D ci-dessous ‘Distributions variables cibles’ :



Data Preprocessing ‘Le prétraitement des données’ est effectué pour améliorer les données avant de créer un modèle d’apprentissage automatique. En effet, la plupart du temps, les données ont des ordres de grandeur différents et parfois même des valeurs manquantes. Ces problèmes peuvent impacter négativement les performances du modèle. Pour pallier cela, on a effectué quelques traitements préparatoires sur les données tels que la suppression des valeurs nulles et la mise à l’échelle qui comprend la standardisation et la normalisation. Aussi l’encodage d’ensemble de données pour les algorithmes d’apprentissage automatique, ils ont changé la colonne avec des éléments textuels en une variable numérique. Ils ont changé la colonne avec des éléments textuels en une variable numérique :

```
[23] df["Type"] = df["Type"].replace("L",0)
      df["Type"] = df["Type"].replace("M",1)
      df["Type"] = df["Type"].replace("H",2)
```

Le prétraitement inclut également la division des données en données d'apprentissage (80%) et en données de test (20%). Les données d'entraînement sont utilisées pour entraîner les modèles. Ces derniers seront testés par la suite à l'aide des données de test pour déterminer l'exactitude de la prédiction.

```
df_target = df.Target.copy()
df_faillure_type = df["Failure Type"].copy()
X_train, X_test, y_train, y_test = train_test_split(df.drop(['Failure Type', 'Target'],axis=1),
                                                  df['Target'], test_size=0.3, random_state=42)
```

Ils importent également les bibliothèques pour les algorithmes d'apprentissage automatique qu'ils utilisent dans leur étude :

```
import time

from sklearn.metrics import accuracy_score, classification_report
classifier=[]
imported_as=[]

#LGBM
import lightgbm as lgb
lgbm = lgb.LGBMClassifier()
classifier.append('LightGBM')
imported_as.append('lgbm')

#MultilayerPerceptron
from sklearn.neural_network import MLPClassifier
mlp=MLPClassifier()
classifier.append('Multi Layer Perceptron')
imported_as.append('mlp')

#BAGGING
from sklearn.ensemble import BaggingClassifier
bc = BaggingClassifier()
classifier.append('Bagging')
imported_as.append('bc')

#GBC
from sklearn.ensemble import GradientBoostingClassifier
gbc = GradientBoostingClassifier()
classifier.append('Gradient Boosting')
```

```
#ADA
from sklearn.ensemble import AdaBoostClassifier
ada = AdaBoostClassifier()
classifier.append('Ada Boost')
imported_as.append('ada')

#XGB
import xgboost as xgb
from xgboost import XGBClassifier
xgb = XGBClassifier()
classifier.append('XG Boost')
imported_as.append('xgb')

# Logistic Regression
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
classifier.append('Logistic Regression')
imported_as.append('lr')

#RFC
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier()
classifier.append('Random Forest')
imported_as.append('rfc')

#KNN
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=1)
classifier.append('k Nearest Neighbours')
imported_as.append('knn')
```

```
#SVM
from sklearn.svm import SVC
svc = SVC()
classifier.append('Support Vector Machine')
imported_as.append('svc')

#Grid
from sklearn.model_selection import GridSearchCV
param_grid = {'C': [0.1, 1, 10, 100, 1000, 2000], 'gamma': [1, 0.1, 0.01, 0.001, 0.0001], 'kernel': ['rbf']}
grid = GridSearchCV(SVC(), param_grid, refit=True, verbose=3)
classifier.append('SVM tuning grid')
imported_as.append('grid')

#STacking
from sklearn.ensemble import StackingClassifier
estimators=[('rf', RandomForestClassifier(n_estimators=10, random_state=42)),
            ('svc', SVC(random_state=42))]
stc = StackingClassifier(estimators=estimators, final_estimator=LogisticRegression())
classifier.append('Stacked (RFR & SVM)')
imported_as.append('stc')

classifiers = pd.DataFrame({'Classifier':classifier, 'Imported as':imported_as})
print('All Models Imported\nModels stored in dataframe called classifiers')
```

