



Apprentissage machine pour la prédiction de l'effet de la réadaptation physique sur le genou pathologique

Mémoire présenté comme exigence de la maîtrise en informatique

Par Patrick Bellerive

Janvier 2023



<http://r-libre.teluq.ca/2853>

REMERCIEMENTS

J'en suis maintenant rendu à écrire l'un des documents les plus importants de ma vie. L'effort soutenu, la volonté de réussir et la passion n'auraient pas été suffisants sans le support de plusieurs personnes. Il est important pour moi de souligner et d'énoncer l'ampleur des sacrifices que touche ce travail.

Premièrement, l'accompagnement de ma directrice de recherche, Neila MEZGHANI. Bien que l'université à distance ne soit pas une nouveauté, le contexte actuel rend même le stage de recherche virtuel. Cet accompagnement constant, sans réserve, a repoussé mes limites. Un sincère remerciement pour ton accompagnement et surtout pour m'avoir permis de voir les données autrement !

Je tiens aussi à remercier Youssef OUAKRIM pour son aide, parfois sur demande. Toujours de bonne humeur et avec des pistes de solutions chaque fois que je me sentais perdu dans ces centaines de lignes de programmation et de chiffres. Nos séances de remue-méninges me manquent déjà.

Je tiens à remercier mes patrons pour la latitude qu'ils ont eue, pour me permettre d'accomplir ce processus.

Évidemment, les encouragements, la patience et l'accompagnement de ma famille. Sans cela, cette maîtrise aurait été un échec. La très grande patience de ma femme et mes enfants pour ces journées où je me suis isolé pour être concentré afin de me dépasser et d'approfondir des connaissances si précieuses.

RÉSUMÉ

L'analyse des signaux biomécaniques comporte une complexité élevée ainsi qu'une grande dimensionnalité à cause des divers facteurs et contraintes que le corps humain génère sur le mouvement. Que ce soit la complexité des articulations, les pathologies ou la condition physique, cela influence les données de chaque individu. La technologie permettant d'obtenir les signaux est présente, mais il faut maintenant mettre à profit l'information récoltée. L'utilisation des arbres de décision est rapide et performante. De plus, ils sont compréhensibles. Cependant, les arbres de décision sont davantage linéaires.

Ainsi, l'analyse de données biomécaniques exige une approche adaptée à leur grande variabilité et leur grande dimensionnalité afin de faire ressortir les attributs significatifs pour effectuer des prédictions précises.

Dans le cadre de cette recherche, l'objectif est la mise en place d'algorithmes qui permettent de mieux cibler les attributs de valeur importants à la classification et la généralisation des algorithmes pour prédire l'effet de la réadaptation sur le genou pathologique, en l'occurrence atteint d'arthrose. Afin d'y parvenir, il y a l'exploration d'algorithmes alternatifs en premier lieu. Par la suite, nous avons l'exploration de techniques de réduction de la dimensionnalité, soit l'extraction des caractéristiques significatives.

Travail connexe

L'apprentissage d'un modèle en intelligence artificielle utilise les attributs d'un individu. Depuis des années, les théories et les algorithmes pour la sélection des caractéristiques importantes se multiplient. Dans sa plus simple expression, l'arbre de décision construit des arbres en sélectionnant les caractéristiques qui représentent la plus grande quantité d'information. Les premiers partages sont généralement plus importants et ne dépendent pas du jeu de données [26].

L'objectif est de réduire les caractéristiques nécessaires, tout en choisissant les plus pertinentes. Dans la littérature, deux concepts bien distincts s'y retrouvent ; soit la sélection de caractéristiques et l'extraction de caractéristiques. Il existe différentes méthodes pour déterminer les caractéristiques qui représentent la meilleure information comme : l'entropie et le pointage de fisher (Gu, Li, and Han 2012).

Dans un autre ordre d'idées, l'extraction des caractéristiques se tourne vers l'utilisation d'auto-encodeurs (Hinton and Salakhutdinov 2006) pour l'extraction des attributs les plus significatifs. Cette méthode non supervisée permet l'extraction des caractéristiques ainsi que la réduction de la dimension. À travers cette thématique de recherche, deux approches distinctes seront mises en place pour déterminer les caractéristiques importantes afin d'effectuer l'entraînement de nos modèles.

TABLE DES MATIÈRES

REMERCIEMENTS.....	III
RÉSUMÉ.....	V
TRAVAIL CONNEXE.....	VI
TABLE DES FIGURES	X
TABLE DES TABLEAUX.....	XII
1. INTRODUCTION	1
1.1. CONTEXTE	1
1.2. OBJECTIF ET METHODOLOGIE.....	1
1.3. PLAN	2
2. LE GENOU ET SA BIOMÉCANIQUE.....	4
2.1. LES ARTICULATIONS	4
2.1.1. ARTICULATION FIBREUSE.....	4
2.1.2. ARTICULATION CARTILAGINEUSE.....	5
2.1.3. ARTICULATION DU GENOU.....	5
2.2. L'ARTICULATION DU GENOU ET SA BIOMECHANIQUE.....	6
2.3. LE FEMUR.....	7
2.4. LA PATELLA	8
2.5. LE TIBIA.....	9
2.6. PATHOLOGIES DU GENOU – ARTHROSE.....	10
3. LE CYCLE DE MARCHE	12
3.1. LE CYCLE DE MARCHE	12
4. APPRENTISSAGE MACHINE.....	15

4.1.	TRAITEMENT DES DONNEES	15
4.1.1.	<i>Nettoyage</i>	15
	• <i>Bruit dans les données</i>	15
	• <i>Valeur manquante</i>	16
	• <i>Valeur aberrante</i>	17
4.1.2.	<i>Réduction des dimensions</i>	17
4.2.	L'APPRENTISSAGE.....	18
4.2.1.	<i>Apprentissage supervisé</i>	19
4.2.2.	<i>Apprentissage non supervisé</i>	20
4.3.	ARBRE DE DECISION	20
4.3.1.	<i>Entropie</i>	22
4.3.1.1.	<i>Indice de Gini</i>	23
4.3.2.	<i>Algorithme</i>	24
4.4.	XGBOOST	24
4.4.1.	<i>Amplification du gradient</i>	25
4.4.2.	<i>XGBOOST, l'algorithme</i>	26
4.5.	RESEAU DE NEURONES.....	28
4.5.1.	<i>Objectif du réseau de neurones</i>	28
4.5.2.	<i>Fonction d'activation</i>	32
4.5.3.	<i>Descente du gradient</i>	36
4.5.4.	<i>Réseau de neurones à propagation avant</i>	38
4.5.5.	<i>Rétropropagation</i>	38
4.6.	AUTO-ENCODEUR.....	39
4.6.1.	<i>Structure d'un auto-encodeur</i>	39
5.	METHODOLOGIE ET RESULTATS EXPERIMENTAUX.....	41
5.1.	METHODOLOGIE	41
	<i>Processus global</i>	42
5.2.	BASE DE DONNEES	43
5.3.	APPRENTISSAGE MACHINE TRADITIONNEL	45
5.3.1.	<i>Analyse des résultats - système de classification – XGBOOST</i>	45

5.4.	APPRENTISSAGE PROFOND.....	54
5.4.1.	<i>Analyse des résultats – Auto-encodeur unique</i>	55
5.4.2.	<i>Analyse des résultats – Auto-encodeur par classe</i>	56
6.	CONCLUSION	59
	BIBLIOGRAPHIE	61

TABLE DES FIGURES

Figure 1 L'articulation du genou [20].....	6
Figure 2 – Mouvement de l’articulation du genou [21].....	7
Figure 3 - Le fémur [29]	8
Figure 4 - La patella (rotule) [30]	9
Figure 5 - Le tibia [30]	10
Figure 6 - Cycle de marche - Phase d'appui [31].....	13
Figure 7 - Cycle de marche - Phase d'oscillation [31].....	13
Figure 8 - Cycle de marche complet	14
Figure 9 - Autoencodeur avec images et réduction de la dimension (https://stackabuse.com/autoencoders-for-image-reconstruction-in-python-and-keras/)	18
Figure 10 - Arbre de décision.....	21
Figure 11 - Formule de l'entropie	22
Figure 12 - Courbe des valeurs de l'entropie [28].....	23
Figure 13 - Représentation visuelle du gradient	25
Figure 14 - XOR Classification non linéaire, orange correspond à VRAI et bleu à FAUX[2]	29
Figure 15 – Perceptron	30
Figure 16 - Xor - classifications avec le perceptron [2].....	31
Figure 17 - Fonction sigmoïde[32].....	33
Figure 18 - Fonction Tanh [33]	34
Figure 19 - Fonction d'activation RELU [34].....	35
Figure 20 - Fonction d'activation LRELU[35]	36
Figure 21 - descente du gradient [2]	37

Figure 22 - Architecture d'un auto-encodeur [22].....	40
Figure 23 - - Les courbes cinématiques des 309 sujets de la base de données.....	44
Figure 24 - Processus d'entrainement de XGBOOST.....	46
Figure 25 - Importance des attributs pour le groupe homme	49
Figure 26 - Importance des attributs pour le groupe femme	50
Figure 27 - Importance des caractéristiques pour le jeu de données homme-femme.....	51
Figure 28 - Processus de construction des données pour l'entrainement	57

TABLE DES TABLEAUX

Tableau 1 - Résultats de la classification avec Xgboost	47
Tableau 2 - Abréviations et détails [17]	48
Tableau 3 - Xgboost - Attributs les plus important selon le sexe...48	
Tableau 4 - Résultats – auto-encodeur unique	55
Tableau 5 - Résultat – auto-encodeur unique avec données à T0	56
Tableau 6 - Résultat – auto-encodeur double avec les 69 attributs	57
Tableau 7 - Résultat – auto-encodeur double avec données à T0	.57

1. INTRODUCTION

1.1.Contexte

Contrairement à l'imagerie rayons X, l'acquisition des signaux biomécanique liés au genou et sa mobilité apporte des informations importantes puisque celle-ci est faite en mouvement. Cependant, le traitement de ces signaux est une tâche complexe puisque sa biomécanique lui permet de bouger dans plusieurs degrés de liberté. Cela nécessite donc des techniques innovatrices pour effectuer l'acquisition des signaux. Cette articulation est soumise à diverses contraintes statiques et dynamiques qui peuvent entraîner une ou plusieurs pathologies du genou, dont l'arthrose [26]. Les articulations atteintes de l'arthrose ressentiront les impacts par un effet sur la mobilité de celles-ci. L'arthrose est une maladie chronique et l'une des plus importantes causes dans l'utilisation des systèmes de la santé [26].

1.2.Objectif et méthodologie

La présente recherche a pour objectif général de développer des méthodes d'apprentissage machine pour la prédiction de l'effet de la réadaptation physique sur le genou pathologique. Plus précisément, ses objectifs spécifiques sont : (1) la mise en place d'un processus d'analyse des données à l'aide de nouvelles méthodes d'évaluation de l'importance des caractéristiques dans le but de développer un système de prédictions de bonne performance, c'est-à-dire dont la

précision et le pouvoir de généralisation sont élevés et (2) la mise en place de canaux de préparation des données en utilisant une méthode non supervisée afin d'extraire les caractéristiques les plus pertinentes. En outre, celle-ci a pour effet la réduction des dimensions afin d'offrir un jeu de données réduit et représentatif à un algorithme d'apprentissage supervisé afin d'en faire l'entraînement.

1.3.Plan

Afin d'exprimer clairement la démarche, le document est divisé en chapitres pour aborder distinctement les thèmes liés à l'évolution de la recherche actuelle. Outre le chapitre de conclusion, cet ouvrage se divise selon les chapitres principaux suivants :

- Le genou et sa biomécanique (Chapitre 1)
- Le cycle de marche (Chapitre 2)
- L'apprentissage machine (Chapitre 3)
- L'analyse des résultats (Chapitre 4)

Les deux premiers chapitres expriment les informations nécessaires sur la provenance des signaux, soit ceux du genou pathologique, acquis pendant l'exécution du cycle de marche.

Le chapitre 3, intitulé Apprentissage machine, énonce la théorie en lien avec les méthodes utilisées.

Le document se termine avec une analyse des résultats des différentes expérimentations, suivi d'une conclusion afin d'exprimer la vision en lien avec cette recherche et les futures.

2. LE GENOU ET SA BIOMÉCANIQUE

Le genou est une articulation complexe, ce qui génère beaucoup de mouvements. Cela se définit par le type d'articulation qu'est le genou. Ainsi, dans ce chapitre, nous abordons les composantes qui constituent le genou et ce qui lui permet sa mobilité.

2.1. Les articulations

Le corps humain est composé de plusieurs articulations. Une articulation est la liaison entre deux ou plusieurs os. Il existe trois types d'articulations :

- Fibreuse (synathrose)
- Cartilagineuse (amphiathrose)
- Synoviale (diathrose)

Plusieurs caractéristiques les distinguent, mais principalement leur mobilité. L'articulation fibreuse est dite immobile ; par exemple, les os du crâne. Les articulations cartilagineuses, un peu plus mobiles se retrouvent par exemple au niveau des côtes. Les articulations complètement mobiles, comme les articulations des doigts, sont des articulations synoviales.

2.1.1. Articulation fibreuse

L'articulation fibreuse se constitue des os adjacents liés par un tissu fibreux. Les os sont directement connectés ensemble. Il n'y a pas de

cavité présente entre les os. Par exemple, le crâne est composé d'articulations fibreuses. Ce type d'articulation n'est pas mobile.

2.1.2. Articulation cartilagineuse

Liés par un cartilage, les os forment ce que l'on appelle une articulation cartilagineuse. Une articulation cartilagineuse, comme celle de la colonne vertébrale est dite semi-mobile. Cette liaison est différente des autres, puisque le tissu qui lie deux os est robuste, mais flexible.

2.1.3. Articulation du genou

Le genou est une des articulations les plus mobiles du corps humain et permet plusieurs mouvements, dont les suivants :

- Flexion : Réduction de l'angle de l'articulation
- Extension : Augmentation de l'angle de l'articulation
- Abduction : Éloignement par rapport au centre du corps
- Adduction : Rapprochement par rapport au centre du corps
- Rotation interne : Mouvement autour d'un axe vers le centre du corps
- Rotation externe : Mouvement autour d'un axe s'éloignant du centre du corps

- Circumduction : Mouvement qui combine les différents mouvements. Le membre distal décrira alors un cercle complet.

2.2.L'articulation du genou et sa biomécanique

Une articulation lie des os. Les os sont catégorisés en cinq grandes familles. Les os longs, les os courts, les os plats, les os irréguliers (formation axis p. 28) et les sésamoïdes.

Le genou fait la jonction entre trois os, soit : le fémur, la patella et le tibia.

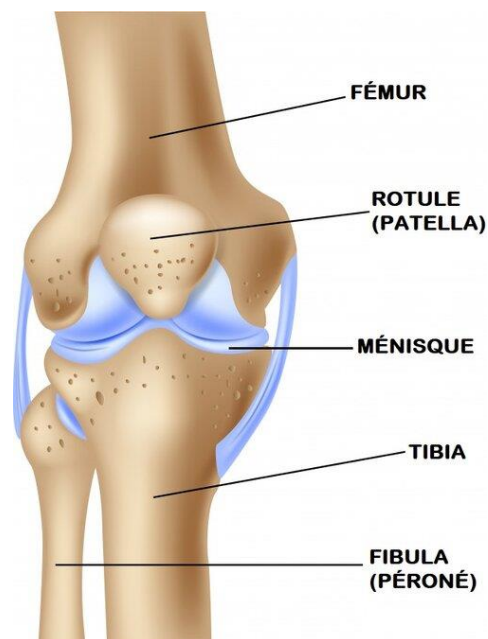


Figure 1 L'articulation du genou [20]

Par sa constitution, le genou effectue des mouvements de flexion-extension, abduction-adduction et rotation interne-externe. Donc, le genou est une articulation à six degrés de liberté.

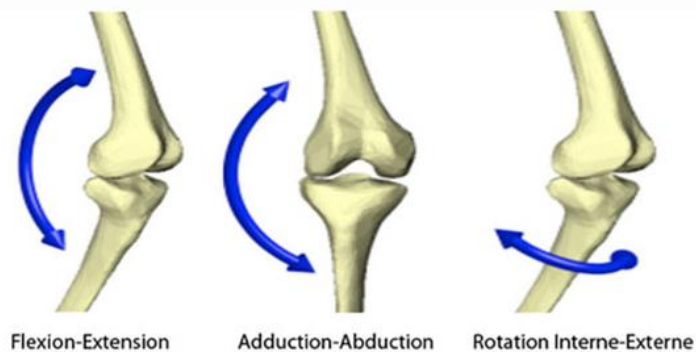


Figure 2 – Mouvement de l'articulation du genou [21]

Tout comme plusieurs articulations, l'efficacité et la précision des mouvements dépendent des muscles. Le genou fait la jonction de trois os de façon imparfaite. C'est-à-dire que l'emboîtement des os dans les cavités n'est pas parfait. Ceci augmente l'importance des muscles afin de solidifier et de contrôler le mouvement.

2.3. Le fémur

Le fémur est l'os le plus long et le plus fort du corps humain¹. Il fait partie de la catégorie des os longs. Il constitue l'articulation de la hanche ainsi que celle du genou avec sa partie distale.

¹ <https://assets.openstax.org/oscms-prodcms/media/documents/AnatomyandPhysiology-OP.pdf> page 328



Figure 3 - Le fémur [29]

2.4. La patella

La patella, plus communément nommée la rotule, se retrouve à la jonction entre le fémur et le tibia. Elle constitue l'articulation du genou. Elle se situe à la partie distale du fémur. La patella fait partie des os sésamoïdes².

² Wikipédia, https://fr.wikipedia.org/wiki/Os_s%C3%A9samo%C3%AFde

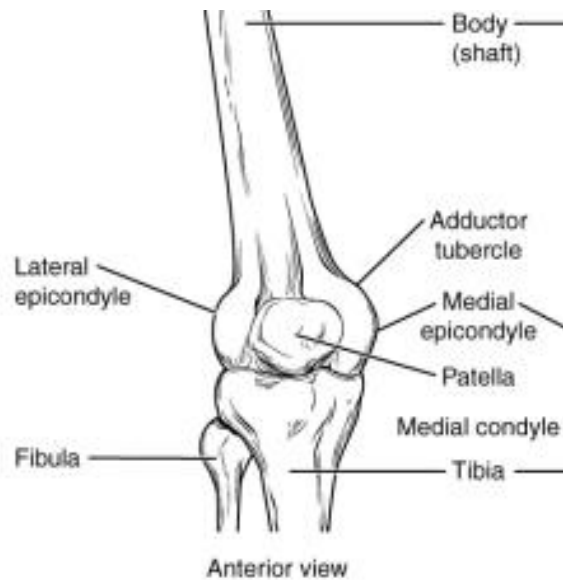


Figure 4 - La patella (rotule) [30]

2.5. Le tibia

Le troisième os principal constituant l'articulation du genou est le tibia. Le tibia fait également partie de la catégorie des os longs ; c'est le deuxième os le plus long du corps. Comparativement au fémur, qui est seul, le tibia est en paire avec la fibula.

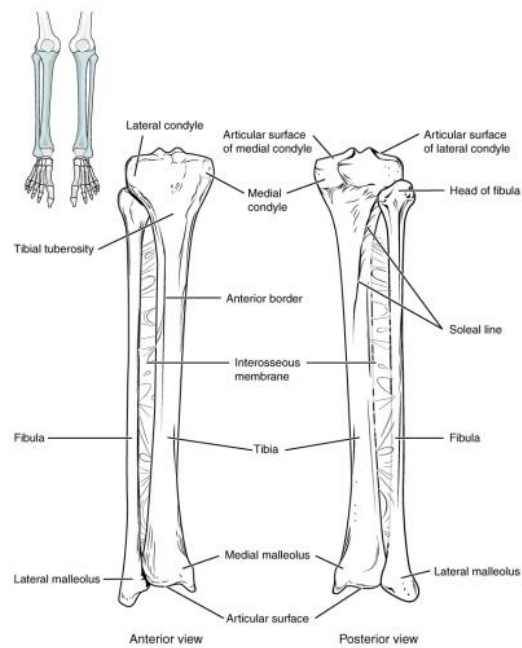


Figure 5 - Le tibia [30]

2.6. Pathologies du genou – Arthrose

L'arthrose est une maladie qui peut toucher toutes les articulations. Cette maladie est évolutive. Elle entraîne la détérioration du cartilage. Une personne atteinte d'arthrose au genou ressentira des douleurs lorsqu'elle effectue des mouvements.

La cause de l'arthrose au genou est l'incapacité du corps à réparer le cartilage endommagé, fissuré. Au fil du temps, des excroissances

osseuses peuvent se former et nuire au mouvement et créer de la douleur³.

L'arthrose au genou est la pathologie la plus susceptible de causer des invalidités fonctionnelles. Elle touche près de quatre millions de Canadiens, soit 13.6 % de la population, selon l'Agence de la santé publique du Canada [24]. D'où notre intérêt à cette pathologie dans ce travail de recherche.

³ Danault D., Dupontgand S. (2018), Certification en conditionnement physique, Anatomie physiologique [Manuel de l'élève]

3. LE CYCLE DE MARCHE

L'activité de la marche est simple pour l'homme puisque de façon générale, celui-ci l'exécute sans réfléchir. Cette activité plus complexe qu'il n'y paraît, soit l'action de marcher, est une succession de cycles. Ainsi, dans ce chapitre, nous discuterons du cycle de marche. À travers celui-ci, le genou et ses composantes traversent différentes positions, ce qui génèrent plusieurs mouvements des différentes composantes du genou.

3.1. Le cycle de marche

Dans son sens général, un cycle est défini par :

« Suite de phénomènes se renouvelant dans un ordre immuable. »⁴

Le cycle de marche représente la succession de mouvements physiques avec un retour à l'état initial. Le cycle de marche se découpe principalement en deux phases majeures :

- La phase d'appui : Phase dans laquelle le pied est en contact avec le sol.

⁴<https://www.larousse.fr/dictionnaires/francais/cycle/21275#:~:text=1.,immuable%20%3A%20Le%20cycle%20des%20saisons.&text=2.&text=Toute%20succession%20r%C3%A9gulier%20r%C3%A8glement%20r%C3%A9p%C3%A9titif%20de,l'ensemble%20du%20monde%20vivant.>



Figure 6 - Cycle de marche - Phase d'appui [31]

- La phase d'oscillation : Phase dans laquelle le pied est décollé du sol.



Figure 7 - Cycle de marche - Phase d'oscillation [31]

Un cycle de marche se termine lorsque le pied de départ revient à sa position initiale. Dans l'image suivante, on parle de la jambe droite.

Les phases d'appui et d'oscillation se produisent en alternance pour chaque membre dans un cycle de marche normal.

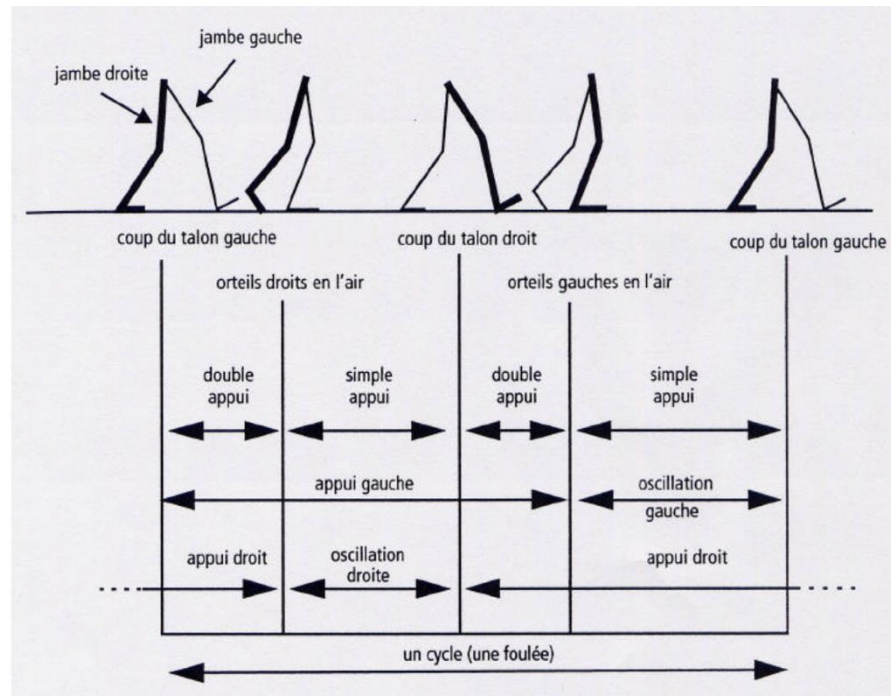


Figure 8 - Cycle de marche complet⁵

Dans chacune des phases, on retrouve les différents mouvements de l'articulation comme cités plus haut : flexion, extension, rotation interne, rotation externe, abduction et adduction.

⁵<http://www2.ift.ulaval.ca/~dupuis/Modelisation%20et%20animation%20par%20ordinateur%20IFT-66819%20et%20IFT-22726/Mod%C3%A9lisation%20du%20corps%20humain/Marche.pdf>

4. APPRENTISSAGE MACHINE

Dans ce chapitre nous discuterons de la théorie derrière l'expérimentation qui constitue cette recherche. Nous présenterons les différentes méthodologies nécessaires dans le processus de mise en place d'un processus d'apprentissage automatique.

4.1. Traitement des données

Le traitement des données est une phase importante. Que ce soit pour nettoyer ou normaliser, les opérations de traitement des données permettent d'améliorer les performances d'un système de prédiction.

4.1.1. Nettoyage

Lors de l'acquisition des données, peu importe la méthode, il est possible de retrouver des erreurs dans les données, des valeurs manquantes ou des valeurs aberrantes.

Le processus de nettoyage permet de corriger les différentes anomalies.

- Bruit dans les données

Lors de l'acquisition, il est possible de se retrouver avec des valeurs qui n'ont rien à voir avec ce que l'on attendait. Soit parce que l'outil

de captation est défaillant, soit parce que la méthode d'acquisition ne permettait pas la réponse attendue. Il existe aussi des processus d'acquisition de données subjectives par l'entremise d'entrevues. Si ce processus est défaillant, on se retrouve avec des erreurs dans notre jeu de données.

Dans le cas où la captation, ou le processus a mal tourné pour un seul individu, ou un petit groupe d'individus, il peut être nécessaire de retirer ces individus du jeu de données.

Si l'erreur de captation se trouve seulement sur un attribut, alors on est dans le contexte de valeur manquante ou aberrante. Le bruit affecte généralement un ensemble de données.

- Valeur manquante

C'est un attribut dont la valeur est nulle, peu importe la raison. Une valeur nulle ne peut pas être acceptée et doit correspondre à un choix dans une liste de valeur concrète selon le jeu de données. Dans ce cas, nous la remplacerons par une valeur adéquate. Le choix de cette valeur peut être fait de différentes façons. Par exemple, on peut remplacer les valeurs nulles par la moyenne des valeurs existantes.

Dans un cas extrême où trop de valeurs d'un individu seraient manquantes, on pourrait décider de retirer l'individu. Nous sommes dans un contexte d'erreurs de captation ou de processus.

- Valeur aberrante

Les données aberrantes (outliers) sont des valeurs extrêmes par rapport à l'ensemble des données à analyser. En effet, parfois, certains attributs, pour un individu en particulier, se retrouvent avec des valeurs extrêmes. Que ce soient des valeurs extrêmes positives ou négatives, il faut une façon de déterminer si ces valeurs ont du sens et si on doit les conserver telles quelles. Celles-ci peuvent avoir un effet significatif sur la qualité de nos modèles.

4.1.2. Réduction des dimensions

L'apprentissage machine peut être un long processus. Celui-ci demande beaucoup de données. Afin d'optimiser la capacité d'apprentissage, on a parfois recours à la réduction des dimensions de notre jeu de données. L'objectif est de réduire le nombre d'attributs nécessaires à l'apprentissage. La réduction n'est pas exécutée de façon aléatoire. Celle-ci est exécutée avec différentes techniques afin de faire ressortir des attributs significatifs. Par exemple, une image à haute résolution contient une plage de couleur très grande. Lorsqu'on utilise un outil de conversion pour réduire la taille de cette image, on effectue une réduction du nombre d'attributs, en l'occurrence les couleurs, pour conserver celles qui sont significatives. Même, si le nombre d'attributs a été réduit, l'image est toujours représentative puisque les attributs significatifs ont été conservés. Tel que présenté dans la figure 9, la partie centrale représente les attributs significatifs qui permettront de reconstruire l'image.

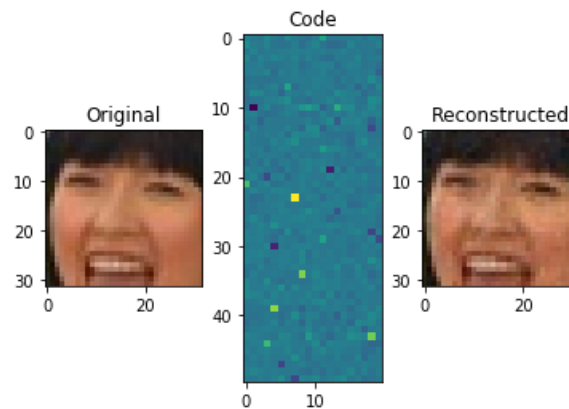


Figure 9 - Autoencodeur avec images et réduction de la dimension
 (<https://stackabuse.com/autoencoders-for-image-reconstruction-in-python-and-keras/>)

4.2. L'apprentissage

Afin de développer un système basé sur l'apprentissage machine, les algorithmes nécessitent généralement deux phases. La première consiste à estimer un modèle à partir d'un jeu de données, appelé observations ou données d'entraînement. Durant cette phase, le modèle fait ressortir les caractéristiques pertinentes qui aideront à faire la classification ou la prédiction. La seconde phase correspond à la mise en production : le modèle étant déterminé, de nouvelles données peuvent alors être présentées afin d'obtenir le résultat correspondant à la tâche souhaitée (ex : classification, prédiction ou regroupement). Pour ce faire, nous avons recours à différents algorithmes. Ces algorithmes se différencient, entre autres, par leurs méthodes d'apprentissage. Nous pouvons catégoriser ces algorithmes en deux catégories principales, soit : les algorithmes supervisés et non supervisés.

4.2.1. Apprentissage supervisé

Lorsqu'on parle d'apprentissage supervisé, cela signifie que l'entraînement des algorithmes doit se faire à partir d'un jeu de données dont la classe de chaque individu est connue d'avance.

L'algorithme peut alors comparer la sortie prédite avec la classe réelle afin d'en faire ressortir un taux d'erreurs et continuer à s'améliorer.

On compte parmi ce type d'algorithmes : la régression, les arbres de décision, Bayes, SVM, ainsi que les réseaux de neurones.

L'avantage principal de ce type d'apprentissage est la précision. Étant donné que l'algorithme connaît les classes d'avance, il est en mesure de se valider. De plus, dans la mesure où l'algorithme n'est pas efficace dans son apprentissage, le taux d'erreurs nous l'indiquera.

Cependant, l'un des désavantages est l'acquisition et la catégorisation d'un jeu de données. Dans certains cas, l'acquisition de données peut être complexe. De plus, pour être efficace, la taille du jeu de données doit être importante. L'objectif est de trouver la régularité dans les données d'entrées⁶. Plus le jeu de données est grand et varié, plus l'algorithme supervisé aura ce qu'il faut pour bien apprendre. Cela s'ajoute au désavantage d'avoir un travail fastidieux, afin d'assigner chaque classe à chaque individu.

⁶ Introduction to machine learning, p.11

4.2.2. Apprentissage non supervisé

L'absence de connaissances des classes des individus (labels ou étiquettes), qui constitue le jeu de données, nécessite des algorithmes dont l'apprentissage est non supervisé. C'est-à-dire des algorithmes n'ayant pas besoin de connaître la classe des individus afin de les associer à un groupe. Comme les classes sont inconnues, les algorithmes non supervisés vont tenter de regrouper les individus similaires ensemble, par eux-mêmes. Selon l'algorithme, la méthode de regroupement sera différente. Cela dit, ces algorithmes se basent généralement sur une mesure de distance intragroupe et intergroupe. Il est à noter que les algorithmes non supervisés sont parfois utilisés afin de générer les classes, pour ensuite passer le jeu de données à un algorithme supervisé.

L'avantage des algorithmes non supervisés est la capacité d'apprendre par eux-mêmes avec un jeu de données brutes, sans connaître la classe de l'individu.

4.3. Arbre de décision

L'arbre de décision est un algorithme d'apprentissage supervisé qui applique le concept de diviser pour régner. L'algorithme sépare récursivement les nœuds à chaque étape.

Un arbre de décision est composé de :

- Une racine.
- Des nœuds : Représente une séparation du jeu de donnée

- Des nœuds terminaux, nommés feuilles, qui représentent la classe ou la valeur dans le cas d'une régression

Comme le démontre la figure 10, chaque nœud participe à la décision en appliquant une fonction de comparaison afin de séparer le jeu de données en différentes classes.

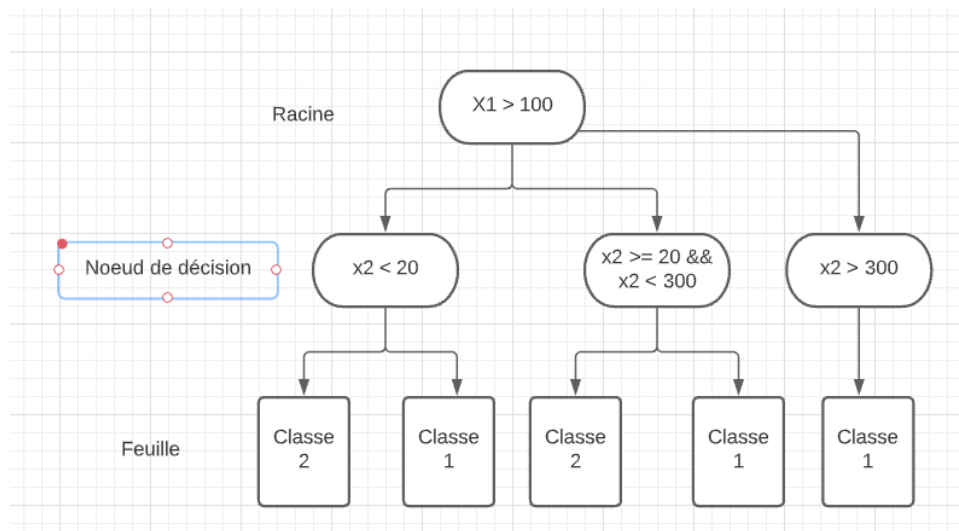


Figure 10 - Arbre de décision

Un arbre de décision peut être utilisé pour la classification. Celui-ci utilise une mesure d'impureté afin d'effectuer la séparation (split) des nœuds. Les feuilles représentent les classes.

Cet algorithme utilise différentes fonctions pour déterminer l'impureté, comme l'entropie et l'indice de Gini.

L'arbre de décision offre quelques avantages ont la simplicité, la rapidité et l'interprétabilité.

Ce dernier avantage est particulièrement le plus important. En effet, dans le domaine de la santé, la possibilité de créer un arbre de décision et de le fournir à une physiatre qui l'utilisera et pourra l'interpréter est très intéressante dans ce domaine pour bien justifier la décision du professionnel. Cela dit, peu d'algorithmes sont interprétables comme l'arbre de décision.

4.3.1. Entropie

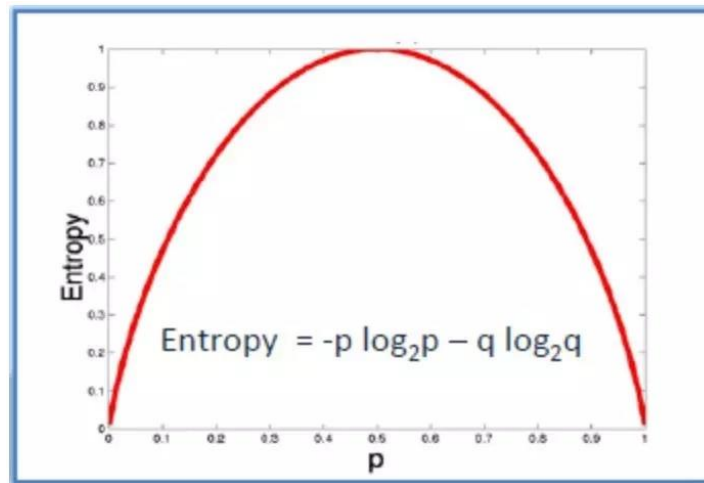
L'entropie est une mesure de la quantité d'informations. Cette mesure est utilisée afin de déterminer le bon partage des nœuds pour la construction de l'arbre. Dans un problème à deux classes, positif et négatif, l'arbre de décision est l'encodage de cette décision. L'information nécessaire pour représenter les deux classes se représente par la formule suivante où p la probabilité :

$$I(p,n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

Figure 11 - Formule de l'entropie⁷

$I(p,n)$ représente l'information d'un sous arbre nécessaire à la construction de l'arbre principal. Donc, lorsqu'il n'y a qu'une seule classe, l'entropie (soit la quantité d'information) est nulle, tel que démontré dans la figure 11.

⁷ Induction of decision tree, quinlan 1986



$$\text{Entropy} = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$$

Figure 12 - Courbe des valeurs de l'entropie [28]

4.3.1.1. Indice de Gini

L'indice de Gini est une autre méthode pour calculer l'impureté. Celui-ci a été introduit par Breiman. Dans le cas d'une classe positive et négative, l'indice de Gini se définit de la façon suivante où A représente l'ensemble des données et A^i le nombre d'éléments i dans cet ensemble.

$$Gini(A) = 1 - \sum_i^n \left(\frac{A^i}{A} \right)^2$$

4.3.2. Algorithme

L'algorithme de base d'un arbre de décision est simple, ce qui est un avantage énorme puisqu'il est interprétable⁸.

1. Si l'entropie de $X < \delta$, créer une feuille avec la classe majoritaire
 - a. Retourner
2. Sinon
 - a. Répartir les attributs en i sous-ensemble
 - b. Trouver le meilleur attribut selon la mesure de l'entropie
 - c. Créer un nœud pour cet attribut
 - d. Recommencer pour créer des nœuds enfants jusqu'à la condition d'arrêt

4.4. XGBOOST

L'algorithme XGBOOST est un algorithme qui mélange plusieurs méthodes afin de pallier la problématique principale des arbres de décision, soit la certitude d'avoir généré le bon arbre. Pour ce faire, il va utiliser une première méthode qui se nomme *l'amplification de gradient*.

⁸ decomite-Arbres.pdf, <https://www.fil.univ-lille1.fr/~decomite/ue/MFFDD/Arbres.pdf>

4.4.1. Amplification du gradient

L'amplification du gradient permet de tenir compte de plusieurs modèles d'un algorithme traditionnel. Cette technique peut être utilisée avec différents algorithmes comme l'arbre de décision et les forêts aléatoires (*random forest*).

L'objectif principal de l'amplification du gradient est d'unifier les modèles dits faibles, afin de les rendre plus fort ensemble. Prenons l'exemple du golfeur qui frappe un grand coup pour atteindre sa cible (prédiction F_0). Une fois près de sa prédiction, une suite de petits coups (delta 1) afin d'atteindre la cible sera effectuée (Figure 12).

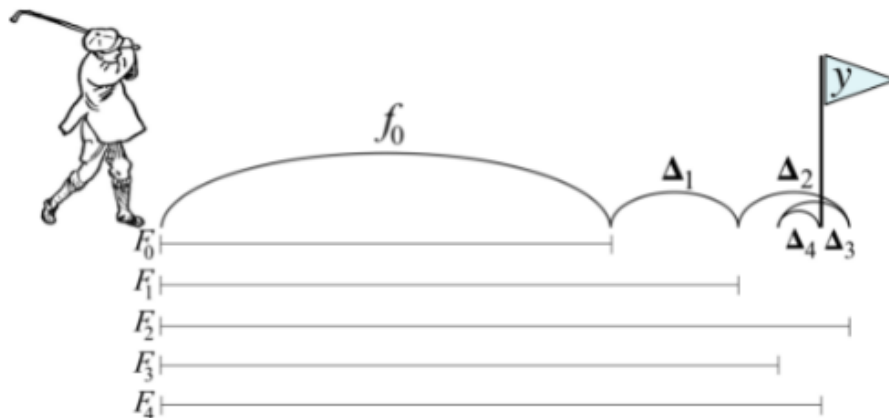


Figure 13 - Représentation visuelle du gradient⁹

Du point de vue mathématique :

⁹ XGBOOST - <https://explained.ai/gradient-boosting/L2-loss.html>

1. F_0
2. $F_1 = F_0 + \text{delta } 1$
3. $F_2 = F_1 + \text{delta } 2$
4. $F_3 = F_2 + \text{delta } 3$
5. $F_4 = F_2 + \text{delta } 4$

Cet apprentissage itératif représente la méthode de l'amplification du gradient. Afin d'être efficace, cet algorithme ajoute un hyperparamètre qui représente le taux d'apprentissage η afin de contrôler la grandeur du delta.

Dans l'exemple précédent, chaque $F_M(x)$ représente un nouvel arbre qui est l'enfant du précédent.

Afin de déterminer jusqu'à quel moment l'algorithme doit continuer de générer des modèles, l'utilisation d'une fonction de perte est nécessaire. Une fonction populaire est l'erreur quadratique moyenne [27]. À chaque itération, celle-ci permet de déterminer à quel point la prédiction de l'algorithme est près de son objectif. La fonction de perte L qui représente l'erreur quadratique des N éléments sur la valeur y , étant la valeur de notre jeu de données.

$$L(y, F_M(X)) = \frac{1}{N} \sum_{i=1}^N (y_i - F_M(x_i))^2$$

4.4.2. XGBOOST, l'algorithme

Xgboost est un algorithme de type amplification du gradient. C'est-à-dire qu'il crée itérativement des modèles faibles afin de les unifier

pour être plus fort et avoir de meilleures prédictions qui généralisent bien.

Il utilise un terme de régularisation de sa fonction, tout comme le gradient boosting.

À chaque itération, il construit des arbres. Afin de construire le bon arbre, il doit s'assurer que les bons attributs ont été choisis. Afin d'évaluer cela, il utilise une fonction dont les termes représentent la fonction de perte (L) et le terme de régularisation Ω qui permet d'ajouter une pénalité pour minimiser les erreurs.

$$\text{obj}(\theta) = L(\theta) + \Omega(\theta)$$

L'ajout de Xgboost se situe au niveau de la modélisation de la complexité du modèle. Au lieu de se fier seulement à l'impureté et assumer que l'arbre généré est le bon, Xgboost ajoute ce nouveau paramètre ($\text{Obj}(\text{téta})$).

Donc, Xgboost formalise la complexité d'un arbre afin de donner des scores à chaque feuille afin de déterminer le bon sous arbre à utiliser.

Ici, afin de calculer la complexité, T représente le nombre de feuilles dans l'arbre en cours d'évaluation. On ajoute à celui-ci la sommation de w . Cette variable est un tableau contenant le score de chaque feuille. Les autres termes sont des termes de régulation [26].

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

Complexité d'un arbre

La complexité utilise w , le score d'une feuille. En théorie, l'objectif est de générer tous les arbres, de trouver le meilleur.

4.5. Réseau de neurones

Les réseaux de neurones sont une imitation du fonctionnement du cerveau humain. La connexion entre les neurones propage la connaissance afin d'en retourner un résultat.

4.5.1. Objectif du réseau de neurones

En intelligence artificielle et plus précisément en apprentissage automatique, l'objectif est d'approximer une fonction. Le contexte dans lequel apparaît la nécessité d'avoir de l'apprentissage automatique se présente lorsque nous avons un historique de données, à partir duquel nous voulons apprendre un modèle afin de prédire des résultats avec de nouvelles valeurs de paramètres. Par exemple, détecter la langue d'un texte. Avec l'historique des mots que je connais, je peux tenter de prédire la langue d'un nouveau texte que je ne connais pas. Un autre exemple classique est la prédiction du prix d'une maison selon certains critères. En fonction

des données et de ce que l'on veut en faire ressortir, il existe différentes techniques. Des méthodes supervisées ou non.

Alors, s'il existe une tonne de méthodes, pourquoi les réseaux de neurones ? L'exemple simple et efficace pour expliquer ceci est la représentation de la fonction XOR que l'on retrouve généralement dans la littérature. La fonction XOR spécifie que la fonction retourne 1 si et seulement si un des deux paramètres vaut 1.

$$0, 0 = 0$$

$$0, 1 = 1$$

$$1, 0 = 1$$

$$1, 1 = 0$$

Voici la représentation graphique des valeurs de XOR

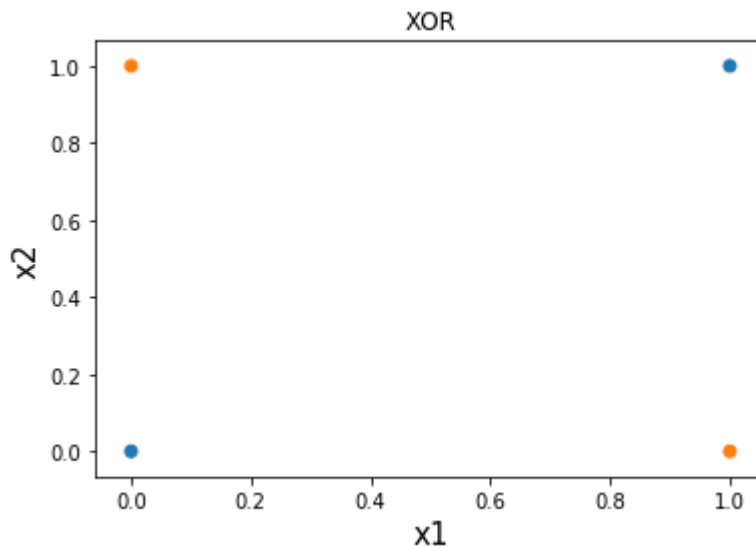


Figure 14 - XOR Classification non linéaire, orange correspond à VRAI et bleu à FAUX[2]

En orange, c'est la valeur de sortie 1 et en bleu, la valeur de sortie 0.

On peut comprendre visuellement que ce n'est pas linéaire. Mais disons que l'on tente d'approximer la fonction XOR à l'aide d'une fonction linéaire, soit la régression linéaire.

$$f(X; W; b) = W * X + b$$

X étant un ensemble d'entrées, les valeurs de W sont les poids synaptiques et b représente le biais.

Cette équation représente une régression linéaire, mais aussi le perceptron, soit un réseau de neurones à sa plus simple représentation.

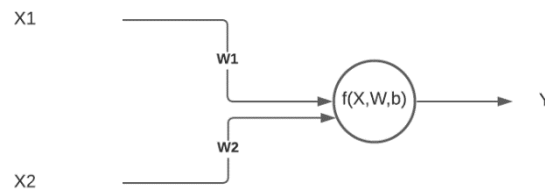


Figure 15 – Perceptron

Le perceptron effectue la sommation pondérée des entrées. Il existe différentes théories pour initialiser les poids. Généralement, les poids synaptiques sont initialisés aléatoirement.

Si on effectue un premier apprentissage et que l'on calcule l'erreur quadratique, le résultat est le suivant :

Prédiction = $[[0. 0. 0. 0.]]$

Delta = $[[0. 1. 1. 0.]]$

Erreur quadratique : 0.5

À la suite de cet échec, on devra ajuster les poids pour tenter de mieux prédire. C'est ce qui correspond à l'apprentissage.

Disons que l'on ajuste le biais à 0.5. On se retrouve avec une prédiction erronée (Figure 15).

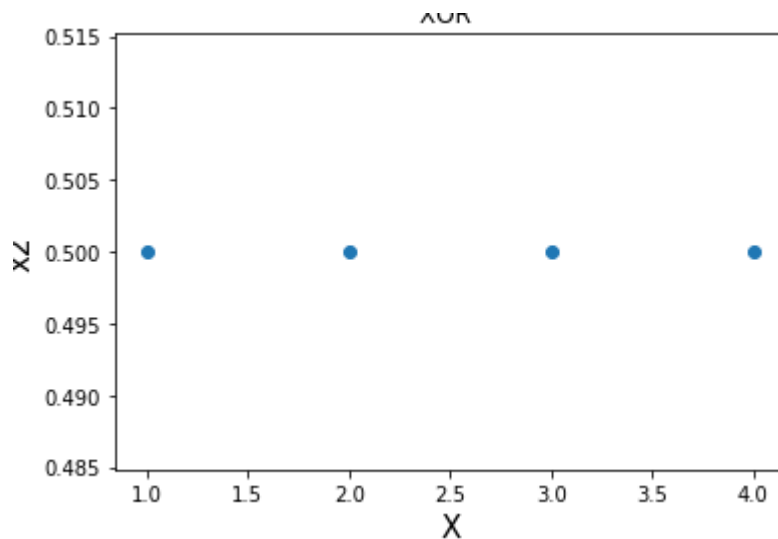


Figure 16 - Xor - classifications avec le perceptron [2]

Ce qui se produit, c'est la démonstration de l'incapacité d'un modèle linéaire à apprendre la fonction XOR. La fonction linéaire sépare en

zones bien distinctes. Elle trace une ligne entre deux zones, 0 et 1. Ce qui n'est pas la représentation de la fonction XOR.

Par cet exemple, on comprend que les approches linéaires ne sont pas suffisantes.

Afin d'obtenir un réseau de neurones efficace, nous ajouterons ce que l'on nomme la fonction d'activation et la descente du gradient.

La fonction g est une fonction d'activation quelconque qui aide à traiter le problème non linéaire. Elle reçoit en paramètre. Le résultat d'une équation linéaire composé de $W.T$, qui est un vecteur de poids. Celui-ci multiplié à x , une matrice qui correspond aux valeurs fournies à l'algorithme. S'additionne b , qui représente le biais. Ce biais est ajusté expérimentalement.

$$h = g(W.T * x + b)$$

4.5.2. Fonction d'activation

La fonction d'activation a pour but premier de convertir la fonction linéaire en fonction non linéaire. Dans la littérature, on parle de différentes fonctions d'activations :

- SIGMOÏDE
- TANH
- RELU (rectifier linear unit)
- Leaky RELU
- SOFTMAX

SIGMOÏDE

La fonction sigmoïde est l'une des premières fonctions utilisées. Cependant, elle est de plus en plus mise de côté et utilisée seulement pour la couche de sortie.

La fonction sigmoïde permet d'obtenir un résultat entre 0 et 1, peu importe la valeur de x . La formule du sigmoïde ou e la constante d'Euler et x , la valeur d'entrée, se représentent de la façon suivante (Figure 16).

$$a = \frac{1}{1 + e^{-x}}$$

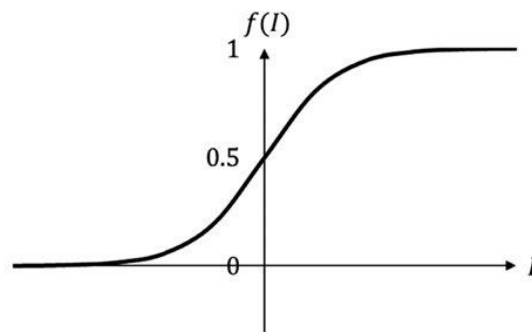


Figure 17 - Fonction sigmoïde[32]

TANH

TANH est une autre fonction qui permet de générer une sortie en -1 et 1. Dans les faits, elle est une version décalée de la fonction SIGMOÏDE.

Cela fait en sorte que les données sont plus centrées sur 0, puisque la fonction TANH passe par le point central. Celle-ci est représentée par l'équation suivante dans laquelle e représente la constante d'Euler et x la valeur d'entrée.

$$\tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} = \frac{e^{2x} - 1}{e^{2x} + 1} = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

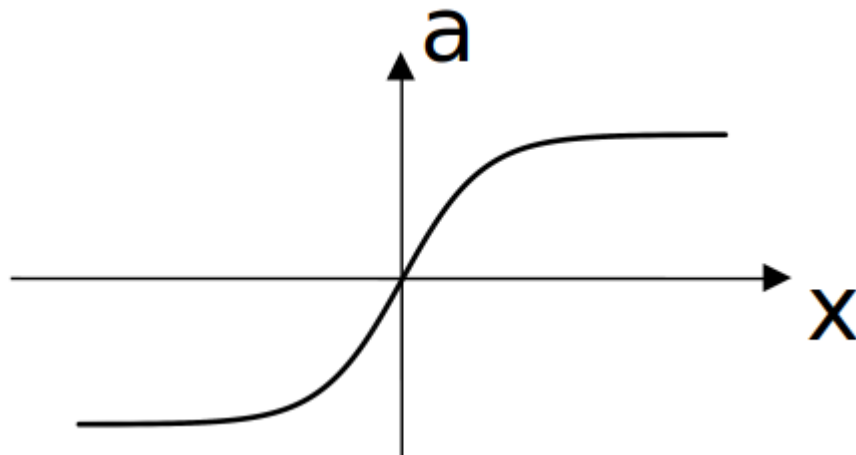


Figure 18 - Fonction Tanh [33]

RELU

La fonction RELU est la nouvelle fonction privilégiée par défaut.

L'avantage principal de la fonction RELU, et qui en fait sa popularité, est sa vitesse de calcul.

La fonction RELU est toujours positive. On constate que la différence principale entre une fonction linéaire et l'unité de rectification linéaire, c'est qu'elle retourne 0 sur la moitié de son domaine [2]. Elle consiste simplement en une fonction qui retourne le maximum entre 0 et x, la valeur d'entrée.

$$a = \max(0, x)$$

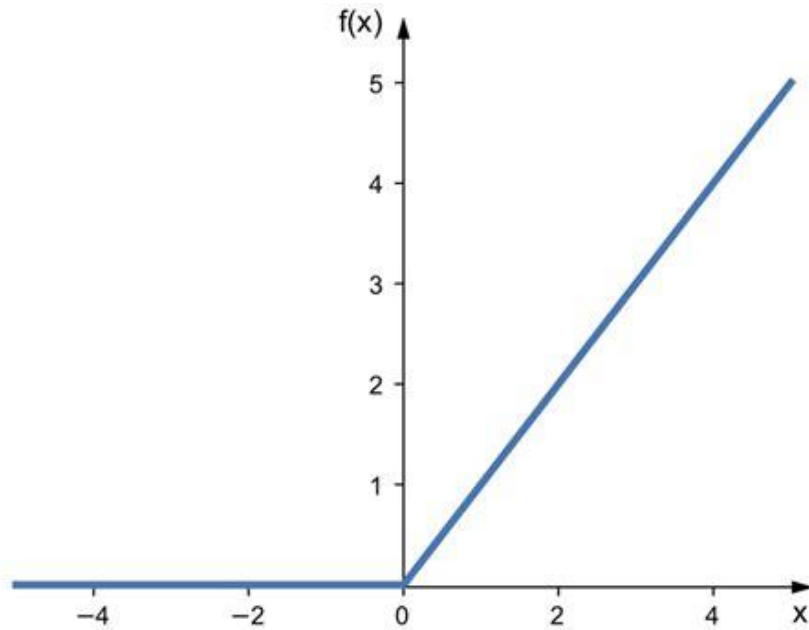


Figure 19 - Fonction d'activation RELU [34]

LEAKY RELU

La fonction LEAKY RELU ajoute un paramètre alpha à la fonction RELU. Dans la littérature [2] on attribue généralement 0.01 à alpha. Alpha peut très bien être un paramètre et même un paramètre qui peut être inféré.

$$a = \max(\alpha * x, x)$$

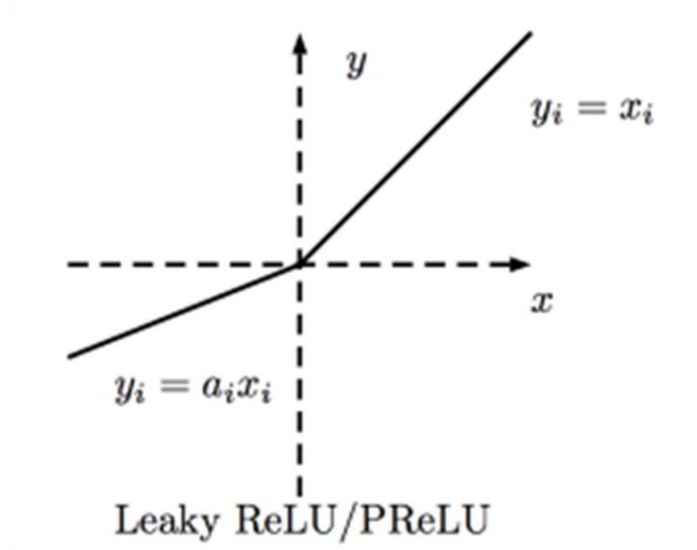


Figure 20 - Fonction d'activation LReLU[35]

4.5.3. Descente du gradient

L'ajout de la fonction d'activation afin de convertir en fonction non linéaire n'est pas suffisant. Comme exprimé précédemment, l'objectif d'un réseau de neurones est d'approximer une fonction. Cette approximation nous donne des valeurs de poids réutilisables sur de nouvelles données afin de prédire une sortie. On parle d'apprentissage ou d'optimisation. L'approximation s'optimise par l'algorithme de la descente du gradient. Généralement, on parle de minimisation de $f(x)$. Ce que nous désirons minimiser est la fonction de coût, parfois appelée la fonction de perte ou d'erreur. La descente du gradient utilise la pente afin de minimiser la fonction. L'algorithme utilise la dérivée afin de déterminer celle-ci.

$$f(x') = \frac{dy}{dx}$$

Selon la pente, cela indique à l'algorithme la façon d'optimiser ses paramètres. Une pente de zéro ne donne aucune information. Dans le cas d'une pente négative ou positive, nous voulons déplacer dans le sens inverse de celle-ci, afin de se rapprocher du minimum. Ce déplacement est fait par petits pas et correspond à un hyperparamètre du réseau de neurones. Celui-ci se nomme le taux d'apprentissage (α -learning rate). La figure suivante du livre Deep Learning exprime très bien la descente du gradient.

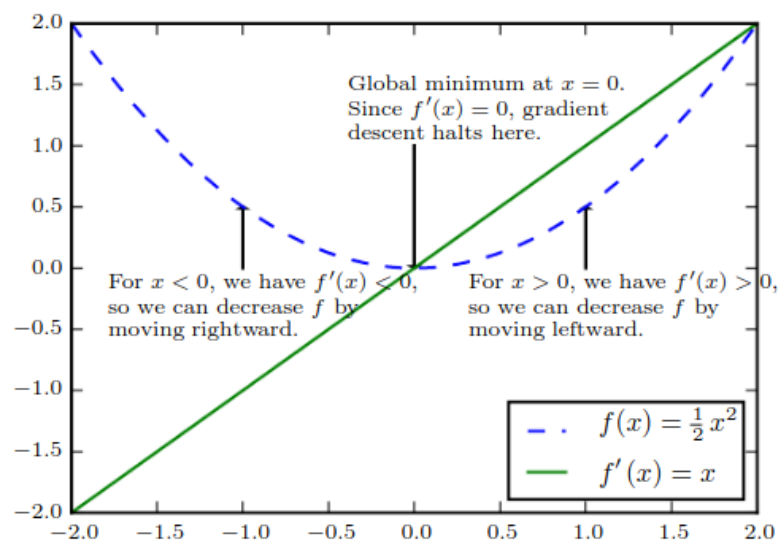


Figure 21 - descente du gradient [2]

En résumé, la descente du gradient nous permet d'optimiser les poids en minimisant la fonction de coût. Avec la notion d'optimisation, nous

pouvons maintenant construire un réseau de neurones plus complet que l'on nomme réseau à propagation avant.

4.5.4. Réseau de neurones à propagation avant

Les réseaux de neurones à propagation avant, parfois nommés perceptrons multicouches, représentent le modèle classique des réseaux de neurones. L'objectif des réseaux de neurones à propagation avant est d'approximer une fonction. On dit qu'ils sont à propagation avant parce qu'ils font circuler l'information X d'entrée vers une couche cachée à l'autre. Chacune de ces couches cachées correspond à une fonction linéaire suivie d'une conversion non linéaire et d'une optimisation. Et cela jusqu'à la sortie Y. Mathématiquement, on peut représenter un réseau de neurones comme un appel de fonction de fonction si le réseau est composé de 4 couches de 1 neurone. Chaque neurone représente une fonction, soit f_1 , f_2 , f_3 et f_4 . Cela s'applique de la façon suivante.

$$A = f_4(f_3(f_2(f_1(\cdot))))$$

4.5.5. Rétropropagation

Afin d'être complet et efficace, le réseau de neurones optimisé par la descente de gradient doit faire la propagation inverse afin d'ajuster les poids. C'est le processus de la rétropropagation. Concrètement, cela veut dire que l'algorithme va effectuer une comparaison avec la sortie attendue et la sortie calculée afin d'obtenir la dérivée et de mettre à jour les poids de la couche N jusqu'à la couche 1. En soi, la rétropropagation est très simple. Différentes méthodes

aident à l'optimisation de l'ajustement des poids, soit optimiser l'apprentissage.

4.6. Auto-encodeur

Il existe plusieurs architectures de réseau de neurones destinés à une fonction précise. L'auto-encodeur est une architecture dont l'objectif est de décoder en sortie l'entrée X .

En effet, le concept de base d'encoder pour décoder semble inutile et mathématiquement très simple à effectuer. Cela dit, l'idée est plutôt d'entraîner l'auto-encodeur à effectuer une tâche de copie afin d'extraire les propriétés caractérisant la donnée d'entrée [2].

L'encodage et le décodage en sortie peuvent être utilisés de différentes façons. Par exemple, nous pouvons entraîner un auto-encodeur à encoder des images bruitées pour ensuite les décoder sans le bruit. Le réseau de neurones sert donc au nettoyage des images.

4.6.1. Structure d'un auto-encodeur

Un auto-encodeur est construit de deux réseaux de neurones à propagation avant : un en entrée pour l'encodage et un autre en sortie pour le décodage.

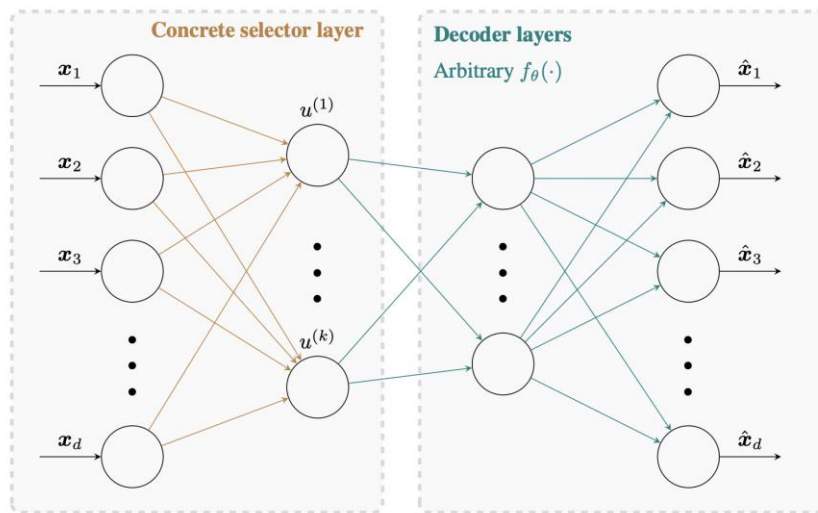


Figure 22 - Architecture d'un auto-encodeur [22]

L'objectif d'un auto-encodeur n'est pas de réussir un décodage parfait, mais de réussir à faire ressortir les attributs de valeur d'un individu, afin de le reconstruire le plus fidèlement possible. L'image démontre une symétrie, mais celle-ci n'est pas un prérequis pour les couches cachées.

Comme ce sont deux réseaux de neurones standard, les couches cachées peuvent être composées des mêmes types de couches « dropout ».

5. Méthodologie et résultats expérimentaux

L'enjeu des données est important en apprentissage automatique. Dans un premier temps, la mise en place d'un processus afin d'utiliser les meilleurs critères a été effectuée par l'entremise d'algorithmes traditionnels et les réseaux de neurones. La capacité d'un algorithme à effectuer un choix en se basant sur les meilleurs attributs tout en considérant les autres est mise en place avec l'algorithme XGBOOST. Celui-ci ajoutant des mesures de complexité aux arbres de décision afin de mieux sélectionner les attributs.

Dans un second temps, la mise en place d'un processus d'extraction des attributs les plus importants par l'entremise d'auto-encodeur a été construite. Celle-ci permettant l'extraction d'attributs afin d'entraîner un algorithme traditionnel.

5.1. Méthodologie

Rappelons que la présente recherche a pour objectif la mise en place d'une méthode d'apprentissage afin de prédire, en termes d'amélioration, l'impact de l'activité physique sur le genou pathologique par le développement de nouveaux processus d'apprentissage. Les arbres de décision étant facilement interprétables, ils sont très intéressants pour les spécialistes. Cependant, la complexité des signaux nous amène à explorer des méthodes alternatives afin de mieux généraliser la prédiction.

Processus global

La méthodologie adoptée dans ce projet de recherche se base sur les étapes principales suivantes (Figure 22) :

1. Acquisition des données
 - a. Signaux biomécaniques
 - b. Évaluation des patients, avant et après l'exécution du programme d'activité physique
2. Analyse et nettoyage des données
3. Extraction des caractéristiques biomécaniques
4. Étiquetage des patients amélioré vs non amélioré
5. Entraînement
6. Prédiction

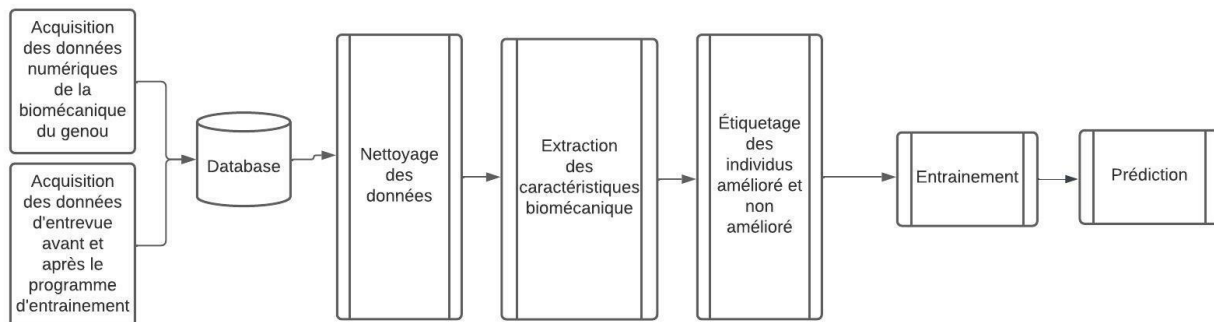


Figure 22 - Diagramme en bloc résumant les différentes étapes de la méthodologie adoptée

5.2. Base de données

En 2010, des scientifiques et cliniciens du réseau québécois de la santé (du CHUM, de l'École de Technologie Supérieure, du Centre de médecine sportive de Laval, de l'Institut de Rhumatologie de Montréal, de l'Université de Sherbrooke) ont contribué avec le partenaire industriel Emovi Inc., à la création d'un programme novateur visant l'amélioration du diagnostic et de la prise en charge de l'arthrose du genou dans le continuum de soins allant de la médecine générale à la médecine spécialisée. S'appuyant sur des données probantes et des standards de soins reconnus internationalement, le programme MonArthrose tire son originalité dans le fait qu'il allie, d'une part, une évaluation biomécanique du genou à l'aide d'une technologie de Graphie cinématique du Genou GCG (système KneeKGTM) et d'autre part, une série d'interventions ciblées et personnalisées visant à réduire les facteurs de risque d'aggravation de la maladie, à promouvoir l'activité physique et à gérer efficacement la douleur. Ce programme novateur a permis la mise en œuvre d'une base de données inédites dans laquelle figure 600 participants.

Dans le cadre de ce projet de recherche, nous avons extrait, parmi les 600 participants, 309 participants souffrant de l'arthrose au genou qui ont tous réalisé un programme d'exercices physiques. Le jeu de données contient différentes informations. Elle contient 69 facteurs biomécaniques qui sont des valeurs calculées à partir des données brutes récoltées avec l'outil d'Emovi tel qu'illustré dans l'image suivante.

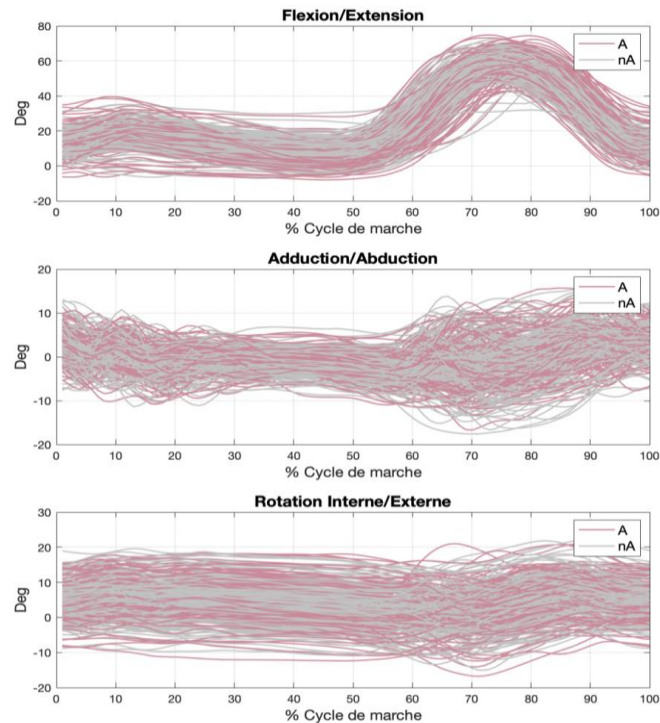


Figure 23 - - Les courbes cinématiques des 309 sujets de la base de données.

Le jeu de données contient aussi un ensemble de valeurs démographiques comme l'âge, le sexe et le poids. Pour terminer, le jeu de données contient l'évaluation KOOS (The Knee Injury and Osteoarthritis Outcome Score (KOOS)). Cette valeur est une cotation internationale pour les pathologies du genou. Le patient remplit un formulaire à partir duquel le KOOS est calculé. Celui-ci se divise en différentes échelles soit, la qualité de vie, la douleur, les symptômes, la raideur, la fonction du genou dans la pratique du sport et des loisirs et la fonction au quotidien. Chaque question se veut attribuer un score entre 1 et 5. Afin d'obtenir un score pour chaque section, par

exemple la douleur, qui contient 9 questions, nous allons utiliser la formule suivante :

$$koos\ pain = 100 - \frac{moyenne(Q1 - Q9) * 100}{4}$$

Chaque section utilise la même formule.

5.3. Apprentissage machine traditionnel

Les techniques d'apprentissage-machine basent leurs apprentissages sur des théories mathématiques. Plus précisément, selon les algorithmes utilisés, les probabilités, les calculs d'erreurs et la théorie de l'information ne sont que quelques exemples des théories mathématiques utilisées. Les arbres de décision utilisent l'entropie qui nous vient de la théorie de l'information. L'utilisation de la théorie de l'information est optimisée par d'autres techniques avec le nouvel algorithme de XGBOOST.

5.3.1. Analyse des résultats - système de classification – XGBOOST

La construction du processus afin d'entraîner l'algorithme XGBOOST se veut similaire au processus général. Celui-ci a l'avantage d'être simple et de conserver une certaine compréhension des choix effectués. La mise en place du système de classification XGBOOST a été effectuée selon le processus cité plus haut.

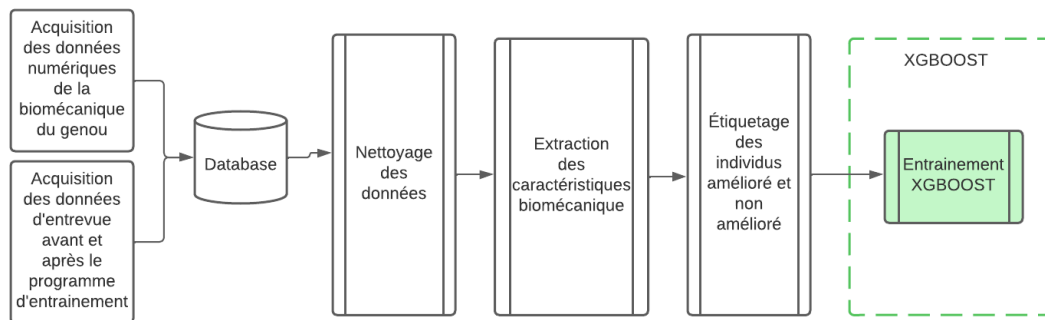


Figure 24 - Processus d'entraînement de XGBOOST

L'apprentissage de XGBOOST a été effectué avec des ensembles d'attributs complets et non discriminés. Comme la force de XGBOOST est la capacité d'évaluer l'importance des attributs, il se veut d'en avoir le plus possible afin d'en faire ressortir ceux qui sont importants. L'expérimentation a été faite avec quatre ensembles d'attributs :

- L'entièreté des courbes, c'est-à-dire les différentes courbes, ce qui correspond à un total de 300 points correspondant aux angles de mouvements durant les cycles de marche.
- Les 69 facteurs biomécaniques qui correspondent à des valeurs d'intérêts mesurées sur les courbes.
- Les 69 facteurs sans les cinq facteurs considérés les plus importants par XGBOOST dans l'expérimentation précédente.
- Extraction des meilleures combinaisons des 69 facteurs biomécaniques pour l'entraînement de XGBOOST.

L'expérimentation avec tous les attributs a été accomplie afin de faire ressortir l'importance des attributs selon Xgboost. Dans l'ensemble,

on constate une prédiction intéressante qui s'élève, par exemple, à 93% dans la population masculine.

Attributs utilisés	HOMME	FEMME	Homme-Femme
Attributs complets	93.75%	89.65%	88%
T0 + 69 facteurs biomécaniques	81.25%	82.75%	73%
Sans le top 5 des attributs	81.25%	81.25%	71.11%
Meilleure combinaison T0	93.75%	89.65%	93.33%

Tableau 1 - Résultats de la classification avec Xgboost

Le tableau suivant résume les attributs qui sont ressortis comme étant les plus significatifs durant l'entraînement selon les différents jeux de données.

Abréviations	Détails
AbsProtPfv1	Valeur absolue de la rotation au contact initial
AbsProtPfv2	Valeur absolue de la rotation à 20% du cycle de marche
CExternalTibialRotationInitialContact	Rotation externe du tibia au contact initial
COAFFixedFlexionDuringStance	Contracture en flexion durant l'appui
Pabd6	Valeur maximale de l'abduction entre le contact initial et le milieu de la phase d'appui
COAFInternalRotationTibiaFemurLoading	Rotation interne du tibia et du fémur durant la phase de chargement
COAFKneeVarusDuringStance_AllMethod * (2e calc)	Varus du genou pendant la phase d'appui
COAFValgusThrustDuringLoading_AllMethod * (1er calc)	Poussée en valgus pendant la phase de chargement
Pflex2	Valeur minimale de la flexion à 20% du cycle de marche
PabdPfv5	Abduction à 69% du cycle de marche
Pabd1	Valeur maximale de l'abduction/adduction au contact initial
Prot1	Valeur maximale de la rotation au contact initial
score_koos_pain_t0	Blessure du genou et score de résultat de l'arthrose (Knee injury and Osteoarthritis Outcome Score, http://koos.nu)

Tableau 2 - Abréviations et détails [17]

Homme	Femme	Homme-Femme
Pabd6	COAFKneeVarusDuringStance_AllMethod * (2e calc)	Pabd1
COAFFixedFlexionDuringStance	Prot1	COAFFixedFlexionDuringStance
CExternalTibialRotationInitialContact	Pflex2	COAFValgusThrustDuringLoading_AllMethod * (1er calc)
AbsProtPfv2	COAFInternalRotationTibiaFemurLoading	AbsProtPfv1
score_koos_pain_t0	PabdPfv5	score_koos_pain_t0

Tableau 3 - Xgboost - Attributs les plus important selon le sexe

Sur la figure suivante (Figure 24), on remarque que plusieurs attributs sont plus significatifs selon Xgboost. Deux expérimentations ont été effectuées. En premier lieu, nous avons effectué l'entraînement avec les 69 facteurs biomécaniques ainsi que les informations à T0. À la suite de l'entraînement de l'algorithme, celui-ci associe une valeur d'importance aux attributs pour déterminer ceux qui sont les plus significatifs selon Xgboost. Dans un deuxième temps, nous avons entraîné à nouveau l'algorithme, mais en retirant les caractéristiques qui avaient été identifiées comme étant les plus significatives.

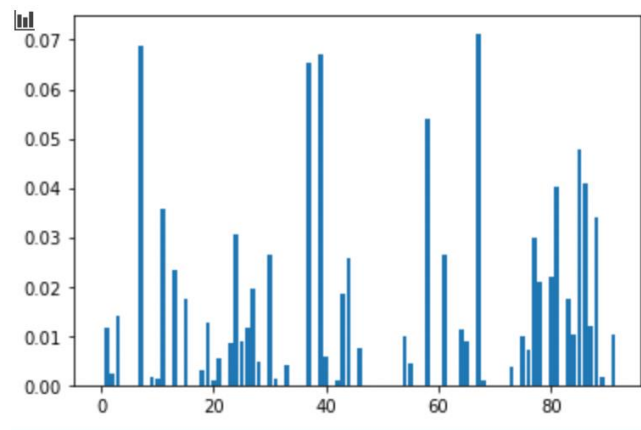


Figure 25 - Importance des attributs pour le groupe homme

Chez les hommes, les cinq meilleurs attributs sont les suivants. L'entraînement initial à T0 et les 69 attributs incluaient ceux-ci et donnent 81,25% de précision :

- Pabd6
- COAFFixedFlexionDuringStance
- CExternalTibialRotationInitialContact
- AbsProtPfv2
- score_koos_pain_t0

Par la suite, l'entraînement de Xgboost sans les attributs les plus significatifs fait ressortir d'autres attributs avec une importance significative chez les hommes :

- AbsPabdPfv2
- PabdPfv2
- COAFKneeVarusDuringStance_AllMethod * (3e méthode calc)
- AbsProtPfv3

- Pflex4

Le même exercice est effectué chez la femme afin d'évaluer l'impact des attributs considérés les plus significatifs.

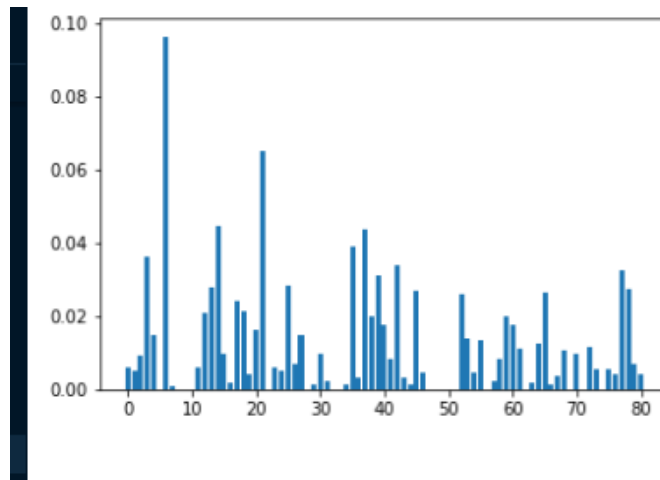


Figure 26 - Importance des attributs pour le groupe femme

Chez la femme, initialement, les facteurs les plus importants sont :

- COAFKneeVarusDuringStance_AllMethod * (2e calc)
- Prot1
- PFlex2
- COAFInternalRotationTibiaFemurLoading
- PabdPfv5

Comme mentionné, ce sont des attributs complètement différents que le groupe des hommes. Une fois retiré du jeu de données, on constate un lot de nouveaux attributs complètement modifiés :

- COAFDecreaseSagittalPlaneRangeMotion
- COAFKneeVarusInitialContact

- COAFVarusThrustDuringLoading_AllMethod * (3e méthode de calc)
- AbsProtPfv2MinusProtPfv1
- COAFValgusThrustDuringLoading_AllMethod * (1e méthode de calc)

On peut constater une très grande disparité entre les deux ensembles d'attributs les plus importants. Même si chez les hommes la disparité est présente, on peut voir une certaine corrélation dans le type d'attribut présent, par exemple : AbsProtPfv2 et AbsProtPfv3

Le jeu de données complet, hommes et femmes, donne des différences importantes.

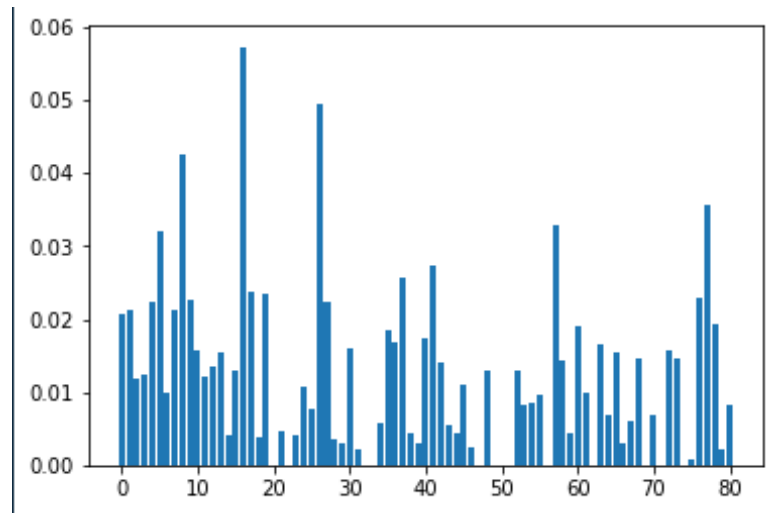


Figure 27 - Importance des caractéristiques pour le jeu de données homme-femme

Les attributs importants pour l'ensemble du jeu de données sont complètement différents. Ce qui démontre une certaine disparité entre le groupe des hommes et celui des femmes. L'algorithme doit

donc apprendre d'autres attributs significatifs pour les deux ensembles :

- COAFFixedFlexionDuringStance
- Pabd1
- COAFValgusThrustDuringLoading_AllMethod * (1 Méthodes de calc)
- score_koos_pain_t0
- AbsProtPfv1

Une fois les attributs importants retirés, il est intéressant de constater que les attributs sont un mélange très près de ce que nous retrouvons précédemment.

- Pflex2
- AbsPflex3
- COAFValgusThrustDuringLoading_AllMethod * (2 Méthodes de calc)
- COAFVarusThrustDuringLoading_AllMethod * (3 Méthodes de calc)
- score_koos_adl_t0

Selon les résultats de recherche [26], l'arbre de décision débute par le score_koos_pain, ce qui démontre l'importance du niveau de douleur [26] pour classifier les patients. L'importance dans l'arbre de décision, basé sur l'indice de Gini, n'a pas la même importance que l'arbre XGBOOST, qui effectue son approche différemment. Dans

l'ensemble le score_koos_pain n'est pas particulièrement plus significatif qu'un autre attribut.

Dans le tableau 1 de résultats, on observe que l'entraînement sans les cinq attributs les plus importants n'affecte pas les résultats de façon significative.

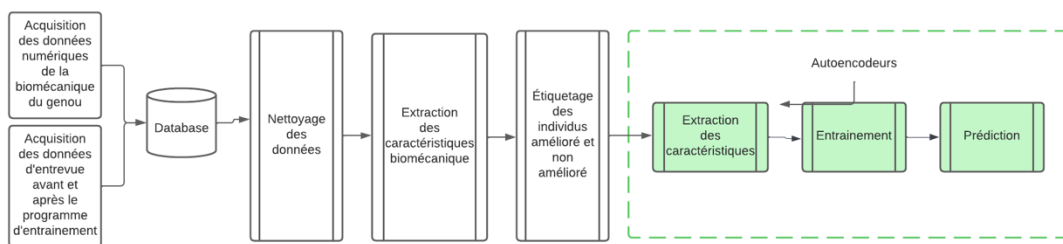
L'impact négligé sur les résultats en retirant les attributs avec le plus d'importance nous amène à regarder si les attributs importants selon l'algorithme sont les mêmes selon le sexe. Nous pouvons constater dans le tableau 2 qu'il n'y a pas de correspondance sauf pour le koos_pain à T0, qui se veut une donnée subjective et non un signal.

À l'exception du score_koos_pain, toutes les valeurs proviennent de calcul basé sur les données brutes. Ce qui soulève au moins quelques questionnements :

1. À quel niveau la différence physiologique de la femme et de l'homme peut complexifier et altérer la précision de l'algorithme ?
2. Considérant les attributs de valeur complètement distincts selon le sexe, quel est l'impact sur la généralisation de l'algorithme ?
3. Est-ce qu'avec le jeu de données complet, hommes-femmes, l'algorithme se repose sur des attributs moins significatifs, ce qui aurait un impact sur la généralisation de XGBOOST ?

5.4. Apprentissage profond

Dans cette section, nous expliquons les résultats obtenus avec la mise en place du système d'extraction des caractéristiques afin d'extraire les caractéristiques de valeurs pour fournir un nouveau jeu de données à un système de classification traditionnel.



Les tableaux qui suivent représentent les taux de classification d'algorithme traditionnel avec un ensemble d'attributs réduit. L'ensemble d'attributs réduit a été généré par un réseau de neurones de type auto-encodeur.

À la suite de l'expérimentation avec XGBOOST, plusieurs constats sur l'importance des attributs sont soulevés. L'utilisation d'un réseau de neurones afin d'extraire les caractéristiques significatives est une approche intéressante, mais moins intuitive.

Deux approches ont été expérimentées avec le jeu donné sans les données d'évaluation avant les exercices et avec les données d'évaluations.

La première approche consiste en un seul auto-encodeur afin de réduire la dimension. Comme nous constatons que les caractéristiques significatives sont différentes selon le sexe, la

deuxième approche utilise le jeu de données afin de faire ressortir les caractéristiques des individus qui se sont améliorés et non améliorés de façon individuelle pour ensuite les unifier et utiliser un algorithme traditionnel pour faire l'apprentissage.

Cette approche n'était pas adéquate avec XGBOOST puisque l'algorithme sert à faire une prédiction et non exclusivement une extraction de caractéristiques.

5.4.1. Analyse des résultats – Auto-encodeur unique

La première approche consiste à un seul auto-encodeur. Les tableaux suivants résument les taux de classification avec et sans les données d'évaluation initiales avec le programme d'exercices. Le tableau contient le pourcentage et l'écart-type (%/écart-type). Les processus d'évaluation ont été effectués avec un 10-fold. Le tableau exprime le meilleur résultat obtenu.

	Arbre de décision			XGBOOST		
Nombre d'attributs extraits	Homme	Femme	Homme-Femme	Homme	Femme	Homme-Femme
5	0.50/0.19	0.52/0.10	0.55/0.11	0.44 / 21	0.589 / 0.09	0.50/0.06
10	0.48/0.22	0.503/0.11	0.52/0.08	0.653/0.10	0.55/0.05	0.54/0.10
15	0.525/0.16	0.584/0.13	0.569/0.08	0.51/0.14	0.609/0.08	0.538/0.08
20	0.655/0.13	0.56/0.14	0.50/0.07	0.5373/0.17	0.594/0.12	0.54/0.08

Tableau 4 - Résultats – auto-encodeur unique

Nombre d'attributs extraits	Arbre de décision			XGBOOST		
	Homme	Femme	Tout	Homme	Femme	Homme-Femme
5	0.55/0.12	0.559/0.09	0.53/0.09	0.5875/0.15	0.612/0.12	0.55/0.06
10	0.62/0.14	0.627/0.08	0.54/0.12	0.601/0.16	0.57/0.09	0.53/0.18
15	0.54/0.12	0.51/0.09	0.61/0.11	0.53/0.18	0.601/0.11	0.61/0.11
20	0.742/0.20	0.579/0.11	0.57/0.08	0.669/0.22	0.615/0.13	0.56/0.08

Tableau 5 - Résultat – auto-encodeur unique avec données à T0

Les résultats ne démontrent pas des taux de classification intéressants. Même si l'auto-encodeur effectuait un bon encodage, on n'arrive pas à effectuer une bonne classification. Nous pouvons constater une légère amélioration avec les données d'évaluation par un spécialiste. Les données du spécialiste sont des données subjectives récoltées par un questionnaire avec le patient.

5.4.2. Analyse des résultats – Auto-encodeur par classe

Le processus mis en place permet d'extraire les caractéristiques distinctes de chaque classe. On sépare les individus améliorés et non améliorés afin de former deux jeux de données. Par la suite, un auto-encodeur distinct est entraîné afin de ressortir les caractéristiques de chaque classe distincte.

Afin d'effectuer l'entraînement du modèle et l'extraction des caractéristiques, nous unifions les deux encodeurs de classe distincte afin d'extraire pour chaque individu les caractéristiques pour ensuite entraîner un nouvel auto-encodeur qui servira à extraire les

caractéristiques globales qui ont le plus de valeur afin d'entraîner un algorithme conventionnel.

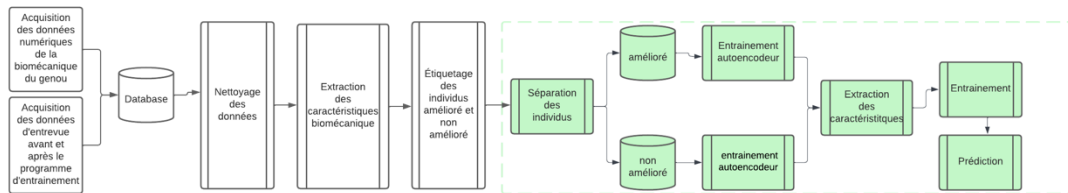


Figure 28 - Processus de construction des données pour l'entraînement

L'utilisation d'un auto-encodeur par classe démontre une amélioration de la généralisation. Les résultats sont significatifs avec le jeu de données qui contient les données d'évaluation à T0.

	Arbre de décision			XGBOOST		
pj159_autoencoder_V2_69_features-demo-2encodeur_extraction						
Nombre de feature extraites	Homme	Femme	Homme-Femme	Homme	Femme	Homme-Femme
5	0.56/0.11	0.57/0.16	0.48/0.11	0.55/0.15	0.49/0.129	0.49/0.07
10	0.53/0.17	0.56/0.129	0.54/0.11	0.55/0.159	0.636/0.07	0.66/0.11
15	0.655/0.176	0.59/0.15	0.64/0.11	0.60/0.08	0.62/0.1655	0.773/0.08
20	0.58/0.15	0.615/0.15	0.637/0.10	0.58/0.15	0.663/0.10	0.787/0.07

Tableau 6 - Résultat – auto-encodeur double avec les 69 attributs

	Arbre de décision			XGBOOST		
pj159_autoencoder_V2_69_features-T0-2encodeur_extraction						
Nombre de feature extraites	Homme	Femme	Homme-Femme	Homme	Femme	Homme-Femme
5	0.59/0.17	0.61/0.09	0.65/0.08	0.6/0.178	0.62/0.10	0.65/0.11
10	0.70/0.07	0.61/0.11	0.61/0.09	0.70/0.11	0.65/0.11	0.755/0.09
15	0.68/0.17	0.62/0.10	0.60/0.04	0.73/0.14	0.705/0.12	0.70/0.08
20	0.74/0.15	0.62/0.11	0.68/0.13	0.70/0.15	0.72/0.09	0.81/0.09

Tableau 7 - Résultat – auto-encodeur double avec données à T0

Les deux algorithmes montrent une amélioration significative par rapport à l'utilisation du jeu de données du tableau 6. Par contre, l'arbre de décision ne semble pas généraliser très bien avec les données d'évaluation à T0. En contrepartie, l'amélioration est significative pour l'algorithme XGBOOST. On peut noter une généralisation intéressante. En effet, que ce soit avec le jeu de données hommes, femmes ou complet, la classification offre une précision intéressante. Cela dit, l'écart-type est significativement plus bas avec le jeu de données complet.

De plus, le taux de classification ainsi que l'écart-type est similaire, peu importe le nombre caractéristique, extrait par l'auto-encodeur pour entraîner les algorithmes traditionnels.

6. CONCLUSION

L'objectif de la recherche était de développer de nouvelles méthodes pour la classification de signaux biomécaniques pour la prédiction de l'impact d'un programme de réadaptation physique. L'utilisation d'arbre de décision offre plusieurs avantages, mais comporte certaines limitations lorsque les données ne sont pas linéaires et très complexes. Afin d'approfondir la compréhension des signaux, il est nécessaire de comprendre la valeur qu'apporte chaque caractéristique à l'apprentissage des algorithmes. De plus, il sera nécessaire d'optimiser la façon dont les algorithmes évaluent l'importance relative des caractéristiques d'un jeu de données.

Au-delà des techniques d'apprentissage machine, le problème de classification non linéaire avec des signaux complexes demande des algorithmes en mesure de déterminer l'importance de chacun afin d'effectuer un apprentissage qui permettra une généralisation suffisamment importante pour une utilisation scientifique.

Pour atteindre cet objectif, la première exploration a été l'utilisation de XGBOOST, un algorithme basé sur le gradient avec une méthode innovatrice pour déterminer l'importance des attributs comparativement à un arbre de décision traditionnel. Nous pouvons constater de bons résultats, mais un jeu de données plus important permettrait de mieux évaluer la généralisation. De plus, la cotation des attributs relève plusieurs questionnements sur la construction des arbres de décision sous-jacents dans XGBOOST. L'attribut qui a la meilleure entropie n'est pas nécessairement celui qui permettra de générer le meilleur arbre. Quelle est la meilleure méthode pour

déterminer l'importance des attributs ? Quel est l'impact de l'utilisation des attributs considéré le plus important sur l'importance des autres attributs et la construction des arbres ? Nous avons constaté que l'absence des attributs les plus importants avait eu peu d'influence sur le résultat.

La deuxième approche a été d'utiliser un réseau de neurones afin d'extraire les caractéristiques importantes. Malgré cette technique, les données restent non linéaires et l'arbre de décision ne procure pas un taux de classification satisfaisant. À l'inverse, l'algorithme XGBOOST a offert un taux et une généralisation intéressante. En revanche, il faut noter que cette réussite se veut avec des données subjectives d'évaluation et non seulement à l'aide des signaux.

La détermination des caractéristiques significatives dans les signaux biomécaniques nécessite d'approfondir diverses techniques. La capacité d'utiliser une méthode différente pour construire les arbres de décision tout en conservant leur interprétabilité reste à explorer. L'amélioration apportée avec XGBOOST est une piste de solution.

De plus, les réseaux de neurones ont démontré une capacité à extraire des caractéristiques significatives. Il est actuellement impossible de comprendre pourquoi ou sur quelle caractéristique de base le réseau s'appuie pour déterminer la nouvelle caractéristique.

Les auto-encodeurs pourraient faire partie de la solution pour comprendre davantage les données. À l'aide de cette compréhension, nous pourrions passer de l'extraction de caractéristiques à la sélection de caractéristiques significative afin de construire des arbres de décision extrêmement robustes.

BIBLIOGRAPHIE

- [1]. Danault D., Dupontgand S. (2018) , Certification en conditionnement physique, Anatomie [Manuel de l'élève]
- [2]. Goodfellow,I., Bendio, J., Courville A. (2018), Deep learning, , consulté octobre 2020
- [3]. Coursera (2021) , Spécialisation en apprentissage profond de <https://www.coursera.org/>
- [4]. Bordet,C. consulté en 2022, Comprende la descente du gradient, <https://www.charlesbordet.com/fr/gradient-descent/>
- [5]. Starmer,J(2018),Bias,<https://www.youtube.com/watch?v=EuBBz3bI-aA> – consulté le 19 décembre 2021
- [6]. Kokkinos, I. (2016),UberNet : Training a 'Universal' Convolutional Neural Network for Low, Mid and High-Level Vision using Diverse Datasets and Limited Memory, <https://arxiv.org/abs/1609.02132>, consulté le 19 décembre 2021
- [7]. He, K., Zhang,X., Ren, S. Sun, J. (2015),Delving Deep into Rectifiers :Surpassing Human-Level Performance on ImageNet Classification, <https://arxiv.org/pdf/1502.01852.pdf>, consulté en octobre 2020
- [8]. Larochelle, H.(2013), Neural Network : Feedforward, <https://www.youtube.com/watch?v=SGZ6BttHMPw&feature=youtu.be&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH>
- [9]. Larochelle, H.(2013),Neural Network :Feedforward – activation function, <https://www.youtube.com/watch?v=tCHIkgWZLOQ&feature=youtu.be&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH>
- [10]. Larochelle, H.(2013), Neural Network : Empirical risk minimization, <https://www.youtube.com/watch?v=5adNQvSIF50&feature=youtu.be&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH>
- [11]. Larochelle,H.(2013), Neural Network : Loss function, <https://www.youtube.com/watch?v=PpFTODTztsU&feature=youtu.be&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH>
- [12]. Larochelle, H.(2013),Neural Network :Backpropagation, https://www.youtube.com/watch?v=_KoWTD8T45Q&feature=youtu.be&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH

- [13]. Alex(2016),Lecture 4 Backpropagation, Neural Networks
<https://www.youtube.com/watch?v=59Hbtz7XgjM>
- [14]. Siraj, R. (2017),Backpropagation in 5 Minutes (tutorial),
<https://www.youtube.com/watch?v=q555kfIFUCM>
- [15]. RimstarOrg (2016),Backpropagation Neural Network -
 How it Works e.g. Counting,
<https://www.youtube.com/watch?v=WZDMNM36PsM>
- [16]. Welch Labs (2014),Neural Networks Demystified [Part 4
 : Backpropagation], <https://www.youtube.com/watch?v=GlcnxUlrtk>
- [17]. Y.Ouakrim, 2011, Y.Ouakrim (Octobre 2011).
 Classification de sujets asymptomatiques et gonarthrosiques en
 fonction des données cinématiques : Comparaison de l'approche
 globale et de l'approche locale. Master's thesis, École De
 Technologie Supérieur.
- [18]. Centre du genou, 2021,
<http://www.centredugenou.com/douleur-au-genou/anatomie-et-fonction-du-genou/>
- [19]. <https://explained.ai/gradient-boosting/L2-loss.html>
- [20]. <https://www.bebecardio.com/blogue/douleurs-genoux-maternite>
- [21]. <http://www.centredugenou.com/douleur-au-genou/anatomie-et-fonction-du-genou/>
- [22]. <http://proceedings.mlr.press/v97/balin19a/balin19a.pdf>
- [23]. Hinton, G. E., and Salakhutdinov, R. R. 2006. Reducing
 the dimensionality of data with neural networks.
- [24]. L'arthrose au Canada, <https://www.canada.ca/fr/sante-publique/services/publications/maladies-et-affections/arthrose.html>
- [25]. Modèle xgboost
<https://xgboost.readthedocs.io/en/latest/tutorials/model.html>
- [26]. M.Mezghani, 2019. Étude de l'effet d'un programme
 d'exercices physiques par des méthodes d'apprentissage
 machine, Unisversité Teluq.
- [27]. Erreur quadratique moyenne,
<https://explained.ai/gradient-boosting/L2-loss.html>

- [28]. [Courbe de l'entropie, https://pic4.zhimg.com/v2-be5deac9e0a9e68572f76a06ac321e2e_r.jpg](https://pic4.zhimg.com/v2-be5deac9e0a9e68572f76a06ac321e2e_r.jpg)
- [29]. [Formation axis, Manuel de l'étudiant](#)
- [30]. Anatomy and physiology, <https://assets.openstax.org/oscms-prodcms/media/documents/AnatomyandPhysiology-OP.pdf>, p.329, 332
- [31]. Biomécanique de la marche, partie 1, https://www.youtube.com/watch?v=RAE_RpfGeUc&t=1s
- [32]. Fonction sigmoïde, https://www.researchgate.net/figure/Standard-sigmoid-function_fig2_336275807
- [33]. Tangente hyperbolique, https://fr.wikipedia.org/wiki/Tangente_hyperbolique
- [34]. Fonction reLu, https://www.researchgate.net/figure/The-plot-of-the-ReLU-function_fig5_335540811
- [35]. LReLU, [http://www.gabormelli.com/RKB/Leaky_Rectified_Linear_Activation_\(LReLU\)_Function](http://www.gabormelli.com/RKB/Leaky_Rectified_Linear_Activation_(LReLU)_Function)