All Models Imported
Models stored in dataframe called classifiers

Ils créent une classe et répertorient le score de précision de chaque algorithme d'apprentissage automatique. Également ajouter que cette méthode, dans laquelle ils exécutent les modèles que ils vont utiliser de manière compacte en créant une seule classe, est une méthode qui peut rendre le travail d'un data scientist beaucoup plus facile, et ils la recommandent particulièrement pour être inclus dans les modèles utilisés dans un flux de projet, autrement dit pour afficher les synonymes des algorithmes ML.

```

class Modelling:
    def __init__(self, X_train, Y_train, X_test, Y_test, models):
        self.X_train = X_train
        self.X_test = X_test
        self.Y_train = Y_train
        self.Y_test = Y_test
        self.models = models

    def fit(self):
        model_acc = []
        model_time = []
        for i in self.models:
            start=time.time()
            if i == 'knn':
                accuracy = []
                for j in range(1,200):
                    kn = KNeighborsClassifier(n_neighbors=j)
                    kn.fit(self.X_train,self.Y_train)
                    predK = kn.predict(self.X_test)
                    accuracy.append([accuracy_score(self.Y_test,predK),j])
                temp = accuracy[0]
                for m in accuracy:
                    if temp[0] < m[0]:
                        temp=m
                        i = KNeighborsClassifier(n_neighbors=temp[1])
                        i.fit(self.X_train,self.Y_train)
            model_acc.append(accuracy_score(self.Y_test,i.predict(self.X_test)))
            stop=time.time()
            model_time.append((stop-start))
            print(i,'has been fit')
    
```

```

self.models_output = pd.DataFrame({'Models':self.models,'Accuracy':model_acc,'Runtime (s)':model_time})

def results(self):
    models=self.models_output
    models = models.sort_values(by=['Accuracy','Runtime (s)'],ascending=[False,True]).reset_index().drop('index',axis=1)
    self.best = models['Models'][0]
    models['Models']=models['Models'].astype(str).str.split("(", n = 2, expand = True)[0]
    models['Accuracy']=models['Accuracy'].round(5)*100
    self.models_output_cleaned=models
    return(models)

def best_model(self,type):
    if type=='model':
        return(self.best)
    elif type=='name':
        return(self.models_output_cleaned['Models'][0])
    def best_model_accuracy(self):
        return(self.models_output_cleaned['Accuracy'][0])
    def best_model_runtime(self):
        return(round(self.models_output_cleaned['Runtime (s)'][0],3))
    def best_model_predict(self,X_test):
        return(self.best.predict(X_test))
    def best_model_matrix(self):
        return(classification_report(self.Y_test,self.best.predict(self.X_test)))
    
```

```
display(classifiers)
```

	Classifier	Imported as
0	LightGBM	lgbm
1	Multi Layer Perceptron	mlp
2	Bagging	bc
3	Gradient Boosting	gbc
4	Ada Boost	ada
5	XG Boost	xgb
6	Logistic Regression	lr
7	Random Forest	rfc
8	k Nearest Neighbours	knn
9	Support Vector Machine	svc
10	SVM tuning grid	grid
11	Stacked (RFR & SVM)	stc

```
[28] models_to_test = [bc,gbc,ada,rfc,mlp,lr,knn,svc]
```

Ils s'adaptent les algorithmes d'apprentissage automatique à utiliser et affichent les scores de précision dans une liste :

```

X_train.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7000 entries, 9069 to 7270
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0    UDI                    7000 non-null   int64
1    Product ID            7000 non-null   object
2    Type                  7000 non-null   int64
3    Air temperature [K]   7000 non-null   float64
4    Process temperature [K] 7000 non-null   float64
5    Rotational speed [rpm] 7000 non-null   int64
6    Torque [Nm]           7000 non-null   float64
7    Tool wear [min]       7000 non-null   int64
dtypes: float64(3), int64(4), object(1)
memory usage: 492.2+ KB

classification = Modelling(X_train,y_train,X_test,y_test,models_to_test)
classification.fit()
BaggingClassifier() has been fit
GradientBoostingClassifier() has been fit
AdaBoostClassifier() has been fit
RandomForestClassifier() has been fit
MLPClassifier() has been fit
LogisticRegression() has been fit
KNeighborsClassifier(n_neighbors=1) has been fit
    
```

```

RandomForestClassifier(n_estimators=10,
                       random_state=42)),
('svr', SVC(random_state=42))],
final_estimator=LogisticRegression()) has been fit
classification.results()

```

	Models	Accuracy	Runtime (s)
0	BaggingClassifier	99.133	2.483880
1	RandomForestClassifier	98.833	0.810921
2	GradientBoostingClassifier	98.767	0.986392
3	StackingClassifier	98.633	1.347955
4	AdaBoostClassifier	97.667	0.320145
5	LogisticRegression	97.400	0.164693
6	MLPClassifier	96.633	1.077893
7	KNeighborsClassifier	96.067	0.147125

Nous pouvons voir les scores de succès de divers algorithmes d'apprentissage automatique dans la liste.

Le meilleur modèle est : BaggingClassifier 'BC'

Précision du modèle : 99.133%

Durée d'exécution en secondes : 2.483880 s

3.3- Conclusion

Selon les analyses effectuées et les résultats obtenus, il est possible de faire quelques considérations concluantes liées à ce projet.

Avant de développer les modèles, nous avons effectué un prétraitement des données pour nous assurer de la validité des hypothèses d'applicabilité des modèles et garantir les meilleures performances. En bref, dans la phase de prétraitement, nous avons supprimé certains échantillons ambigus, nous avons appliqué un encodage d'étiquette aux colonnes catégorielles, puis nous avons effectué la mise à l'échelle des colonnes avec StandardScaler.

À la fin, nous pouvons conclure que, les modèles choisis fonctionnent très bien. Le meilleur modèle est BaggingClassifier 'BC' et le pire est RandomForestClassifier 'RCF' ; cependant, le temps de réponse de RandomForestClassifier 'RCF' est instantané tandis que BaggingClassifier 'BC' prend plus de temps. Le choix du modèle dépend des besoins de l'entreprise : pour une application plus rapide, on peut utiliser RandomForestClassifier 'RCF' tandis que si l'on se soucie davantage de la précision, on peut utiliser BaggingClassifier 'BC'.

Conclusion générale

Le travail présenté dans ce mémoire nous a permis d'approfondir nos connaissances au domaine d'intelligence artificiel en générale et le domaine de la maintenance prédictive utilisant les techniques d'intelligence artificiel spécifiquement.

Pour bien cerner notre thème, nous avons choisi de répartir notre travail en trois chapitres. Dans le premier, nous présentons globalement ce qu'est la maintenance prédictive. Le second chapitre du travail est consacré à une présentation de l'apprentissage automatique, enfin nous décrivons dans le troisième chapitre un modèle proposé pour classer l'erreur qu'une machine puisse commettre en une sorte d'erreur et prédit une défaillance de la machine, consacré à l'implémentation et à l'évaluation de modèle proposé. Nous tentons de mettre en relief les points suivants : la présentation des différentes bibliothèques Python que nous avons utilisé, les prétraitements que nous avons effectués sur la collection de données...

Bibliographies

- [1] Z. Simeu-Abazi, M. DI Mascolo, D.M. Pham, maintenance pré-conditionnelle, laboratoire d'Automatique de Grenoble.
- [2] H.Procaccia, R.Cordier, Application of Bayesian statistical decision theory for a maintenance optimisation problem, *Reliability Engineering and System Safety* . 55(1997)143-149.
- [3] R. Sarker, A. Haque, Optimisation of maintenance and spare provisioning policy using simulation, *Applied Mathematical Modelling* 24(2000) 751-760.
- [4] (Source : EPRI Electric Power Research Institute).
- [5]. Goguelin, P, La prévention des risques professionnels, Que sais-je ? PUF, Paris, 1996.
- [6]. E Garcia, H Guyennet, J.C Lapayre, N Zerhouni, A new industrial cooperative télémaintenance platform. *Computers & Industrial Engineering*, Volume 46, Issue 4, July 2004, Pages 851-864.
- [7] ANPE, Entretien, maintenance, coll. ROME, Description et évolution des métiers, Paris, La Documentation française,
- [8] (Extrait norme NF EN 13306X60-319).
- [9]. Lung B, Morel G, Leger J.B, Proactive maintenance stratégie for Harbour crane opération improvement, *Robotica*, vol 21, 2003, pp 313-324.
- [10] NASA., Reliability Centered Maintenance Guide for Facilites and Collatéral Equipment, February, 2000.
- [11] Despujols A., Optimisation de la maintenance par la fiabilité (OMF), MT9310, *Techniques de l'ingénieur*, Octobre, 2004.
- [12] Zwingelstein G., La Maintenance basée sur la fiabilité - Guide pratique d'application de la RCM, Editions HERMES, 1996.
- [13] Canizo, M., Onieva, E., Conde, A., Charramendieta, S., & Trujillo, S. (2017). Real-time predictive maintenance for wind turbines using Big Data frameworks. 2017 IEEE International Conference on Prognostics and Health Management (ICPHM). doi: 10.1109/ACCESS.2020.2968615
- [14] Borad, A. (2019). addressing challenges associated with imbalanced datasets in machine learning. Repéré
- [15] Cambridge University Press. (2021). Prédicatif. Dans *Dictionnaire Cambridge*
- [16] <https://www.kaggle.com/code/serkanp/predict-diabetes-with-ml?cellIds=4&kernelSessionId=106464730>
- [17] <https://www.kaggle.com/datasets/nancyalaswad90/review>
- [18] <https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification>