

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

**PRÉVISION À LONG TERME DU DÉVELOPPEMENT DES ATHLÈTES
DE BASEBALL À L'AIDE DE L'INTELLIGENCE ARTIFICIELLE
DANS UN ENVIRONNEMENT DE L'INFORMATIQUE QUANTIQUE**

**MÉMOIRE PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN GÉNIE ÉLECTRIQUE**

**PAR
BAOWENDE STEPHANE LEANDRE OUEDRAOGO**

MARS 2025

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

UNIVERSITÉ DU QUÉBEC
UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES
MAÎTRISE EN GÉNIE ÉLECTRIQUE (M. SC. A.)

Direction de recherche :

Prof. Adam W. Skorek

Directeur de recherche

Jury d'évaluation

Prof. Adam W. Skorek

Évaluateur interne

Prof. Ahmed Lakhssassi

Évaluateur externe

Université du Québec à Outaouais

Prof. Alben Cardenas

Évaluateur interne

Université du Québec à Trois-Rivières

REMERCIEMENT

La réalisation de ce document n'a été possible qu'avec les concours multiformes de bon nombre de personnes. Nous tenons à leur exprimer toute notre gratitude.

Nos remerciements vont à :

Monsieur Adam W. Skorek, notre professeur ;

Tout le corps enseignant et le personnel administratif de l'UQTR, pour leur appui à notre formation ;

À Mitacs Accélération, pour la confiance et le soutien dont ils ont fait preuve ;

À LANEK Solutions, pour son intérêt.

Je tiens à remercier mes parents, mes frères et sœurs pour leur aide et leurs conseils ;

Je remercie aussi tous mes amis qui m'ont aidé dans la rédaction de ce document.

RÉSUMÉ

L'objectif de ce projet est de créer un système de calcul fondé sur l'informatique quantique qui permettra de prévoir les performances des athlètes de baseball. Notre objectif est de prévoir leur développement à long terme. Nous avons remarqué qu'en matière de choix des meilleurs joueurs de baseball du Québec, l'état de l'art actuel se fait généralement de manière subjective. De plus, il y a beaucoup de jeunes athlètes qui sont passionnés par le baseball, mais qui n'ont pas les moyens de s'inscrire dans un programme pour avoir une chance d'être sélectionnés.

Dans le cadre de cette recherche, nous avons rassemblé et évalué des données, en plus d'avoir examiné diverses techniques de prévision traditionnelles, telles que le LSTM et l'ARIMA. Nous avons établi que le système quantique mis en place (QLSTM) s'avère plus performant en matière de prévisions que les méthodes courantes.

TABLE DES MATIÈRES

REMERCIEMENT	iii
RÉSUMÉ	iv
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX.....	viii
LISTE DES FIGURES.....	ix
LISTE D'ABRÉVIATIONS	xiii
Chapitre 1 - INTRODUCTION	1
1.1 But de la recherche	2
1.2 Contribution de la recherche	2
1.3 Structure du document.....	3
Chapitre 2 - TRAVAUX DE RECHERCHE CONNEXES.....	4
2.1 Baseball	4
2.1.1 Définition	4
2.1.2 Équipements de jeu	6
2.2 Informatique quantique	15
2.2.1 Bits quantiques ou qubits	18
2.2.2 Quelques portes quantiques.	20
2.3 Conclusion partielle.....	26
Chapitre 3 - SYSTÈME DE PRÉVISION	28

3.1	Traitement de données.....	29
3.1.1	Cueillette de données	29
3.1.2	Analyse de données.....	31
3.2	Systèmes de prévision	31
3.2.1	Réseaux de neurones artificiels.....	32
3.2.2	Réseaux de neurones récurrents.....	34
3.2.3	Modèle LSTM.....	35
3.2.4	Modèle ARIMA	37
3.2.5	Quantum LSTM.....	39
3.3	Conclusion partielle.....	41
Chapitre 4 - MODÉLISATION ET RÉSULTATS		42
4.1	Modélisation du système quantique	43
4.1.1	Prétraitement des données.....	43
4.1.2	Modèle QLSTM.....	53
4.2	Résultats et analyses des modèles QLSTM, LSTM et ARIMA.....	66
4.3	Conclusion partielle.....	68
Chapitre 5 - CONCLUSION.....		70
BIBLIOGRAPHIE		71
ANNEXE A – DIAGRAMME DE TERRAIN IMPRIMABLE		76

ANNEXE B : EXIGENCE CHAMPIONNATS NATIONAUX 2024.....	78
ANNEXE C : ENSEMBLES D'ÉTAPES INTUITIFS ET REPRODUCTIBLES POUR LES FLUX DE TRAVAIL QUANTIQUES	79

LISTE DES TABLEAUX

Tableau 2.1: Tableau résumant la mesure des qubits.....	26
Tableau 3.1: Exemple de fiche de récolte	30

LISTE DES FIGURES

Figure 2.1 : Terrain de baseball ainsi que les espaces réservés à chaque joueur (source : nike.com).....	5
Figure 2.2: Ordinateur Quantique (Source: IBM Canada).....	16
Figure 2.3: Représentation des modèles de particules (source : Micro-stage-IMACA)	17
Figure 2.4: Schémas vectoriel d'état quantique (Élaboration personnelle réalisée avec Microsoft PowerPoint).....	19
Figure 2.5: Représentation de la porte H sur un circuit quantique et l'état résultant dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).....	21
Figure 2.6: Représentation de la porte X sur un circuit quantique et l'état résultant dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).....	22
Figure 2.7: Représentation de la port CCNOT avec la porte H sur un circuit quantique et les états résultants dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).	23
Figure 2.8: Représentation de la porte Z sur un circuit quantique et l'état résultant dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).....	24

Figure 2.9: Exemples de portes sur un circuit quantique (générée avec Qiskit Composer développé dans cette étude).	25
Figure 2.10: Mesure d'un qubit dans une superposition (source : Micro- stage- IMACA).....	25
Figure 3.1 : Processus de traitement des données, intégrant une architecture quantique (Élaboration personnelle réalisée avec Microsoft PowerPoint).	28
Figure 3.2: Schéma d'une partie d'un réseau de neurones artificiels (ANN) (Élaboration personnelle réalisée avec Microsoft PowerPoint).	33
Figure 3.3: Schéma d'une partie d'un réseau de neurones récurrent (RNN) (Élaboration personnelle réalisée avec Microsoft PowerPoint).	34
Figure 3.4: Schéma d'une cellule LSTM (Source : [24] © [2023] IEEE).....	36
Figure 3.5: Schéma d'une cellule QLSTM (Source : Samuel Yen-Chi Chen, © [2022], IEEE)	39
Figure 3.6: Circuit d'implémentation des VQC (source : IBM Quantum Learning Platform)	40
Figure 4.1: Architecture de travail (Élaboration personnelle réalisée avec Microsoft PowerPoint).	42
Figure 4.2: Niveau de corrélation (Élaboration personnelle réalisée avec Microsoft PowerPoint).	44

Figure 4.3: Matrice de corrélation (générée avec un script Python développé dans cette étude).	45
Figure 4.4: Comparaison des valeurs normalisées OPS avant et après traitement (générée avec un script Python développé dans cette étude).....	46
Figure 4.5: Matrice de corrélation après traitement (générée avec un script Python développé dans cette étude).	48
Figure 4.6: Dendrogramme des variables après traitement (générée avec un script Python développé dans cette étude).....	49
Figure 4.7: Séquences anormales (générée avec un script Python développé dans cette étude).	51
Figure 4.8: Séquences lissées (générée avec un script Python développé dans cette étude).....	52
Figure 4.9: Représentation textuelle et graphique d'un circuit quantique (générée avec un script Python et PennyLane développé dans cette étude).....	56
Figure 4.10: Historique de l'optimisation (générée avec un script Python développé dans cette étude).....	58
Figure 4.11: Importances des hyperparamètres (générée avec un script Python développé dans cette étude).	58
Figure 4.12: Courbes d'apprentissage MSE (générée avec un script Python développé dans cette étude).....	61

- Figure 4.13: Comparaison des prévisions QLSTM avec les valeurs réelles
(générée avec un script Python développé dans cette étude). 62
- Figure 4.14: Prévisions futures vs données historiques (avec mise en
évidence de la prévision pour le 1er janvier 2025) (générée
avec un script Python développé dans cette étude). 64
- Figure 4.15: Comparaison des prévisions LSTM avec les valeurs réelles
(générée avec un script Python développé dans cette étude). 66
- Figure 4.16: Comparaison des prévisions ARIMA avec les valeurs réelles
(générée avec un script Python développé dans cette étude). 67

LISTE D'ABRÉVIATIONS

ANN	Artificial Neuronal Network
ARIMA	AutoRegressive Integrated Moving Average
CCNOT	Porte CCNOT
CNOT	Porte CNOT
CSA	Cross-Sectional Area
H	Hadamard
I	Porte Identité
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
NISQ	Noisy Intermediate-Scale Quantum
NOT	Porte NOT
OPS	On-base Plus Slugging
QLSTM	Quantum Long Short-Term Memory
QSVM	Quantum Support Vector Machine
R	Runs
R^2	Coefficient of Determination
RMSE	Root Mean Squared Error
RNN	Recurrent Neuronal Network
SWAP	Porte SWAP
TB	Total Bases

UQTR	Université du Québec à Trois Rivières
VQC	Variational Quantum Circuits
WPA	Win Probability Added
X	Porte Pauli X
Z	Porte Pauli Z
VS	Versus

Chapitre 1 - INTRODUCTION

Le baseball est l'un des sports les plus populaires et les plus appréciés en Amérique du Nord. C'est un sport d'équipe où chacun a son rôle à jouer. Il est à la fois explosif et subtil. C'est un sport qui exige beaucoup de technique et de stratégie. Pour maîtriser l'art de la pratique du baseball, il faut du temps, de la répétition, de l'engagement et un don de soi. Des sportifs ont commencé à le pratiquer vers 1840, en établissant des règles en 1845. Il est clair qu'il occupe une place importante dans la société, notamment américaine. Au-delà d'être un sport de loisirs, le baseball est aussi très compétitif, complexe et constitue une source de revenus pour plusieurs personnes, dont les joueurs, les entraîneurs, les propriétaires d'équipes et même les parieurs.

Dans notre étude, nous nous pencherons plus particulièrement sur les athlètes de baseball, notamment sur leur processus de sélection et de développement. Nous chercherons à répondre à cette question : « Comment peut-on prévoir le développement à long terme d'un athlète de baseball ? » Nous allons répondre à cette question avec l'aide de l'intelligence artificielle.

En effet, l'intelligence artificielle est une discipline qui vise à créer des machines capables de penser comme des êtres humains.

À travers plusieurs études déjà réalisées, on constate que l'informatique quantique représente une avancée majeure dans le domaine de la technologie et de l'information. L'informatique quantique s'appuie sur les lois de la mécanique quantique pour résoudre

des problèmes complexes. Ce projet vise à explorer l'application de l'intelligence artificielle au baseball dans un contexte d'informatique quantique.

1.1 But de la recherche

Dans un monde où les avancées technologiques révolutionnent de nombreux domaines, le sport de haut niveau ne fait pas exception. Cette recherche s'inscrit dans une approche novatrice qui vise à combiner l'informatique quantique et l'intelligence artificielle pour prévoir les performances des joueurs de baseball. L'objectif est d'abord d'utiliser des algorithmes d'apprentissage automatique avancés pour analyser des données, identifier des schémas complexes et fournir des prévisions fiables. Ensuite, il s'agit de les comparer à des modèles classiques.

1.2 Contribution de la recherche

La recherche présente plusieurs avantages, dont les suivants :

- Amélioration de l'analyse : Grâce à l'informatique quantique, cette recherche permet des prévisions plus précises sur les performances sportives. Cela permet aux équipes et aux entraîneurs de prendre des décisions stratégiques éclairées en se basant sur des données.
- Convergence technologique révolutionnaire : La combinaison audacieuse de l'intelligence artificielle et du calcul quantique dans un champ aussi intrigant que le baseball représente une percée interdisciplinaire sans précédent.

- Amélioration des performances athlétiques : En tirant parti de modèles de prévision sophistiqués, cette étude met en évidence des pistes pour optimiser les résultats des joueurs, que ce soit au niveau individuel ou collectif.
- Base pour d'autres disciplines sportives : Le système développé pourrait servir de modèle pour d'autres sports ou d'autres domaines nécessitant une analyse complexe des données.

1.3 Structure du document

Le contenu du document se compose de diverses parties, dont :

- Chapitre 2 : Études connexes

Ce chapitre couvre non seulement les aspects généraux du baseball et de l'informatique quantique, mais également une analyse des travaux antérieurs abordant la problématique.

- Chapitre 3 : Systèmes de prévision

Ce chapitre offre un aperçu des méthodes de prévision couramment utilisées ainsi qu'un résumé du projet.

- Chapitre 4 : modélisation et résultats

Ce chapitre se concentre sur la modélisation et l'analyse des résultats des modèles LSTM, ARIMA et QLSTM.

- Chapitre 5 : Conclusion générale

Ce dernier chapitre est une conclusion générale, et il est suivi de recommandations.

Chapitre 2 - TRAVAUX DE RECHERCHE CONNEXES

2.1 Baseball

2.1.1 *Définition*

Sur un terrain, on trouve une balle, une batte et quatre bases. Ces éléments sont les composantes du baseball, un sport qui remonte à des siècles, principalement dans un jeu anglais appelé « rounders ». Les Américains avaient déjà un jeu similaire avec des bases, connu sous le nom de « town ball », qui existait après la révolution. Dans les années 1840, des amateurs de sport dans plusieurs villes du Nord-Est se rassemblaient régulièrement pour jouer à différentes versions qu'ils appelaient "base ball". En 1845, Alexander Cartwright et Daniel « Doc » Adams, responsables du Knickerbocker Base Ball Club, ont établi les premières règles officielles du jeu. Bien que ces règles aient rapidement évolué au fil des ans, elles étaient déjà très proches de la forme actuelle du baseball, ce qui en fait la « naissance » de ce sport.

Voici un schéma d'un terrain de baseball avec les différentes positions occupées par les joueurs.

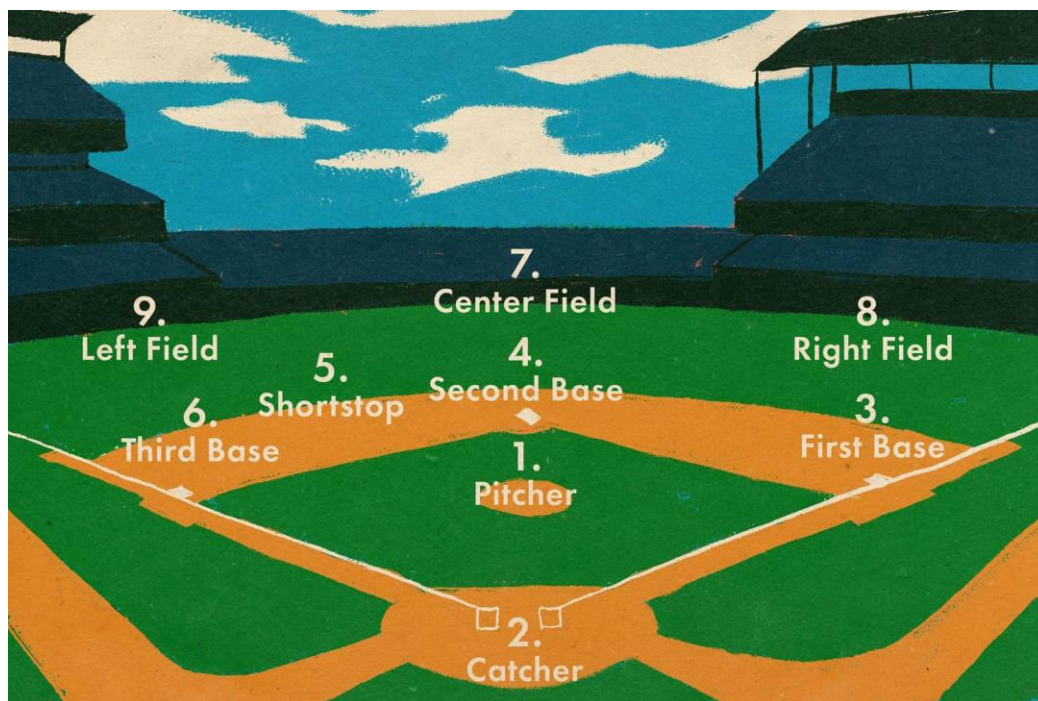


Figure 2.1 : Terrain de baseball ainsi que les espaces réservés à chaque joueur
(source : nike.com)

On observe dans la Figure 2.1 neuf postes de jeu :

- En 1 (Pitcher) : le lanceur
- En 2 (Catcher) : le receveur
- En 3, 4, 5, et 6 (respectivement : First Base, Second Base, Shortstop, et Third Base) : les joueurs de champs intérieurs.
- En 7, 8 et 9 (respectivement : Center Field, Right Field et Left Field) : les joueurs de champs extérieurs.

2.1.2 Équipements de jeu

2.1.2.1 Les lanceurs

Le monticule du lanceur est le cœur de l'univers du baseball. Dans un match de baseball, tout commence avec le lanceur qui effectue son premier lancer. C'est lui qui a pour tâche de faire sortir les frappeurs adverses, mais il y a une grande différence entre dire ça et le faire. Selon Andrew Stone, entraîneur adjoint en baseball à l'Université d'État du Michigan, « le lanceur est le seul joueur, toutes disciplines confondues, qui puisse contrôler le rythme du match » [1].

2.1.2.2 Les receveurs

Le joueur qui travaille le plus fort dans une équipe de baseball est le receveur. Il doit rester accroupi derrière le marbre pendant neuf manches, attraper tout ce que les lanceurs envoient aux frappeurs, supporter d'être frappé par des pointes et des bâtons, et être prêt à effectuer des lancers parfaits pour attraper les voleurs de bases potentiels.

Jouer au receveur dans une ligue de baseball junior exige autant d'efforts physiques et mentaux que la version professionnelle. Comme leurs modèles de la Major League, les jeunes receveurs portent tous les équipements de protection disponibles, y compris le masque, le casque, un protège-poitrine, des protège-tibias et une coque de protection. Parfois, les jeunes joueurs ont besoin d'un certain temps pour s'adapter à cet équipement, mais après quelques conseils sur les fautes, ils verront qu'il en vaut la peine.

Les receveurs communiquent avec les lanceurs en utilisant des signaux manuels pour indiquer quel lancer ils doivent effectuer. Chaque équipe développe son propre ensemble

de signaux, mais la liste standard est un doigt pour une balle rapide, deux pour une courbe, trois pour un changement et quatre pour tout autre lancer, tel qu'un glissement. Les signaux sont modifiés lorsque les coureurs sont sur les buts pour empêcher les frappeurs de les décoder et de les utiliser pour prévenir les lanceurs. Les receveurs peuvent aussi faire des signaux pour les lancements ou les interceptions.

2.1.2.3 Les joueurs de champs extérieurs et intérieurs

Une équipe de baseball est répartie en deux zones principales : l'intérieur du terrain et le champ extérieur. Les quatre joueurs qui se tiennent près des buts forment le champ intérieur, tandis que les trois autres joueurs, situés au-delà des buts, composent le champ extérieur. Le lanceur et le receveur sont officiellement considérés comme faisant partie du champ intérieur. Chacune des quatre positions, première base, seconde base, arrêt-court et troisième base, a une zone de responsabilité distincte et des compétences spécialisées dans cette zone. En revanche, les trois champs extérieurs, champ gauche, champ central et champ droit, sont plus semblables. Chacun d'entre eux couvre environ un tiers du terrain. Peu importe leur poste, tous ces athlètes ont un rôle à jouer lorsque la balle est en jeu : éliminer les coureurs et empêcher les coureurs de marquer.

Les joueurs de champ intérieur doivent rester bas et équilibrés, regarder la balle tomber dans leur gant, puis effectuer un lancer rapide et précis vers la base. Lorsque le batteur frappe la balle, ils se déplacent pour l'intercepter, la rattrapant avec la main qui lance.

Les athlètes adaptent leur position en fonction du frappeur ou de la situation du jeu. Quelques exemples : Face à un frappeur droitier, l'arrêt-court, le voltigeur de gauche, et le joueur de champ centre peuvent tous se déplacer vers la droite. Lorsqu'un coureur se trouve sur la troisième base et qu'il reste moins de deux retraits, les joueurs de l'intérieur (arrêt-court, deuxième but, premier but et receveur) jouent « à l'intérieur », c'est-à-dire près du marbre, sur la pelouse. Sur un jeu d'amorti, les joueurs de première et de troisième base se dirigent vers le marbre pour exécuter l'amorti et le lancer à une base.

2.1.2.4 Les batteurs

Les joueurs de baseball les plus redoutables au bâton ne réussissent qu'environ trois fois sur dix. Un basketteur avec ce taux de réussite serait au chômage ; un batteur de champ ne ferait pas partie de l'équipe de football. Cependant, atteindre un score de 0,300 est considéré comme une réussite remarquable. En l'espace d'une fraction de seconde, les batteurs doivent prendre la décision de s'élancer, commencer leur course, puis, une fraction de seconde plus tard, s'élancer une seconde fois. Encore une fraction de seconde plus tard, ils doivent réussir à frapper une balle sphérique qui tourne à toute vitesse et qui se dirige vers eux à une vitesse pouvant atteindre plus de 90 miles par heure. C'est difficile à réaliser, mais quand on y parvient, c'est particulièrement stimulant.

2.1.2.5 Les coureurs de base

Dès qu'un batteur atteint la première base, il devient un coureur de base. Être un bon coureur de base est presque aussi important que d'être un bon frappeur. Les coureurs

avancent d'une base à l'autre lorsque leurs coéquipiers mettent la balle en jeu. Ils peuvent également avancer en volant une base ou en profitant d'une balle passée ou d'un lancer sauvage, c'est-à-dire lorsqu'un lancer passe à travers le receveur. Chaque coureur de base a un objectif ultime : atteindre la plaque et marquer un point pour son équipe. Chaque lancer amène des variables différentes, ce qui rend la concentration tout aussi cruciale pour un coureur de fond que la vitesse et la technique.

Lorsqu'on court vers une base, glisser sur le sol est le seul moyen d'arrêter son élan et d'éviter de la dépasser. On doit garder la tête haute, la jambe supérieure pointée vers la base, et la jambe inférieure repliée en dessous.

2.1.2.6 Les arbitres

Le baseball est un sport qui fonctionne selon des règles strictes. Les personnes responsables de faire respecter ces règles sont appelées des arbitres. Les arbitres décident si un lancer est une balle ou une prise, si un coureur est sauf ou éliminé, et si une balle frappée est bonne ou fausse. Lors des matchs de la saison régulière de la Ligue majeure, il y a quatre arbitres, mais pour les séries éliminatoires et les World Series, il y en a six.

Les arbitres doivent avoir une vue dégagée et nette du marbre. L'arbitre du marbre se munira d'une petite brosse pour nettoyer la plaque de temps à autre. Les arbitres tourneront toujours le dos au terrain avant de se pencher et de dépoussiérer.

L'arbitre du marbre a la tâche la plus ardue de l'équipe d'arbitres. Il doit prendre des décisions instantanées pour déterminer si une balle est un strike, pour juger d'un amorti, s'il est juste ou faux, et pour prendre des décisions sur la base. Chaque arbitre de baseball

développe son propre style personnel pour annoncer les frappes. Certains sont subtils, d'autres sont bruyants et théâtraux.

La MLB (Majeur League baseball) fournit des statistiques sur les performances des athlètes, appuyant les textes mentionnés, pour chacun des joueurs par poste [2].

À travers un résumé de ces études, on verra que beaucoup d'énergie a été investie dans le suivi des athlètes afin qu'ils puissent obtenir de meilleures performances.

2.1.2.7 Impact des travaux sur le baseball

Les règles du baseball révèlent sa profondeur et sa subtilité. Plusieurs études ont été réalisées pour améliorer la qualité et la sécurité des athlètes. Comprendre la constitution du baseball et celle de ses athlètes nous aide à améliorer le fonctionnement de notre système de prévision.

Une étude portant sur quarante-sept (47) joueurs de baseball a été menée pour déterminer les caractéristiques de la masse du membre supérieur, de la CSA (Cross-Sectional Area) musculaire et de la raideur chez les joueurs de baseball adolescents ayant ou non subi une blessure au coude. Les variables anthropométriques et de performance sont importantes pour distinguer les joueurs de baseball professionnels à différents niveaux de compétition. Il établit un lien entre des attributs tels que le poids, la vitesse, la puissance du bas du corps et la force de préhension avec les performances spécifiques au baseball. Bien que les compétences spécifiques au sport demeurent cruciales pour la réussite, une étude met de l'avant l'importance croissante de la force du bas du corps, de la vitesse et de l'agilité dans la détermination des performances au baseball. Toutefois, la

contribution de la force de préhension semble moins significative, éclipsée par l'influence d'autres attributs physiques [3],[4].

L'âge a un impact sur les performances physiques des joueurs de baseball professionnels, ce qui révèle des informations importantes. Des études antérieures avaient établi un lien entre la force, la puissance et la vitesse et les performances sur le terrain. Cependant, cette étude met l'accent sur l'âge comme facteur crucial dans le maintien de ces attributs. Notamment, cela indique que la puissance du bas du corps se maintient jusqu'à la fin de la vingtaine et au début de la trentaine, tandis que la vitesse, l'agilité et la force de préhension persistent chez les joueurs de plus de 35 ans. Les résultats montrent des tendances divergentes entre les joueurs de position et les lanceurs, suggérant l'influence de compétences de résumé sur ces derniers. De plus, la masse corporelle, la masse maigre et la force de préhension évoluent avec l'âge, ce qui affecte les performances des athlètes et la durée de leur carrière. Cette étude fournit des informations précieuses sur les défis uniques du vieillissement chez les athlètes professionnels, soulignant l'importance de poursuivre des recherches longitudinales dans ce domaine [5], [6], [7].

Il va sans dire que chaque athlète de baseball a besoin d'un entraînement spécifique correspondant à son poste de jeux. Il faut donc un système capable de s'adapter à chaque athlète pour qu'il soit efficace.

Explorons maintenant le monde captivant des systèmes de prévision dans le domaine du sport, où les bâtons et les balles de baseball ne sont pas seulement des objets physiques, mais également des sujets d'étude et de prévision.

2.1.2.8 Impact des travaux sur l'intelligence artificielle

L'étude entre l'intelligence artificielle et l'apprentissage traditionnel montre qu'il y a une utilisation inefficace de ces deux méthodes en éducation. L'étude suggère d'employer des fonctions de prévision des données pour résoudre des problèmes complexes et améliorer les systèmes. Cette expérience se concentre sur l'utilisation de la fonction de contrôle pour classer automatiquement les données, en particulier dans le domaine du sport. L'objectif est de créer un système d'analyse des résultats passés des athlètes pour optimiser les programmes d'entraînement physique. En comparaison avec les méthodes de test conventionnelles, l'approche expérimentale avec un modèle de calcul d'ensemble permet d'améliorer considérablement la précision prévisionnelle et de réduire les erreurs et le temps nécessaire [8].

En 2019, les modèles LSTM surpassent d'autres méthodes, car ils permettent de réaliser des prévisions plus précises. Ils sont considérés comme une approche prometteuse pour améliorer les performances des joueurs de baseball, pouvant devenir une nouvelle référence dans ce domaine [9]. Une recherche met en évidence un cadre conceptuel hiérarchique qui explore le comportement de l'utilisateur à travers diverses actions, activités, comportements intrinsèques à une tâche et interactions entre celles-ci. Une architecture d'apprentissage profond basée sur les LSTM est proposée pour modéliser le comportement interactif. Ce modèle probabiliste permet de prédire les actions de l'utilisateur et de détecter les comportements anormaux. L'évaluation implique plusieurs architectures avec différents nombres de prévisions d'actions [10].

ARIMA (Autoregressive Integrated Moving Average) est une méthode couramment utilisée pour analyser des séries chronologiques et faire des prévisions. Cela consiste à déterminer les valeurs futures d'une série en fonction des données du passé. Elle combine deux approches distinctes : les modèles autorégressifs (AR) et les modèles à moyenne mobile (MA) [11]. Cette combinaison permet d'intégrer leurs forces respectives pour améliorer la précision des prévisions. Ce modèle est utilisé dans divers projets de prévision, notamment dans le domaine du sport. L'association des modèles ARIMA et LSTM constitue une méthode novatrice dans l'analyse et la prévision de séries chronologiques. Bien que le modèle ARIMA soit réputé pour sa capacité à saisir les tendances linéaires et les schémas saisonniers dans les données, le réseau neuronal LSTM se distingue par sa capacité à modéliser des relations complexes et non linéaires grâce à son architecture sophistiquée. En combinant les avantages de ces deux techniques complémentaires, il est possible de concevoir un système prédictif fiable, capable de traiter les composantes simples et complexes des séries temporelles. Cette synergie offre des perspectives prometteuses pour des applications variées, allant de la prévision des ventes à la détection d'anomalies dans des systèmes dynamiques, en passant par l'analyse des signaux physiologiques [12]. Dans une étude visant à comparer les performances des modèles ARIMA et LSTM pour la prévision des séries chronologiques, on a mis en évidence leurs forces et leurs faiblesses. Alors que l'ARIMA, très répandu dans les domaines de l'économie et des affaires, est restreint par sa supposée linéarité et son exigence minimale de points de données, le LSTM se démarque par sa capacité à gérer des données hétérogènes [13].

Toutefois, dans l'immensité du paysage de l'innovation, une nouvelle frontière s'ouvre avec l'avènement des systèmes de prévision quantique, proposant une perspective captivante et révolutionnaire dans le domaine de la prévision. L'étude démontre l'utilisation d'une méthode de prévision du comportement fondée sur l'informatique quantique (QSVM) qui présente une amélioration de la précision par rapport aux algorithmes traditionnels dans un émulateur d'ordinateur quantique. Toutefois, les ordinateurs quantiques actuels sont entravés par des qubits bruyants, ce qui entraîne une détérioration des résultats. Cependant, l'étude révèle que les résultats s'amélioreront considérablement à mesure que la technologie de l'informatique quantique évoluera [14].

Dans cette continuité, une publication a présenté une nouvelle étude proposant une méthodologie de conception de cellules Quantum Long Short-Term Memory (QLSTM) pour l'apprentissage automatique quantique. Cette technique adapte les fonctionnalités des réseaux de neurones artificiels LSTM aux principes de la physique quantique, permettant ainsi à la QLSTM de simuler le comportement d'une LSTM conventionnelle lorsqu'il s'agit d'apprendre des fonctions cycliques. Des simulations numériques illustrent cette transformation, mettant de l'avant la possibilité d'une implantation matérielle efficace sur des dispositifs munis de portes quantiques améliorées. Il propose également d'étendre l'apprentissage en utilisant des dispositifs quantiques tolérants aux pannes, tout en reconnaissant les défis tels que le calcul des dérivées et la gestion des qubits dans les circuits associés [15]. Cela est complété par une étude récente visant à optimiser les hyperparamètres du modèle QLSTM, ce qui permet de minimiser les pertes de test tout en prévoyant les niveaux de pH, d'humidité et de température du sol. Une étude comparative

a été menée entre le modèle LSTM et le modèle QLSTM. Le modèle LSTM traditionnel connaît une décroissance exponentielle marquée par une forte instabilité à travers les âges. En revanche, le modèle QLSTM affiche une baisse exponentielle plus stable. Les résultats obtenus avec notre modèle hybride QGA-PSO QLSTM sont très encourageants et représentent une solution polyvalente [16].

En effet, que ce soit dans le baseball ou dans d'autres sports, la prévision fiable dépend de l'analyse de nombreuses données. La gestion de vastes ensembles de données représente des défis importants dans le contexte actuel.

Examinons de plus près la particularité de l'informatique quantique en tant que solution pour le traitement des données en masse.

2.2 Informatique quantique

La mécanique quantique est une discipline scientifique de base qui permet grâce à sa précision de comprendre le comportement du monde microscopique en particulier pour les objets techniques tels que les ordinateurs et les téléphones [17].

La figure ci-dessous montre un exemple remarquable d'ordinateur quantique, illustrant sa complexité interne.



Figure 2.2: Ordinateur Quantique (Source: IBM Canada)

Cette illustration (Figure 2.2) représente une photo d'un ordinateur quantique d'IBM.

Il existe plusieurs distinctions entre le système classique et le système quantique. Notamment la superposition, qui permet à un qubit d'être dans deux états simultanément, et l'étrangement. Dans l'informatique classique, les bits ne peuvent pas être en superposition d'états, contrairement à l'informatique quantique. Ils ne peuvent prendre que les valeurs 0 ou 1 [18].

La physique quantique est l'une des plus importantes révolutions du vingtième siècle. Bien qu'elle soit une science complexe à comprendre et que certains de ses phénomènes semblent contredire notre compréhension du monde physique classique, la théorie quantique est solidement étayée par de nombreuses expériences. De plus, elle trouve des applications concrètes dans notre vie quotidienne. Elle rassemble des théories qui décrivent le comportement des atomes et des particules, tout en éclairant diverses caractéristiques du rayonnement électromagnétique [19].

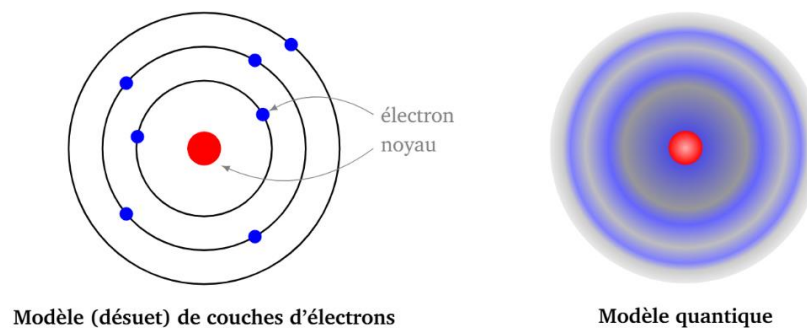


Figure 2.3: Représentation des modèles de particules (source : Micro-stage-IMACA)

Cette illustration (Figure 2.3) met en évidence la distinction entre deux conceptions de l'atome :

- Le modèle des couches électroniques (désuet) : Ce modèle représente les électrons comme des particules orbitant autour du noyau de l'atome de manière similaire à la manière dont les planètes orbitent autour du Soleil. Ce modèle a d'abord été utilisé pour expliquer certains phénomènes atomiques,

mais il est maintenant dépassé parce qu'il ne tient pas compte des propriétés quantiques des électrons.

- **Modèle quantique :** Dans ce modèle quantique, les électrons ne sont plus considérés comme des particules sur des orbites fixes, mais plutôt comme des nuages de probabilité. Ces nuages (ou orbitales atomiques) révèlent les régions où il est le plus probable de détecter un électron à un instant donné. Ce modèle reflète la nature ondulatoire des électrons décrite par la mécanique quantique et constitue la base de notre compréhension moderne des atomes.

La physique quantique englobe tous les domaines de la physique où l'application des principes de la mécanique quantique est cruciale pour comprendre les phénomènes observés. La mécanique quantique sert de théorie de base pour décrire les particules qui composent la matière de l'univers ainsi que les champs de force qui les régissent.

2.2.1 Bits quantiques ou qubits

Les qubits sont les éléments de manipulation de base de l'information dans les ordinateurs quantiques. On migre d'un monde déterministe à un monde probabiliste.

Dans l'informatique classique, les bits correspondent à des charges électriques circulantes qui traduisent le passage ou l'absence d'un courant électrique. Un bit est de valeur 1 si le courant passe soit de 0 si le courant ne passe pas. La lecture d'un bit donne 1 ou 0. Elle est déterministe, à savoir que si l'on répète l'opération de lecture plusieurs fois, ou l'opération de lecture après une réédition du calcul, on obtiendra normalement le même résultat.

En classique chaque information est codée en binaire. Pour les nombres l'information, peut être décoder sous la forme de :

$$nombre = \sum_{i=0}^{n-1} b_i 2^i \quad (2.1)$$

Avec b_i = l'état de bit et $i, n \in \mathbb{N}$.

Pour n bit est de $N = 2^n$ combinaisons possibles pour coder l'information. Et pour N objets on obtient comme nombre de bit nécessaire $n = \log_2(N)$

Un bit quantique est un vecteur unitaire dans un espace vectoriel complexe à deux dimensions qui a été fixé avec une particulière, notée $|0\rangle$ et $|1\rangle$. Il sert de base à l'information quantique [20].

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ et } |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (2.2)$$

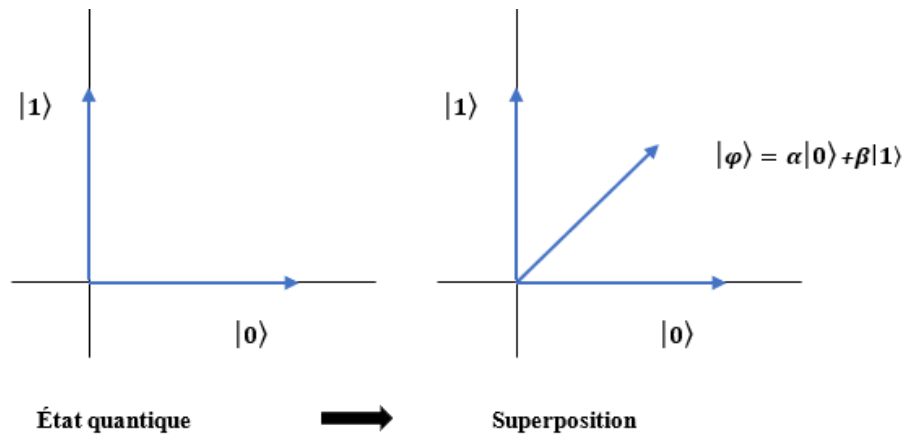


Figure 2.4: Schémas vectoriel d'état quantique (Élaboration personnelle réalisée avec Microsoft PowerPoint).

Cette illustration (Figure 2.4) représente les états quantiques dans un espace vectoriel.

Le qubit est caractérisé généralement par plusieurs propriétés qui sont :

- La superposition : qui permet au qubit d'être dans plusieurs états en même temps c'est-à-dire à l'état $|0\rangle$ et $|1\rangle$ à la fois
- L'intrication : qui permet de synchroniser les qubits.

2.2.2 Quelques portes quantiques.

La manipulation des qubits se fait par le biais de portes quantiques. Avant chaque calcul, on initialise le qubit dans l'état 0.

Les vecteurs en quantique sont des membres d'un espace vectoriel complexe. Pour un qubit, bit quantique : $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ avec α et β qui sont des nombres complexes et désigne également les probabilités respectives des états $|0\rangle$ et $|1\rangle$. Chaque état du qubit $|\varphi\rangle$ est associé à un vecteur colonne $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

$$|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle = |\varphi\rangle = \alpha\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta\begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (2.3)$$

Exemples de portes quantique :

- La porte Hadamard (H) : elle permet de mettre un qubits dans deux états en même temps.

$$H|0\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \quad (2.4)$$

Dans cette équation on constate que l'application de la porte Hadamard sur le qubits 0 le répartit en deux probabilités égales.

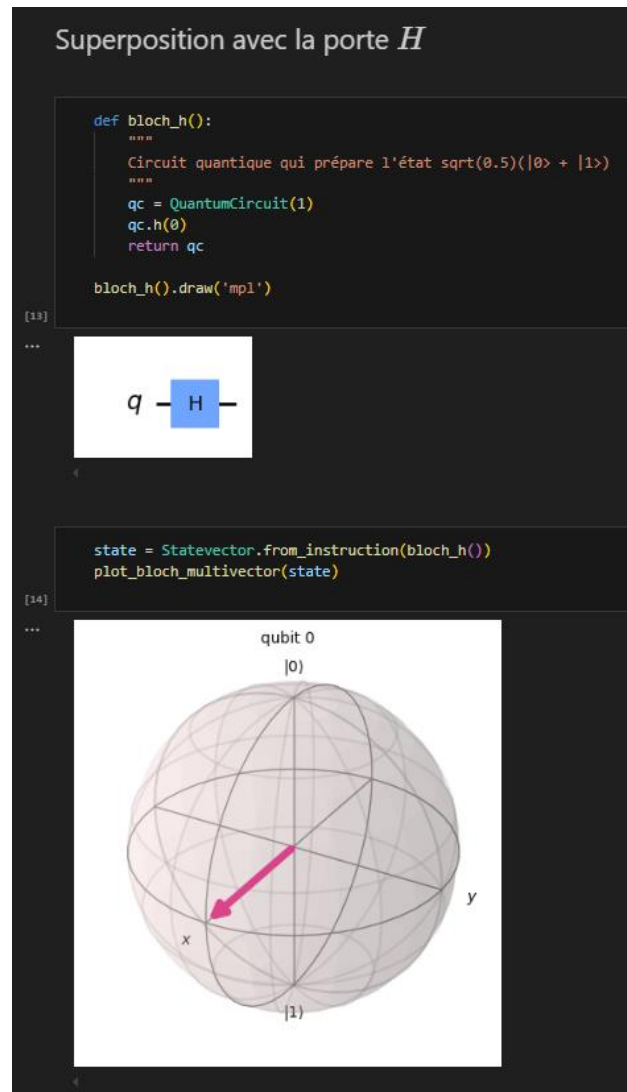


Figure 2.5: Représentation de la porte H sur un circuit quantique et l'état résultant dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).

- La porte X , également appelée porte Pauli X , inverse l'état $|0\rangle$ en $|1\rangle$ et vice versa. Elle est équivalente à la porte RX pour un angle de π .

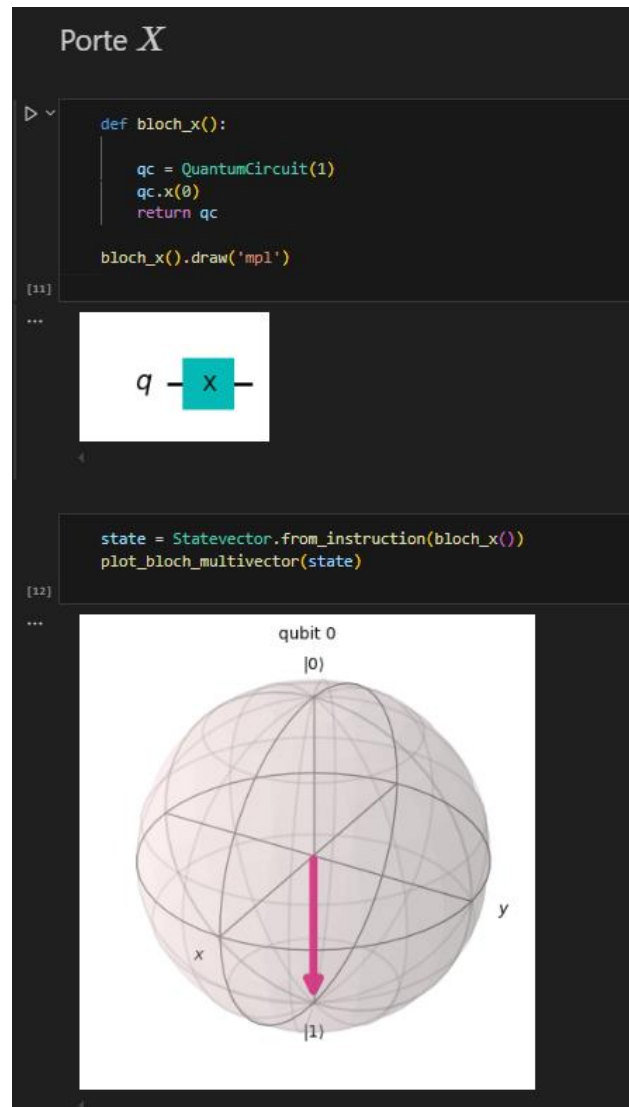


Figure 2.6: Représentation de la porte X sur un circuit quantique et l'état résultant dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).

- La porte NOT, également connue sous le nom de CNOT, agit sur une paire de qubits, où l'un des qubits sert de qubit de contrôle et l'autre, de qubit cible. Si le

qubit de contrôle est $|0\rangle$, alors le qubit cible reste inchangé. Sinon, une porte X agit sur le qubit cible.

- La porte Toffoli, aussi appelée CCNOT, agit sur trois qubits. Elle possède deux qubits de contrôle et un qubit cible. Les deux qubits de contrôle doivent être à l'état $|1\rangle$ pour qu'une porte X puisse agir sur le qubit cible.

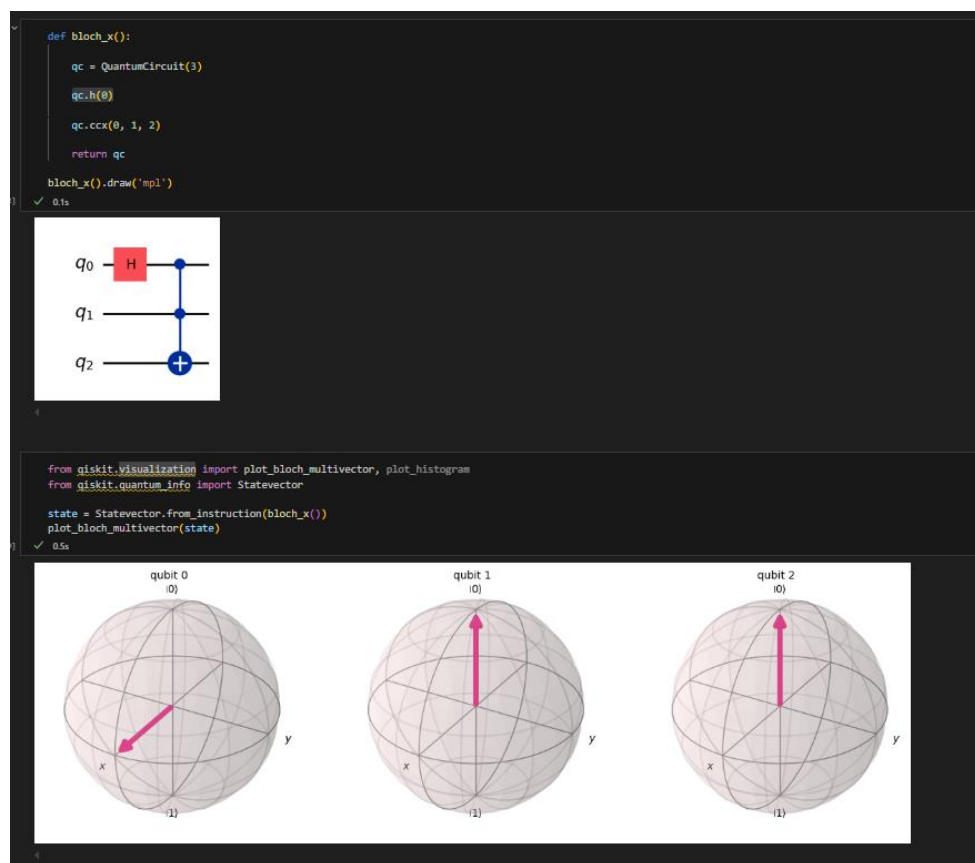


Figure 2.7: Représentation de la porte CCNOT avec la porte H sur un circuit quantique et les états résultants dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).

- La porte SWAP permute l'état de deux qubits.
- La porte identité ou porte I qui traduit l'absence de porte quantique.

- La porte Z ou Pauli Z qui ajoute un signe négatif à l'état $|1\rangle$.

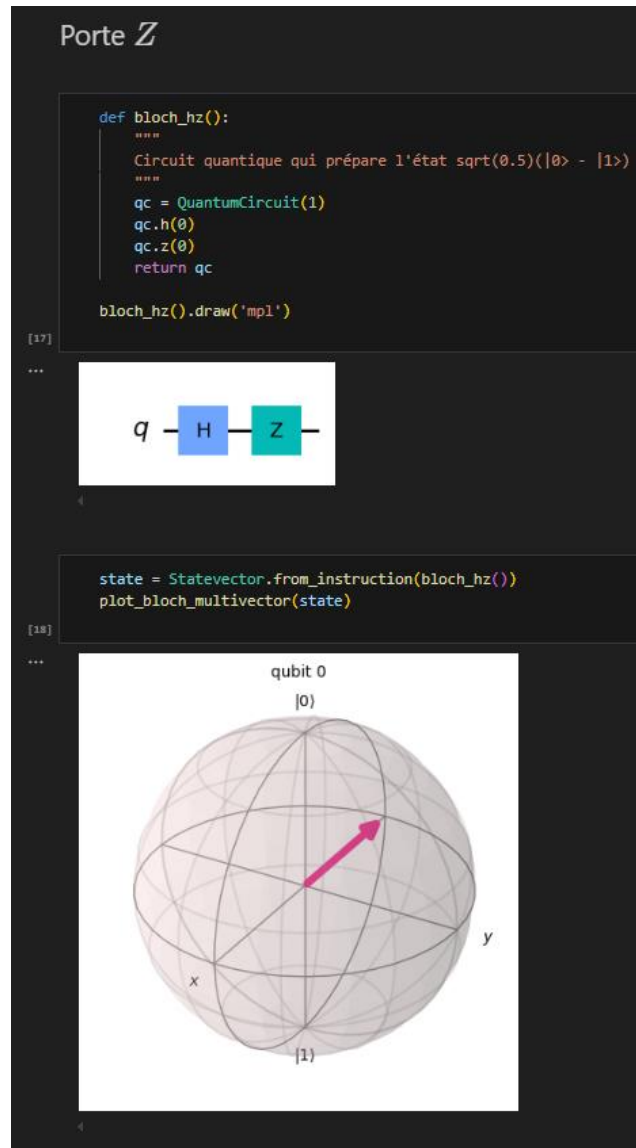


Figure 2.8: Représentation de la porte Z sur un circuit quantique et l'état résultant dans la sphère de Bloch (générée avec un script Python et Qiskit développé dans cette étude).

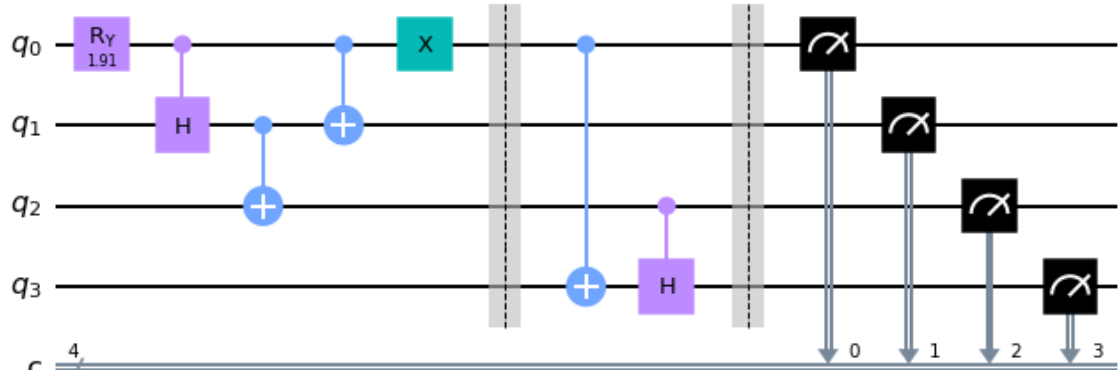


Figure 2.9: Exemples de portes sur un circuit quantique (générée avec Qiskit

Composer développé dans cette étude).

Ce circuit quantique se compose de quatre qubits, représentés par quatre lignes horizontales parallèles, sur lesquelles sont placées sept portes. Chacune de ces portes modifie l'état des qubits en fonction de son rôle. À la fin de chaque étape, des mesures sont effectuées.

La mesure d'un qubit est un processus fondamental dans le calcul quantique, où l'état quantique d'un qubit est projeté sur l'une des bases de mesure.

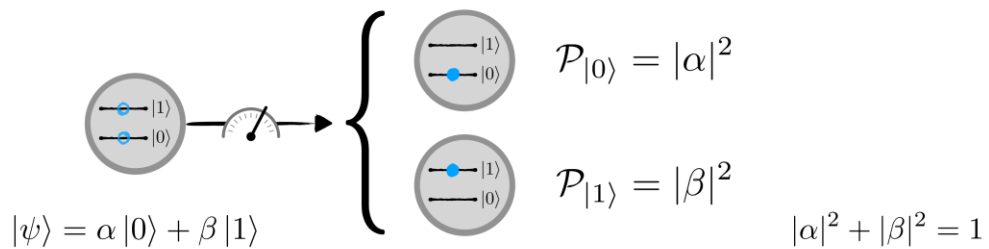


Figure 2.10: Mesure d'un qubit dans une superposition (source : Micro-stage-IMACA)

Cette illustration (Figure 2.10) résume le principe fondamental de la mécanique quantique : un état quantique ne peut pas être observé directement sans être perturbé. La mesure perturbe le système en le forçant à adopter un état classique. La vraisemblance de chacun des résultats est régie par les amplitudes quantiques α et β .

Tableau 2.1: Tableau résumant la mesure des qubits

	$\rho 0\rangle$	$\rho 1\rangle$
$ \Psi\rangle = 1\rangle$	0%	100%
$ \Psi\rangle = \frac{1}{\sqrt{2}} 0\rangle + \frac{1}{\sqrt{2}} 1\rangle$	50%	50%
$ \Psi\rangle = \frac{2\sqrt{2}}{3} 0\rangle - \frac{1}{3} 1\rangle$	88.89%	11.11%

2.3 Conclusion partielle

Ce chapitre a débuté par une introduction au Baseball, en mettant en évidence ses caractéristiques distinctives par rapport aux autres activités physiques. Nous avons vu que chaque poste exige un encadrement et un entraînement spécifiques pour les athlètes. Nous avons examiné des études visant à mieux comprendre les athlètes et à identifier les facteurs clés pour chacun d'entre eux.

Cette étude a également exploré les avancées de l'intelligence artificielle dans le domaine du sport et d'autres secteurs. Les modèles LSTM et ARIMA offrent des perspectives de prévision intéressantes, mais ils présentent des limites.

Pour surmonter ces obstacles, l'informatique quantique émerge comme une solution potentielle. En effet, elle permet d'appliquer des modèles de prévision dans un environnement quantique, ce qui ouvre la voie à des prévisions plus fiables et plus précises.

Chapitre 3 - SYSTÈME DE PRÉVISION

Le système de prévision, regroupe plusieurs sous-études, notamment la récolte et le traitement de données, le système de calcul et l'organisation, et les résultats de prévision.

L'image suivante illustre la méthode d'analyse de données, combinant une structure quantique :

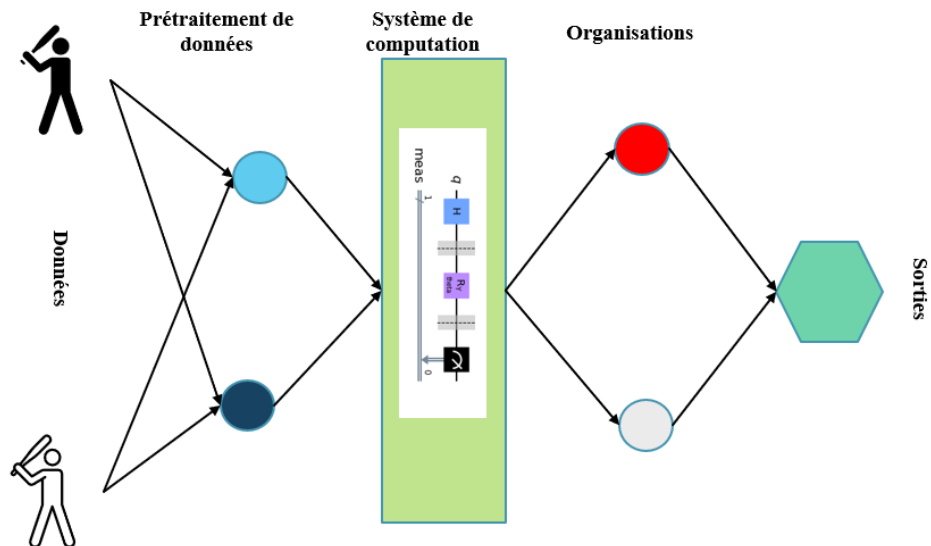


Figure 3.1 : Processus de traitement des données, intégrant une architecture quantique (Élaboration personnelle réalisée avec Microsoft PowerPoint).

Le graphique (Figure 3.1) se présente comme suit :

- **Données** : Les données initiales, telles que les statistiques des athlètes de baseball, sont symbolisées par les formes à gauche.

- **Prétraitement des données :** Les cercles bleus illustrent les différentes étapes de ce processus. Ces étapes consistent à choisir ou à extraire les caractéristiques pertinentes et à les normaliser ou les adapter à une échelle spécifique.
- **Système de computation :** Le rectangle vert central symbolise le système de calcul, dans ce cas un ordinateur quantique. Ce système exploite les propriétés quantiques (telles que le parallélisme) pour traiter les données de manière efficace.
- **Organisations :** Les cercles rouges et gris représentent des décisions ou des catégorisations fondées sur les résultats du calcul quantique. Ces étapes organisent les sorties en catégories ou groupes afin d'effectuer une analyse ultérieure.
- **Sorties :** L'hexagone à droite illustre les résultats finaux du système. Les sorties peuvent inclure des prévisions, des classifications ou des décisions exploitables dans divers contextes, comme l'amélioration des performances sportives.

Néanmoins, notre étude se concentrera sur la prévision des données des athlètes.

3.1 Traitement de données

3.1.1 Cueillette de données

La collecte de données s'effectue selon différents axes. Elle constitue un élément crucial dans le processus d'analyse et de mise en place de solutions. En effet, on peut

considérer cette collecte de données comme une étude à part entière. Dans notre cas, elle concerne celles des athlètes.

Tableau 3.1: Exemple de fiche de récolte

Catégories	Données	Unité de mesure	Période
• Physique	• Taille	• Cm, kg	• /mois
• Psychologie	• Motivation	• Niveau sur 10	• /semaine
• Technique	• Vitesse	• Etc.	• /jour
• etc.	• Etc.		• Etc.

Ce visuel est purement indicatif.

Les données recueillies peuvent être classées comme suit :

- Numériques : elles représentent des valeurs numériques.
- Catégorielles : elles représentent des groupes ou des classes qui n'ont pas d'ordre intrinsèque.
- Ordinales : elles représentent des groupes ou des classes qui ont un ordre intrinsèque.
- Binaires : elles représentent des valeurs qui ne peuvent prendre que deux états.
- Temporelles : elles représentent des valeurs dont la signification change avec le temps.

3.1.2 Analyse de données

L'analyse de données est cruciale dans l'élaboration d'un système de prévision. Elle permet de mieux structurer les informations, ce qui a un impact direct sur les résultats du modèle prédictif. Pour prévoir les performances à long terme des athlètes de baseball, il est important de prendre en compte l'évolution temporelle des données.

Différentes méthodes existent pour comprendre la nature et les tendances des données.

Dans notre étude, nous avons examiné plusieurs d'entre elles afin de :

- Maîtriser les données clés d'un athlète.
- Comprendre l'évolution de ces données au fil du temps.
- Identifier les éventuelles anomalies dans les données.
- Examiner le comportement des données pendant une période donnée.
- Détecter les cycles dans la série de données.
- Vérifier la stationnarité des données.

3.2 Systèmes de prévision

Au cours de la dernière décennie, l'intérêt pour l'apprentissage automatique, ou machine learning, a explosé, devenant omniprésent dans les programmes informatiques, les applications mobiles et les discussions scientifiques. Les débats actuels se concentrent sur les capacités réelles de l'apprentissage automatique par rapport aux attentes parfois exagérées. Fondamentalement, cette approche utilise des algorithmes pour extraire des informations exploitables à partir de données brutes, permettant d'inférer de nouvelles informations sur des données non modélisées. Les réseaux de neurones, modèles

d'apprentissage automatique existant depuis plusieurs décennies, représentent l'une des approches fondamentales. Ils sont constitués de nœuds, métaphores des neurones biologiques, et leurs connexions évoluent dans le temps, reflétant ainsi le concept d'apprentissage [21].

Les quatre architectures de base sont :

- Réseaux à apprentissage non supervisé;
- Réseaux de neurones convolutifs;
- Les réseaux de neurones récurrents;
- Réseaux de neurones récurrents;

Cependant, pour notre étude, nous nous concentrerons sur un sous-ensemble de ces architectures.

3.2.1 Réseaux de neurones artificiels

Les réseaux de neurones constituent un modèle de calcul partageant certaines caractéristiques avec le cerveau animal, où de nombreuses unités simples opèrent en parallèle, sans nécessité d'une unité de contrôle centralisée [21].

Le neurone est l'élément de base du système nerveux, qui est responsable de la réception, du traitement et de la transmission de l'information sous forme de signaux électrochimiques.

Les poids qui régissent les connexions entre ces unités représentent les principaux mécanismes de stockage de l'information à long terme au sein des réseaux de neurones. La mise à jour de ces poids constitue le principal processus par lequel le réseau de

neurones acquiert de nouvelles informations. Ils visent à modéliser le fonctionnement du cerveau dans le but de réaliser des tâches de calcul plus rapidement que les systèmes traditionnels [22].

Les principales applications incluent la reconnaissance et la classification de formes, l'approximation, l'optimisation, ainsi que le regroupement de données.

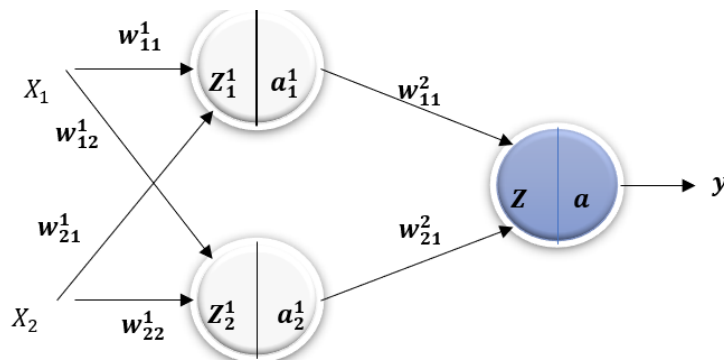


Figure 3.2: Schéma d'une partie d'un réseau de neurones artificiels (ANN)

(Élaboration personnelle réalisée avec Microsoft PowerPoint).

La Figure 3.2 illustre une partie du réseau de neurones artificiels (ANN), qui met en évidence les liens et les transformations entre les couches d'entrée, cachées, et de sortie. Les éléments de cette illustration sont X , w , Z , a et y qui sont respectivement les entrées, les poids, les fonctions des neurones, les fonctions d'activation et les sorties.

3.2.2 Réseaux de neurones récurrents

Les réseaux de neurones récurrents, qualifiés de Turing-complets, ont la capacité de simuler des programmes complexes [21].

En les considérant comme des outils d'optimisation, on peut les envisager comme des facilitateurs « d'optimisation de programmes ». Leur utilité se manifeste surtout dans la modélisation de fonctions avec une dépendance temporelle entre les vecteurs d'entrée et/ou de sortie, grâce à l'intégration de cycles dans le réseau, assurant ainsi une saisie efficace de l'aspect temporel des données.

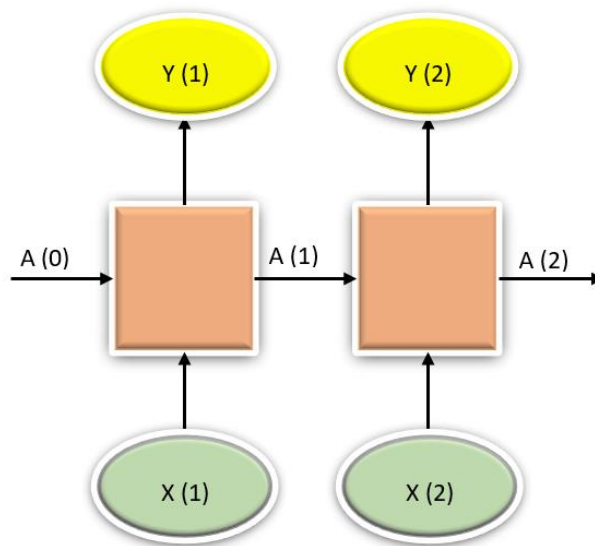


Figure 3.3: Schéma d'une partie d'un réseau de neurones récurrent (RNN)

(Élaboration personnelle réalisée avec Microsoft PowerPoint).

Cette représentation visuelle (Figure 3.3) illustre un modèle séquentiel dans lequel chaque étape prend en entrée un élément (X) pour en produire un autre (Y), tout en

transmettant un état (A) au bloc suivant. Les réseaux de neurones récurrents (RNN) ou les analyses séquentielles, comme les LSTM, sont des exemples de structures de ce type. Ils utilisent des informations passées pour influencer les calculs futurs. Les entrées évoluent progressivement à travers les modules, produisant des sorties correspondant à chaque étape, tout en saisissant les dépendances au sein des séquences.

On modélise un processus par la somme d'une partie déterministe et d'une partie aléatoire (modèle additif), ou par le produit d'une partie déterministe et d'une partie aléatoire (modèle multiplicatif). La séparation en partie déterministe et partie aléatoire est arbitraire [23].

3.2.3 *Modèle LSTM*

L'architecture LSTM (mémoire à court et long terme), développée par Hochreiter et Schmidhuber, fait partie de la famille des réseaux de neurones récurrents, permettant l'analyse de séquences chronologiques.

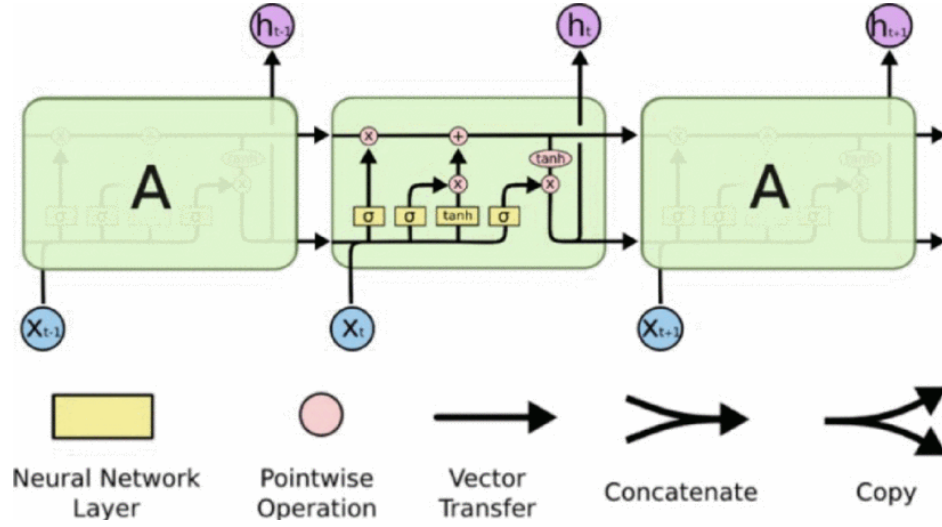


Figure 3.4: Schéma d'une cellule LSTM (Source : [24] © [2023] IEEE).

Les cercles roses représentent les opérations ponctuelles, comme l'addition ou la multiplication de vecteurs (Figure 3.4). Les cases jaunes correspondent aux couches du réseau neuronal utilisées pour l'apprentissage [24].

Dans la cellule LSTM, on a la porte d'oubli, la porte de mémoire, et la porte sortie.

- La porte d'oubli représente un état de transformation des valeurs passées. Une décision est prise quant à la conservation des informations. Cette décision se traduit par un état binaire 1 ou 0 ;

$$f_t = \sigma(w_f \times [h_{t-1}, x_t] + b_f) \quad (3.1)$$

- La porte de mémoire est aussi un état de transformation. Il s'agit d'une mise à jour des nouvelles valeurs. À ce stade, la porte décidera des valeurs stockées ;

$$i_t = \sigma(w_i \times [h_{t-1}, x_t] + b_i) \quad (3.2)$$

$$\hat{c}_t = \tanh(w_c \times [h_{t-1}, x_t] + b_c) \quad (3.3)$$

- Il existe une étape de transformation intermédiaire qui met à jour les valeurs passées ;

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t \quad (3.4)$$

- La porte de sortie est également un état de transformation des valeurs. Dans cette étape la cellule LSTM configure les valeurs futures ;

$$o_t = \tanh(w_o \times [h_{t-1}, x_t] + b_o) \quad (3.5)$$

$$h_t = o_t \times \tanh(C_t) \quad (3.6)$$

- $f_t, i_t, \hat{C}_t, C_t, o_t, h_t$ = fonction des états de transformations ;
- σ, \tanh = les fonctions d'activation ;
- w_f, w_i, w_c, w_o = les poids des différents états de transformations ;
- b_f, b_i, b_c, b_o = biais des différents états de transformations ;
- h_{t-1}, x_t = respectivement les valeurs passées d'une cellule LSTM, et les valeurs d'entrée ;
- h_t = les valeurs futures de la cellule LSTM ;

3.2.4 Modèle ARIMA

Le modèle ARIMA (AutoRegressive Integrated Moving Average) est un modèle utilisé pour le traitement des séries temporelles. ARIMA est un modèle autorégressif de moyenne mobile intégrée différentielle, également connu sous le nom de modèle autorégressif de moyenne mobile intégrée (le déplacement peut également être appelé glissement) [25].

Dans ARIMA, on a :

- AR (AutoRegressive) : C'est une composante importante se basant sur les valeurs passées pour établir des valeurs futures avec p comme nombre d'autorégressifs ;

$$X_t = c + \sum_{i=1}^p \Phi_i X_{t-i} + \epsilon_t \quad (3.7)$$

- X_t, X_{t-i} , représentent respectivement les valeurs actuelles et passées de la série ;
- c , est une constante ;
- Φ_i , sont les coefficients de la composante ;
- ϵ_t , désigne l'erreur ;

- I (Integrated) : l'intégration indique la différenciation aux données;

$$Y_t = X_t - X_{t-1} \quad (3.8)$$

- MA (Moving Average) : C'est une composante permettant d'établir des dépendances linéaires entre les erreurs passées et les observations avec q comme nombre de terme mobiles ;

$$X_t = \mu + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j} \quad (3.9)$$

- μ : moyenne mobile ;
- ϵ_t : erreur ;
- θ_j : coefficient de la moyenne mobile ;

3.2.5 Quantum LSTM

Le QLSTM (Quantum Long Short-Term Memory) est une combinaison du LSTM classique avec les circuits quantiques.

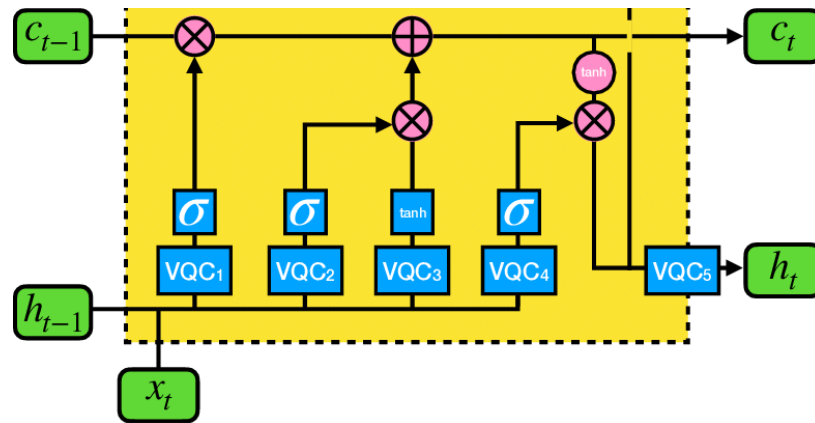


Figure 3.5: Schéma d'une cellule QLSTM (Source : Samuel Yen-Chi Chen, © [2022], IEEE)

On observe dans cette disposition (Figure 3.5) des ensembles de blocs précédant l'activation. Il s'agit des VQC (Variational Quantum Circuits). Les algorithmes variationnels sont constitués de plusieurs modules qui peuvent être combinés et optimisés en fonction des avancées algorithmiques, logicielles et matérielles. Cela comprend une fonction de coût qui définit un problème particulier en utilisant un ensemble de paramètres, une approche pour représenter l'espace de recherche à l'aide de ces paramètres, ainsi qu'un algorithme d'optimisation pour explorer cet espace de manière itérative. Chaque itération, l'algorithme évalue la fonction de coût avec les paramètres actuels et choisit les paramètres pour l'itération suivante, jusqu'à ce qu'il atteigne une

solution optimale. La nature hybride de cette famille d’algorithmes provient du fait qu’elle évalue les fonctions de coût à l’aide de ressources quantiques, mais les optimise avec des ressources classiques [26].

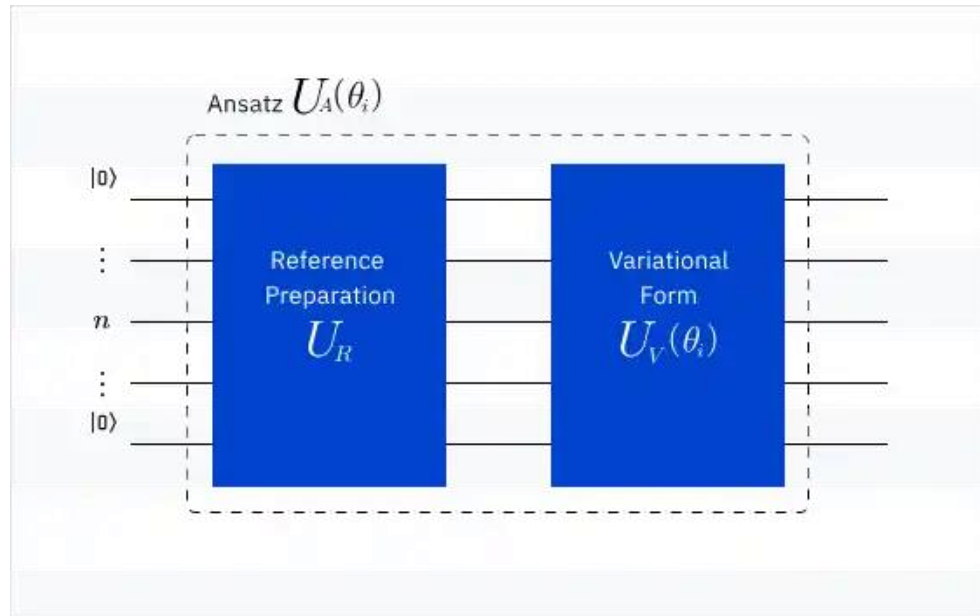


Figure 3.6: Circuit d'implémentation des VQC (source : IBM Quantum Learning Platform)

U_R symbolise l’encodeur de données, tandis que $U_V(\theta_i)$ représente le circuit variationnel (Figure 3.6).

L’encodeur de données quantique permet de stocker des informations en format classique sous forme quantique, afin de tirer parti du matériel quantique. On compte actuellement plusieurs méthodes d’encodage quantique.

$$f_t = \sigma(VQC_1(v_t)) \quad (3.10)$$

$$i_t = \sigma(VQC_2(v_t)) \quad (3.11)$$

$$\hat{C}_t = \tanh(VQC_3(v_t)) \quad (3.12)$$

$$c_t = f_t * c_{t-1} + i_t * \hat{C}_t \quad (3.13)$$

$$o_t = \sigma(VQC_4(v_t)) \quad (3.14)$$

$$h_t = VQC_5(o_t * \tanh(c_t)) \quad (3.15)$$

Ces équations représentent les équations d'une cellule LSTM classique combinée avec des algorithmes variationnels quantiques.

3.3 Conclusion partielle

Dans ce chapitre, nous avons passé en revue différents systèmes de prévisions couramment utilisés dans le domaine du sport et dans d'autres secteurs, en nous basant sur notre architecture comme ligne directrice. Cette architecture sert de cadre de référence pour le développement de notre système. Cette revue a mis en évidence les principes de fonctionnement et l'utilisation de modèles tels qu'ARIMA, LSTM et QLSTM.

Chapitre 4 - MODÉLISATION ET RÉSULTATS

Chaque athlète possède des caractéristiques différentes, telles que sa condition physique, son historique de performances ou encore sa morphologie. Ces différences rendent difficile l'utilisation d'un modèle unique avec des hyperparamètres fixes. Il est donc crucial de développer un système capable de s'adapter dynamiquement aux spécificités de chaque athlète pour fournir des prévisions fiables.

L'implantation du système s'est effectuée en plusieurs étapes. Avant sa mise en place, il a fallu d'abord élaborer une architecture de travail représentant notre démarche. La figure ci-dessous illustre le travail accompli.

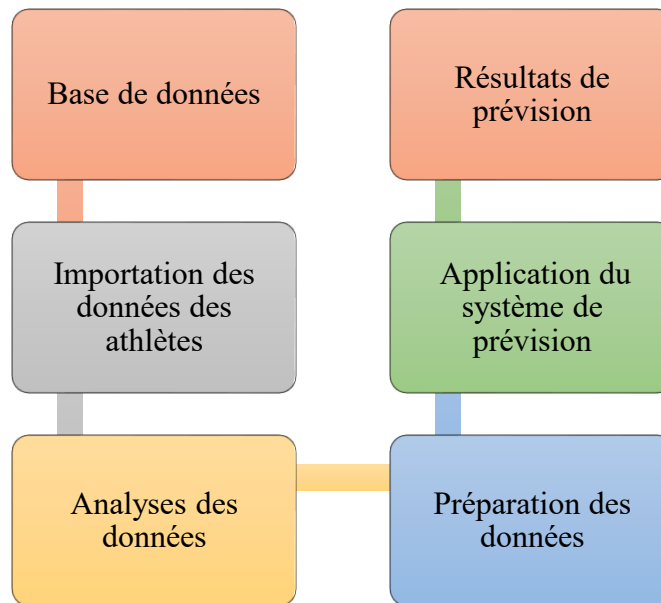


Figure 4.1: Architecture de travail (Élaboration personnelle réalisée avec Microsoft PowerPoint).

L'image (Figure 4.1) illustre le processus de gestion et d'analyse des données des athlètes afin de faire des prévisions. Cette séquence assure un flux structuré, depuis la collecte des données jusqu'à leur exploitation dans le cadre de ces prévisions.

L'informatique quantique représente une avancée majeure dans le domaine des technologies de l'information, promettant de résoudre des problèmes complexes bien au-delà des capacités des ordinateurs classiques. Pour exploiter pleinement cette technologie, des outils et des environnements spécifiques ont été développés.

4.1 Modélisation du système quantique

4.1.1 *Prétraitement des données*

Le prétraitement est essentiel pour réduire les biais causés par la différence des données.

Au départ, l'analyse des données a commencé par leur importation et leur structuration. Ces données proviennent du site web suivant : <https://stathead.com/>.

L'importation des données constitue le premier jalon crucial de tout projet d'analyse. Cette opération permet de charger les informations brutes dans un format exploitable par les outils de traitement et d'analyse. Sans une importation correcte, les étapes subséquentes ne peuvent pas être réalisées. Elle assure également que les données sont correctement formatées pour éviter les erreurs futures.

Après l'importation des données une première matrice de corrélation est faite. L'analyse de la corrélation des variables avec le temps permet d'identifier les variables qui évoluent de manière significative au fil du temps. Cette étape consiste à identifier les

variables qui capturent le mieux les évolutions dans le temps, ce qui peut aider à sélectionner les principales caractéristiques pour la modélisation.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (4.1)$$

Avec :

- x_i = Valeurs d'une variable;
- y_i = Valeurs de l'autre variable;
- \bar{x} et \bar{y} = Valeurs moyennes des variables respectives;
- r = Coefficient de corrélation ;

Enfin, choisir les variables.

Pour identifier les variables présentant les liens les plus étroits avec les autres variables dans notre matrice de corrélation, il est nécessaire d'analyser les coefficients de corrélation les plus élevés dans cette matrice.

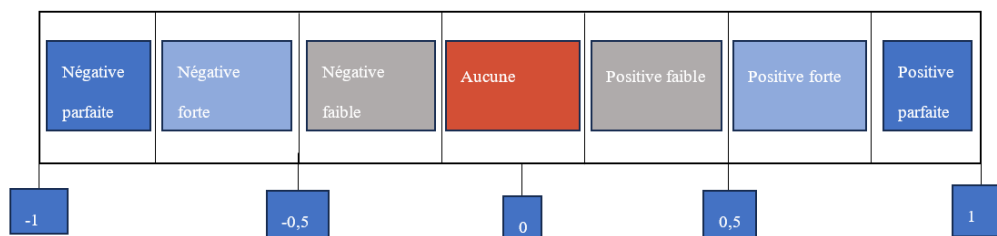


Figure 4.2: Niveau de corrélation (Élaboration personnelle réalisée avec Microsoft PowerPoint).

Ce diagramme (Figure 4.2) montre les niveaux différents de corrélation.

Voici la matrice de corrélation, montrant les relations entre les différentes variables.

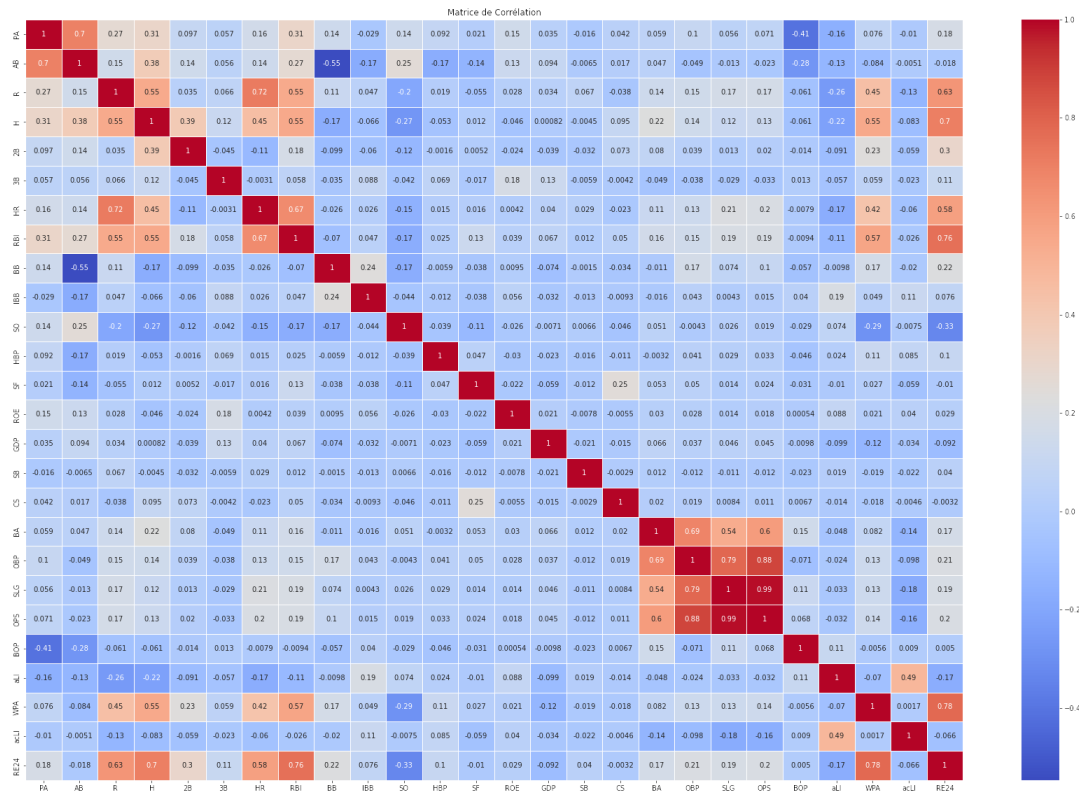


Figure 4.3: Matrice de corrélation (générée avec un script Python développé dans cette étude).

Cette matrice de corrélation (Figure 4.3) met en évidence les liens entre diverses variables d'un ensemble de données. Chaque case révèle l'intensité de la dépendance entre deux variables, allant de -1 (corrélation négative profonde) à 1 (corrélation positive profonde).

- Les zones teintées de rouge signalent des corrélations positives marquées : quand l'une augmente, l'autre tend à augmenter également;
- Les zones bleutées montrent des corrélations négatives : quand l'une augmente, l'autre a tendance à diminuer;

- Les zones blanches indiquent un faible ou nul niveau de corrélation;

Cette visualisation permet de repérer rapidement les variables fortement liées ou indépendantes, ce qui est utile pour la modélisation et la prise de décision.

Après la première matrice de corrélation l'étape suivante concernait la suppression des valeurs aberrantes. Les valeurs aberrantes peuvent fausser les analyses et les résultats des modèles prédictifs. Cette étape consiste à les détecter et à les éliminer. En supprimant ces anomalies, on améliore la qualité des données et on s'assure que les modèles entraînés ne sont pas influencés de manière disproportionnée par des points extrêmes. L'écart interquartile (IQR) est utilisé pour identifier les valeurs aberrantes. Les données situées en dehors de l'interquartile de 1,5 fois sont considérées comme des valeurs aberrantes et sont éliminées.

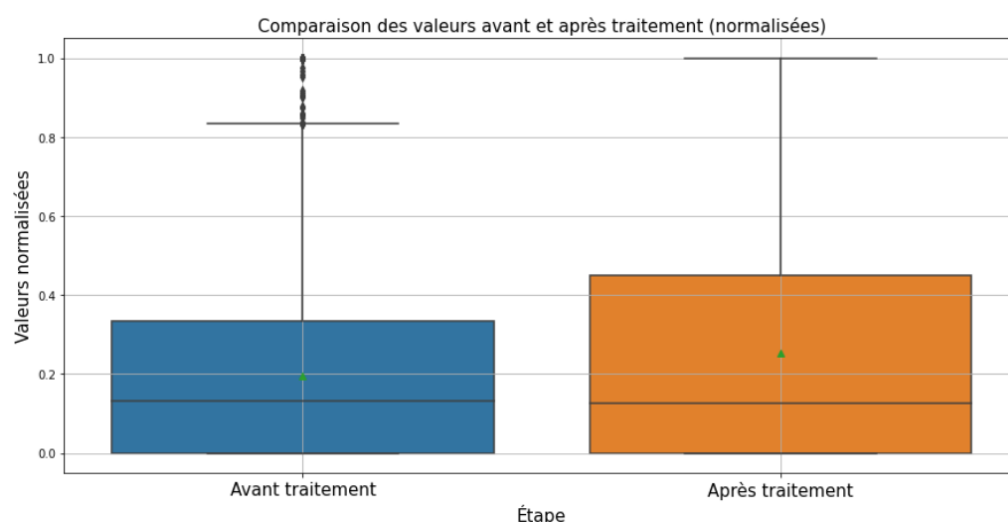


Figure 4.4: Comparaison des valeurs normalisées OPS avant et après traitement (générée avec un script Python développé dans cette étude).

Cette illustration (Figure 4.4) met en évidence une comparaison des valeurs avant et après le traitement des valeurs aberrantes, avec une normalisation appliquée dans les deux cas. Avant traitement, les données montrent plusieurs valeurs aberrantes situées au-delà du troisième quartile. La plage des valeurs est relativement large, avec des valeurs extrêmes qui peuvent influencer les analyses et les modèles si elles ne sont pas traitées. Après avoir éliminé les valeurs aberrantes, les données sont devenues plus homogènes, ce qui a permis d'offrir un tableau plus représentatif. Les valeurs extrêmes observées avant traitement ont été ajustées pour se situer dans la plage acceptable définie par l'IQR.

Après avoir éliminé les valeurs extrêmes, une nouvelle matrice de corrélation est créée et un dendrogramme est inclus pour représenter les liens mis à jour entre les variables. Cela permet de détecter des regroupements logiques et de faciliter la sélection de la variable la plus adaptée pour la prévision.

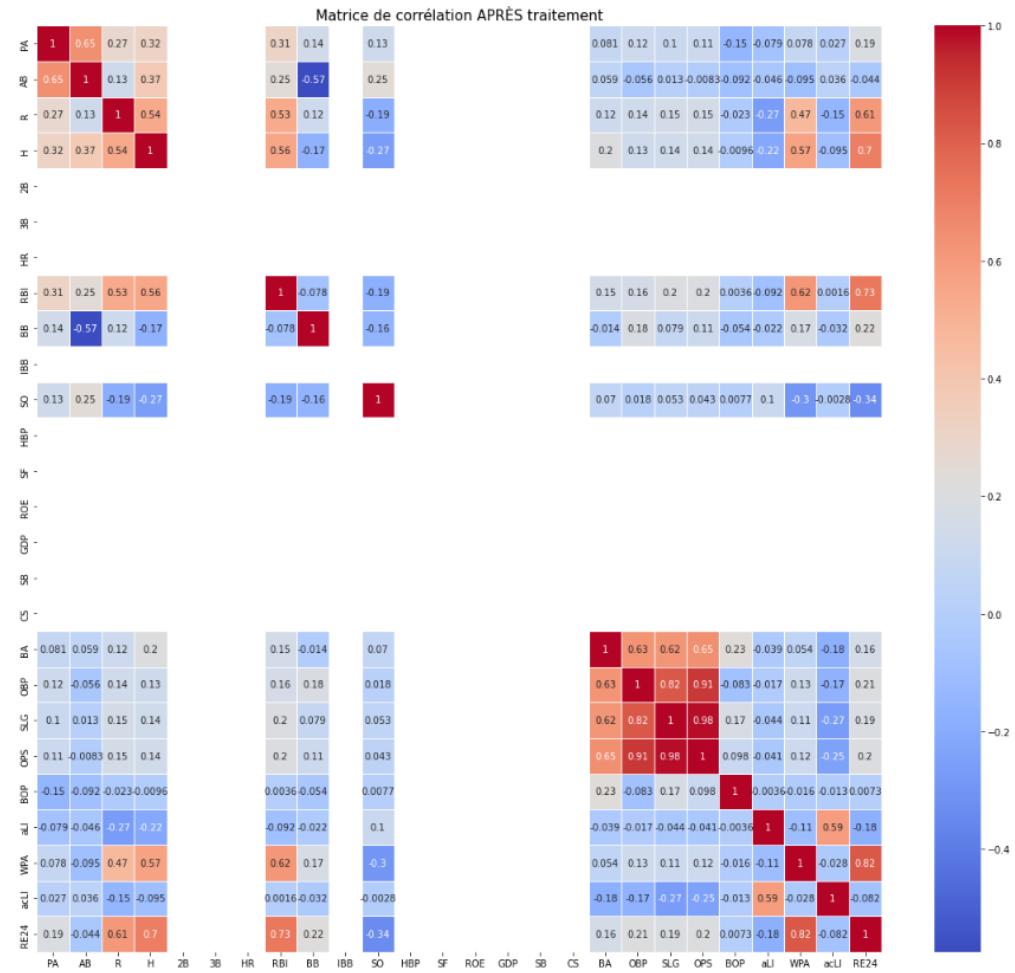


Figure 4.5: Matrice de corrélation après traitement (générée avec un script Python développé dans cette étude).

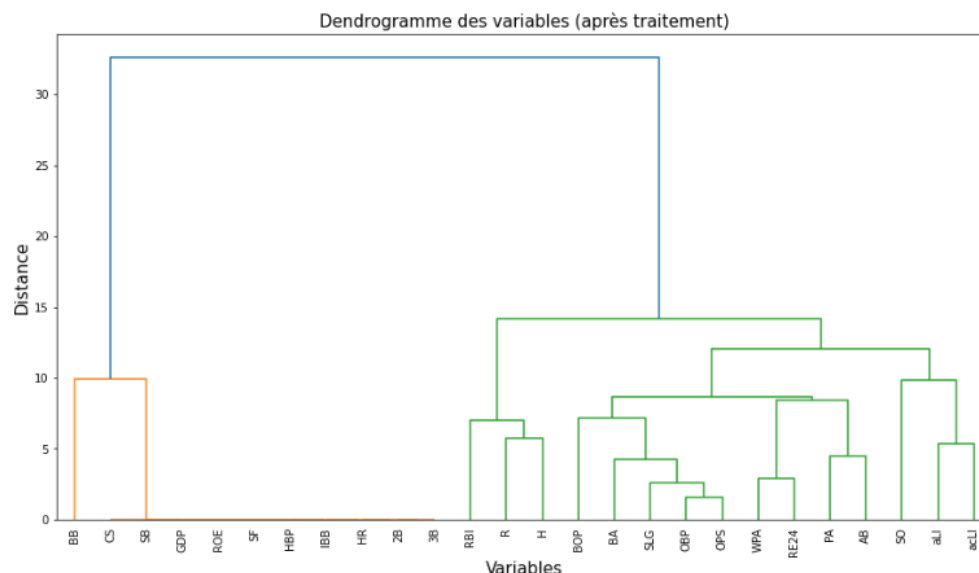


Figure 4.6: Dendrogramme des variables après traitement (générée avec un script Python développé dans cette étude).

La matrice de corrélation et le dendrogramme (Figure 4.5 et Figure 4.6) illustrent les étapes du processus d'analyse des données après le traitement des valeurs aberrantes.

La matrice de corrélation traitée révèle les liens linéaires entre les différents attributs des données après élimination des valeurs aberrantes. L'objectif est de détecter les attributs les plus fortement corrélés.

Le dendrogramme des variables traitées illustre la hiérarchisation des attributs par regroupement. Dans le dendrogramme, les variables proches, celles qui sont rattachées à des niveaux bas, présentent des traits similaires ou une forte corrélation. Cela permet de regrouper des ensembles de variables, facilitant une sélection rationnelle des variables pertinentes.

La matrice et le dendrogramme montrent que certaines variables, comme OPS (On-base Plus Slugging), R (Runs) et WPA (Win Probability Added), entretiennent une

relation significative avec d'autres variables du jeu de données. Ces informations aident à identifier des variables redondantes, à éliminer celles peu pertinentes et à choisir les plus influentes pour la prévision ou l'analyse. De cette manière, ils facilitent considérablement le choix des variables les plus appropriées, ce qui contribue grandement à simplifier les modèles statistiques.

En ce qui concerne cet athlète, l'objectif est un modèle prévisionnel de performance globale de l'athlète. OPS est le meilleur choix, car il offre une vision globale.

Cette étape est essentielle pour garantir l'exactitude des résultats et la fiabilité des indicateurs de rendement, en évitant que des données aberrantes ne les influencent.

Après avoir éliminé les valeurs aberrantes et choisi la variable cible, les données sont organisées en séquences afin de les rendre compatibles avec un modèle séquentiel, tel que le QLSTM.

Après la création des séquences, on y applique un lissage afin de minimiser les éventuelles fluctuations soudaines. Le lissage est une technique pour nettoyer les données, qui permet d'enlever le bruit tout en préservant les tendances. Les données brutes peuvent contenir des fluctuations dues à des erreurs de mesure ou à des variations aléatoires. Le bruit dans les données peut rendre plus difficile l'apprentissage des modèles, qui ont alors du mal à déterminer la tendance réelle. Un lissage permet de mieux visualiser les tendances générales sur un graphique. Il améliore aussi la précision et la stabilité des modèles prédictifs, tout en facilitant leur interprétation.

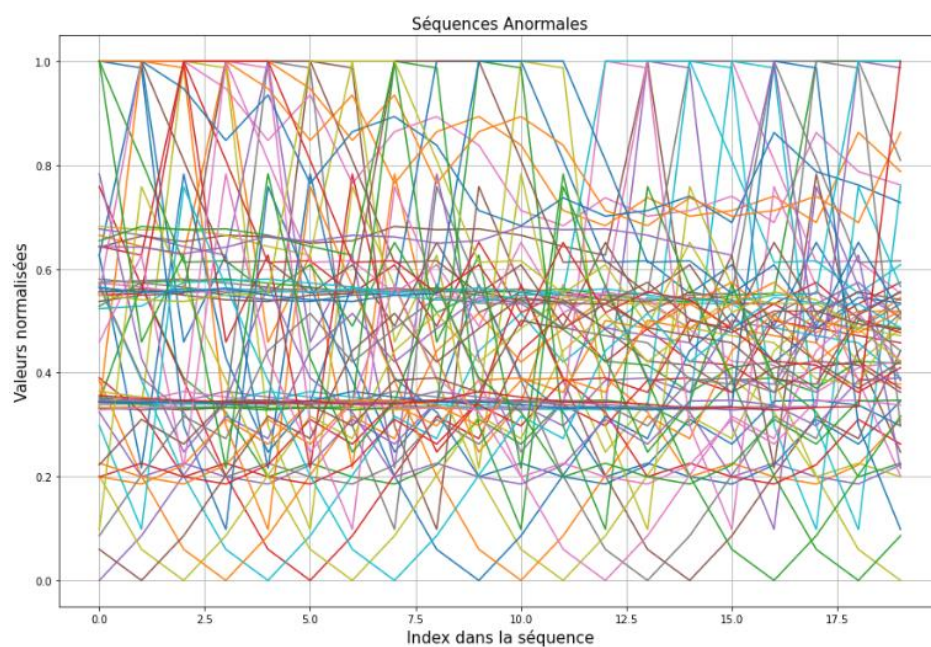


Figure 4.7: Séquences anormales (générée avec un script Python développé dans cette étude).

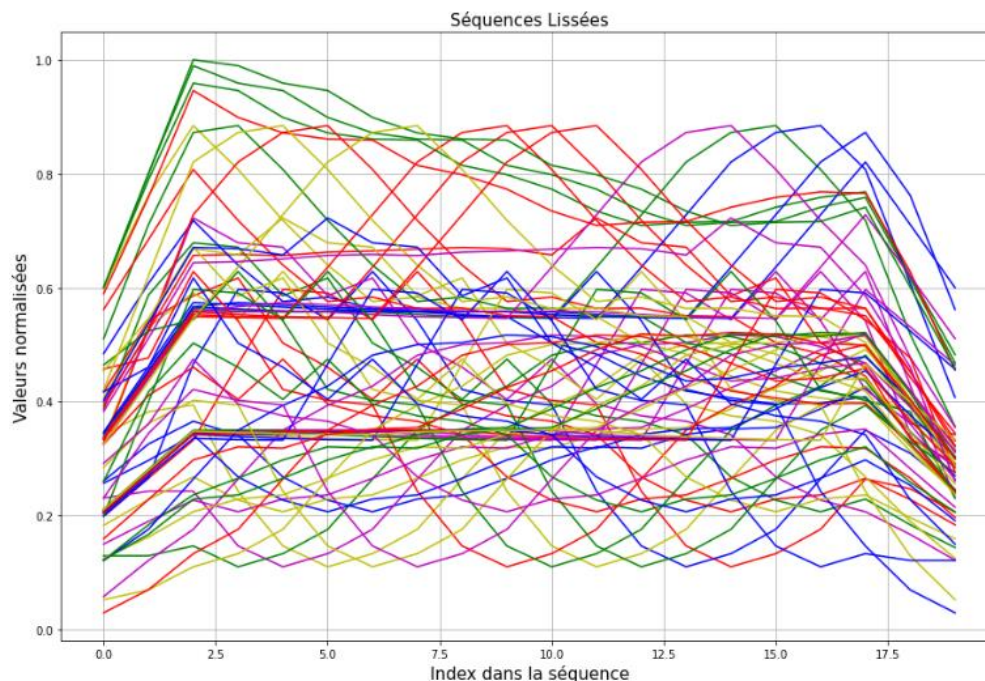


Figure 4.8: Séquences lissées (générée avec un script Python développé dans cette étude).

Les illustrations (Figure 4.7 et Figure 4.8) montrent les séquences anormales et les séquences lissées.

Le graphique des séquences anormales illustre les séquences détectées comme anormales avant le lissage. Les séquences initiales présentaient des variations irrégulières et des anomalies significatives entre les séquences.

Après l'application d'un lissage, les séquences montrent des variations plus uniformes.

Le lissage a permis de mettre en évidence les tendances sous-jacentes aux données, ce qui les rend plus uniformes. De plus, elles deviennent plus faciles à interpréter, ce qui facilite leur analyse.

Il s'est avéré que le traitement des données anormales par lissage est crucial pour obtenir des résultats de haute qualité. Cette amélioration assure une meilleure fiabilité et efficacité des modèles prédictifs, tout en facilitant la compréhension des tendances des données.

Ces étapes ont permis d'identifier la variable la plus importante pour l'athlète, qui servira de base aux prochaines analyses et au développement des modèles prédictifs. La prochaine étape consiste à exploiter ces données améliorées pour élaborer et former un modèle efficace capable de prévoir les résultats futurs de l'athlète.

4.1.2 Modèle QLSTM

Le déploiement du modèle QLSTM a été effectué en suivant une démarche structurée. Tout d'abord, l'architecture du modèle QLSTM (Quantum Long Short-Term Memory) a été établie, en intégrant des couches linéaires et des transformations quantiques. Ce modèle associe des fonctionnalités traditionnelles de LSTM à des éléments quantiques, tels que des qubits et des couches quantiques, pour saisir des liens complexes dans les données. Cette étape s'appuie sur deux aspects cruciaux : la conversion des données classiques en états quantiques manipulables par le modèle grâce à l'encodage quantique, et l'exécution de transformations d'états quantiques à l'aide de circuits quantiques pour mettre en évidence des liens complexes entre les variables temporelles. Ces étapes permettent au modèle d'optimiser l'utilisation des principes de la mécanique quantique dans l'apprentissage profond.

- Encodage quantique (ajout de la composante quantique) : Les caractéristiques extraites par la partie LSTM classique sont combinées avec l'état caché actuel pour former une entrée destinée au circuit quantique. Cette entrée est encodée dans les qubits en utilisant des techniques telles que AngleEmbedding, qui associent les valeurs classiques (par exemple, des vecteurs latents issus du LSTM) aux rotations des qubits.

Un qubit est un vecteur d'état dans un espace à deux dimensions. Pour encoder une valeur classique, AngleEmbedding utilise des rotations dans cet espace. Les rotations sont appliquées aux qubits selon les axes X, Y ou Z. Si une donnée classique est x , alors le qubit correspondant est manipulé par une porte de rotation :

- $R_X(x)$ = Rotation autour de l'axe X;
- $R_Y(x)$ = Rotation autour de l'axe Y;
- $R_Z(x)$ = Rotation autour de l'axe Z;

On peut représenter mathématiquement la rotation d'un qubit dans l'espace de Hilbert de la manière suivante :

$$R_k(x) = \exp\left(-i \cdot \frac{x}{2} \cdot \sigma_k\right) \quad (4.5)$$

Avec :

- $k \in \{X, Y, Z\}$: axe de rotation;
- σ_k : matrice de Pauli pour l'axe k .
- x : donnée classique à encoder

- **Traitement quantique :** Après avoir codé les données dans les qubits, le circuit quantique applique des transformations complexes en utilisant des couches telles que les `BasicEntanglerLayers`. Ces couches introduisent un enchevêtrement entre les qubits, ce qui permet de capturer des relations non linéaires complexes. Les transformations appliquées sur les qubits explorent des espaces de recherche plus vastes, difficiles à modéliser uniquement avec des réseaux classiques.

Représentation textuelle du circuit :

```

0: -AngleEmbedding(M0)-BasicEntanglerLayers(M1)- <Z>
1: -AngleEmbedding(M0)-BasicEntanglerLayers(M1)- <Z>
2: -AngleEmbedding(M0)-BasicEntanglerLayers(M1)- <Z>
3: -AngleEmbedding(M0)-BasicEntanglerLayers(M1)- <Z>
4: -AngleEmbedding(M0)-BasicEntanglerLayers(M1)- <Z>
5: -AngleEmbedding(M0)-BasicEntanglerLayers(M1)- <Z>

M0 =
[ 0.4476304  0.47950133  0.6846912 -0.2087166 -0.18194232 -0.39117011]
M1 =
[[-7.96082302e-01  6.72561936e-01  1.48039508e-02  6.58545601e-01
  -5.84087370e-01  2.77730565e-01]
 [-7.13709661e-01 -3.20423841e-01 -4.24938563e-01  6.58010337e-01
  1.07045125e-01 -5.69086330e-01]
 [-2.22928349e-04  9.64465926e-02  4.75932646e-02 -7.85117641e-01
  -1.77855698e-01  3.18050138e-02]]

```

Représentation graphique du circuit :



Figure 4.9: Représentation textuelle et graphique d'un circuit quantique (générée avec un script Python et PennyLane développé dans cette étude).

Cette illustration (Figure 4.9) montre un réseau quantique composé de qubits et de couches de transformations quantiques, qui est utilisé dans le modèle QLSTM. On observe des variables M0 et M1 dans le circuit quantique. M0 représente les données classiques encodées par la méthode d'encodage AngleEmbedding, tandis que M1 définit les poids

paramétriques des rotations et des interactions dans les couches BasicEntanglerLayers.

Après les transformations, chaque qubit est mesuré avec l'observable Z (Pauli-Z).

Ce circuit sert à traiter les séries temporelles dans la cellule quantique du QLSTM. Il permet d'enrichir le modèle grâce aux propriétés quantiques.

Après avoir créé le modèle, la recherche et l'optimisation des hyperparamètres les plus performants revêtent une importance cruciale pour améliorer les résultats du modèle QLSTM. Cette étape vise à déterminer les paramètres idéaux, tels que la taille des couches cachées, le nombre de qubits, le taux d'apprentissage et le nombre de couches quantiques, pour assurer une convergence rapide et une précision accrue.

La recherche des hyperparamètres s'est effectuée par une méthode systématique, en particulier par l'optimisation bayésienne, qui explore efficacement l'espace des hyperparamètres pour minimiser la perte de validation. De plus, une validation croisée a été intégrée pour évaluer la robustesse des choix sur différents sous-ensembles de données.

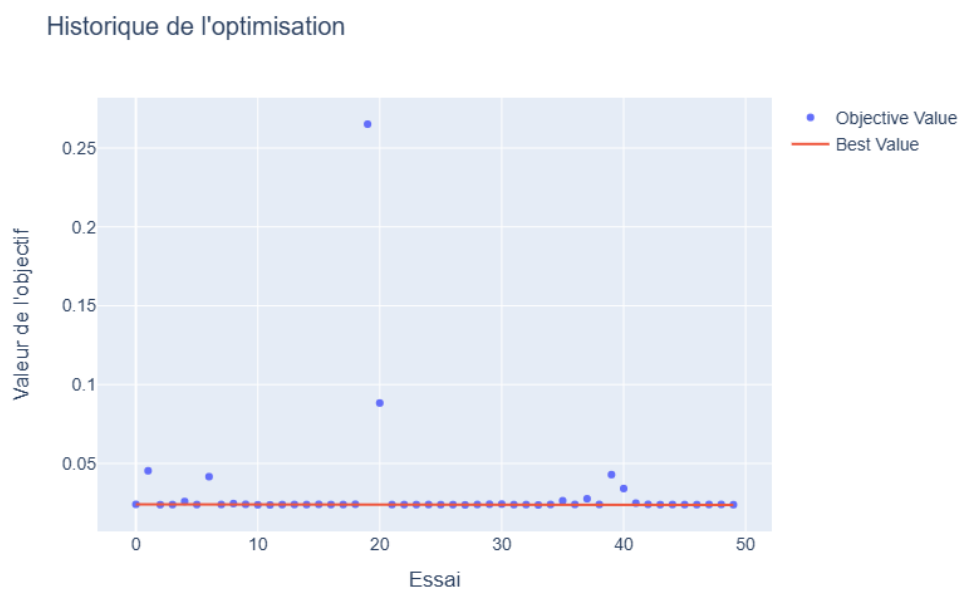


Figure 4.10: Historique de l'optimisation (générée avec un script Python développé dans cette étude).

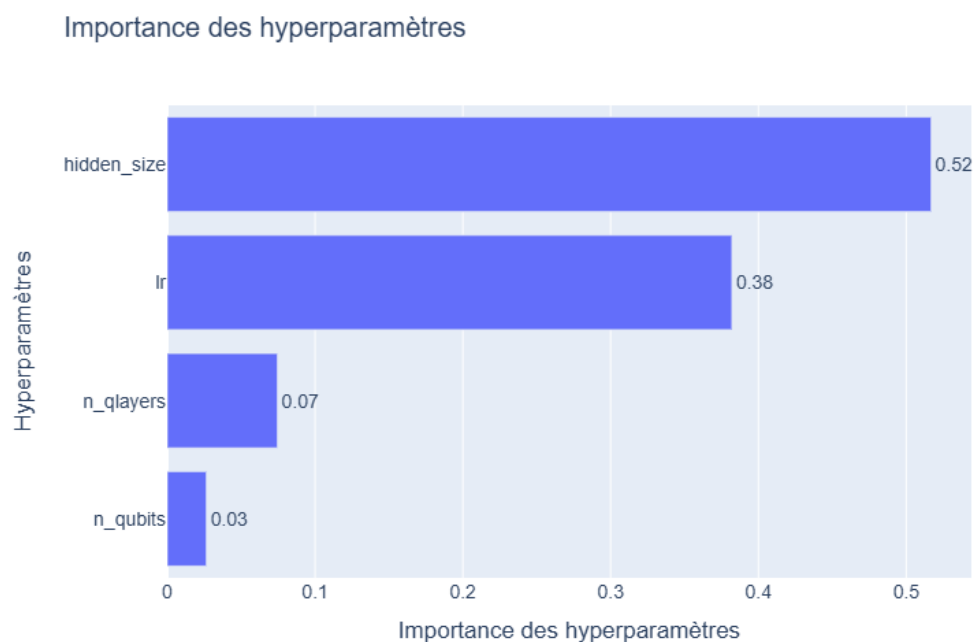


Figure 4.11: Importances des hyperparamètres (générée avec un script Python développé dans cette étude).

Le premier graphique (Figure 4.10) montre l'évolution de l'optimisation des hyperparamètres lors des différentes tentatives pour minimiser une fonction cible. L'axe horizontal indique les diverses itérations d'optimisation. Chaque point correspond à une combinaison unique d'hyperparamètres soumise à l'évaluation. L'axe vertical, quant à lui, montre la valeur de la fonction objective pour chaque combinaison d'hyperparamètres. Ce paramètre évalue la qualité de l'élaboration (par exemple, l'inexactitude ou le déclin du modèle).

Chaque point bleu représente les résultats d'une combinaison d'hyperparamètres lors d'une tentative. Un point plus bas indique de meilleures performances.

La ligne rouge, quant à elle, représente la meilleure valeur atteinte à chaque itération. Elle illustre l'évolution de l'optimisation et la manière dont les hyperparamètres s'améliorent.

Le second graphique (Figure 4.11) met en évidence l'importance relative des hyperparamètres dans l'optimisation du modèle QLSTM. Deux d'entre eux, `hidden_size` (taille des couches cachées) et `Lr` (taux d'apprentissage), sont particulièrement importants, représentant respectivement 52 % et 38 % de l'influence totale. En revanche, `n_layers` (nombre de couches quantiques) et `n_qubits` (nombre de qubits) ont un impact limité, avec des contributions de 7 % et 3 %. Cela souligne que l'accent doit être mis sur l'optimisation de `hidden_size` et `Lr` pour améliorer significativement les performances du modèle.

Cette optimisation assure que le modèle atteint un équilibre optimal entre sous ajustement et surajustement, garantissant ainsi des performances généralisées pour la prévision des séquences complexes des données de l'athlète.

Après l'optimisation, il y a la phase d'entraînement et de validation du modèle QLSTM. Cette étape est cruciale pour évaluer ses performances et affiner ses prévisions.

Pendant cette phase, le modèle est entraîné sur les données prétraitées en utilisant l'ensemble d'entraînement. Le processus repose sur une fonction de perte, telle que l'erreur quadratique moyenne (MSE), et sur un optimiseur pour ajuster les poids et les biais du modèle itération après itération.

En même temps, une vérification régulière est effectuée sur l'ensemble de validation pour surveiller la capacité du modèle à généraliser ses prévisions sur des données inédites. Cela permet d'éviter le surajustement en interrompant l'entraînement si les performances sur l'ensemble de validation commencent à se détériorer, grâce à des critères de convergence.

L'objectif principal de cette étape est de garantir que le modèle, optimisé par ses hyperparamètres, offre des prévisions précises tout en maintenant une bonne généralisation sur de nouvelles séquences de données.

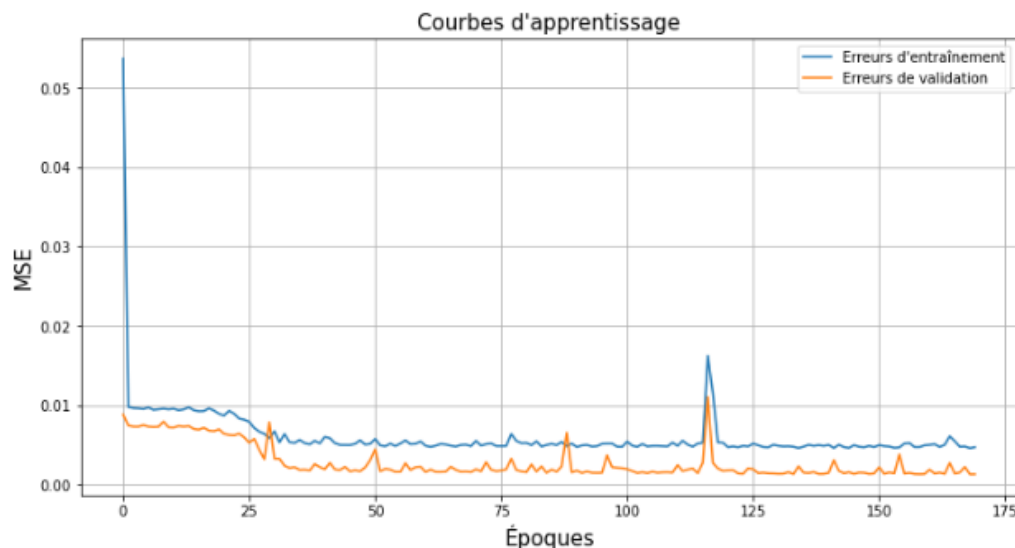


Figure 4.12: Courbes d'apprentissage MSE (générée avec un script Python développé dans cette étude).

Cette figure (Figure 4.12) représente la progression des pertes d'entraînement et de validation au cours des différentes itérations. Au départ, les erreurs diminuent rapidement, ce qui indique un apprentissage efficace. Elles se stabilisent ensuite, indiquant que le modèle a convergé. Toutefois, une hausse soudaine des erreurs survient vers l'époque 125, ce qui nécessite une analyse plus approfondie. Dans l'ensemble, le modèle démontre une bonne capacité de généralisation, avec des erreurs de validation légèrement inférieures à celles d'entraînement.

Après l'entraînement vient la phase de visualisation des prévisions, ainsi que celle où on implante une fonction qui produit des prévisions pour des dates spécifiques.

Dans cette étape, le modèle, une fois qu'il a été entraîné et validé, est soumis à un dernier test avec des données inédites. Cette méthode permet de mesurer l'exactitude

véritable des prévisions. Ces visualisations permettent d'identifier les éventuels écarts et de mieux comprendre le rendement du modèle.

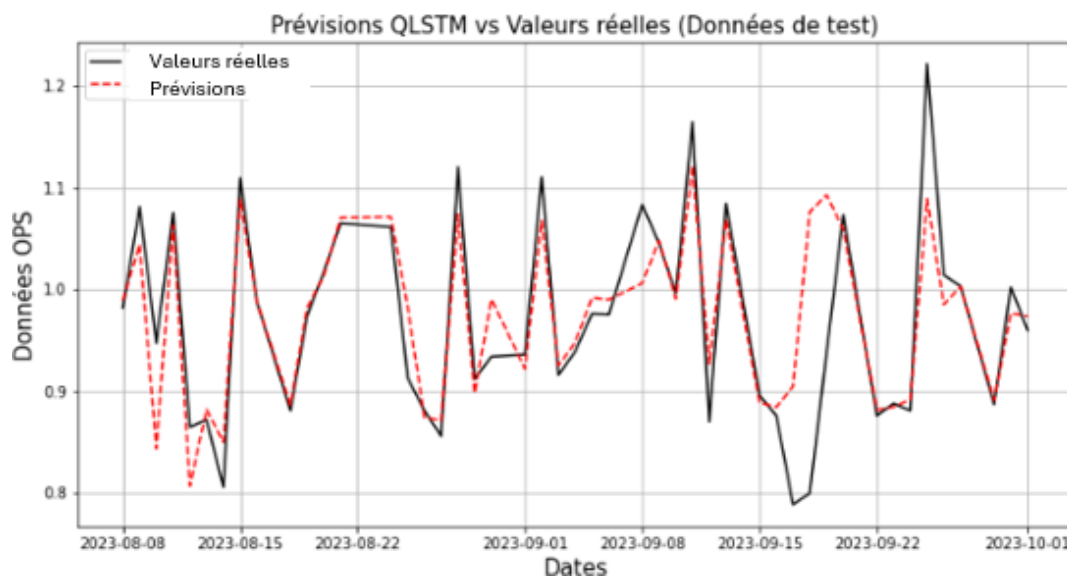


Figure 4.13: Comparaison des prévisions QLSTM avec les valeurs réelles (générée avec un script Python développé dans cette étude).

Cette illustration graphique (Figure 4.13) compare les résultats réels des données OPS (représentés en noir) et les prévisions du modèle QLSTM (représentées par une ligne ondulée de couleur bordeaux). L'objectif principal est d'évaluer les capacités du modèle à être précis et à suivre les tendances des données.

Les prévisions suivent globalement la tendance des valeurs réelles, avec des segments où les courbes se chevauchent, ce qui témoigne de la capacité du modèle à saisir les dynamiques sous-jacentes des données. On observe toutefois quelques divergences, notamment lors des pics ou des baisses soudaines (par exemple, vers le 15 septembre). Le modèle s'avère efficace pour prévoir les tendances générales des données OPS, tout en

mettant en évidence des domaines à améliorer, notamment la gestion des variations rapides. Cette analyse démontre la capacité du QLSTM à s'adapter aux variations progressives des valeurs réelles, ce qui indique un apprentissage adéquat sur les données de séquence. Le modèle offre des résultats prometteurs pour des applications, mais un affinement supplémentaire est nécessaire pour réduire les écarts et améliorer la fiabilité globale des prévisions.

En outre, une fonction flexible est mise en place pour permettre de générer des prévisions à des dates spécifiques, ce qui répond aux besoins d'anticipation ou de planification. Cela assure que le modèle peut être utilisé de manière opérationnelle pour faire des prévisions dans des contextes réels, tout en permettant une interprétation claire et visuelle des résultats.

Entrez une date future (AAAA-MM-JJ) : 2025-01-01

Prévisions futures :

	Date	Prédiction
0	2023-10-02	0.988621
1	2023-10-03	0.987155
2	2023-10-04	0.985756
3	2023-10-05	0.984399
4	2023-10-06	0.983081
..
453	2024-12-28	0.937070
454	2024-12-29	0.937070
455	2024-12-30	0.937070
456	2024-12-31	0.937070
457	2025-01-01	0.937070

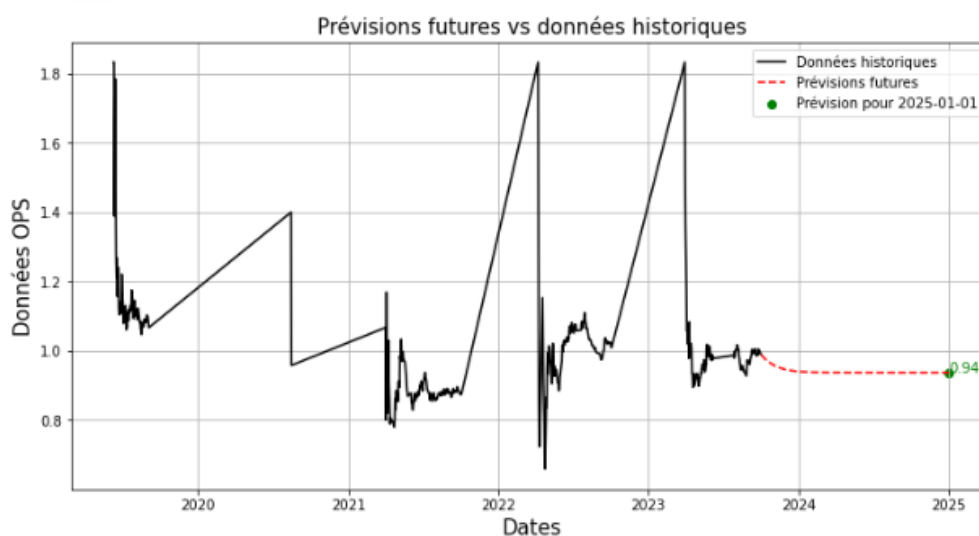


Figure 4.14: Prévisions futures vs données historiques (avec mise en évidence de la prévision pour le 1er janvier 2025) (générée avec un script Python développé dans cette étude).

Cette illustration (Figure 4.14) superpose les données historiques réelles (ligne noire) et les prévisions futures (ligne rouge pointillée), facilitant la compréhension de l'évolution prévue de la variable cible (OPS). Il accorde une importance particulière à la date choisie par l'utilisateur, ici le 1er janvier 2025, indiquée par un symbole vert et accompagnée de son estimation correspondante (0,937 070).

L'analyse montre une stabilité dans les prévisions futures après une certaine période, ce qui suggère que le modèle prévoit une stabilisation progressive des valeurs au fil du temps.

Ce type de représentation graphique est bénéfique pour évaluer la capacité du modèle à produire des prévisions à long terme et pour vérifier sa conformité avec les tendances observées dans les données historiques.

Pour assurer un fonctionnement optimal, le système exige que certaines conditions soient remplies. Tout d'abord, il est important de noter que l'utilisation d'un simulateur quantique fonctionnant sur un processeur classique (unité centrale de traitement) peut entraîner une baisse significative des performances du programme, surtout lorsqu'il manipule des séries temporelles complexes. Cela signifie que la création des jeux de données ainsi que l'entraînement des modèles peuvent nécessiter beaucoup de mémoire et de puissance de traitement.

Pour pallier ces limites, on suggère d'employer un simulateur plus puissant ou un matériel quantique réel, comme celui offert par IBM Quantum. Toutefois, il est crucial de garder à l'esprit que, bien qu'il soit prometteur, le calcul quantique n'est pas encore largement disponible. Par conséquent, demander l'utilisation de matériel quantique, comme celui d'IBM, peut entraîner des temps d'attente significatifs lors de l'exécution des opérations.

4.2 Résultats et analyses des modèles QLSTM, LSTM et ARIMA

Dans cette section, nous présenterons et évaluerons les conclusions tirées grâce aux méthodes QLSTM, LSTM et ARIMA, en les comparant sur la base des données examinées. La variable OPS est utilisée comme entrée de données dans les modèles servant à l'entraînement et à la prévision du comportement des systèmes. Cette analyse mettra en évidence les points forts et les points faibles de chaque approche, ainsi que leur capacité à répondre aux besoins de modélisation et de prévision.

Le graphique ci-dessous compare les résultats réels des données OPS aux prévisions du modèle LSTM (Long Short-Term Memory) sur la période étudiée. Cette visualisation permet d'évaluer la capacité du LSTM à saisir les tendances et les variations dans les données au fil du temps.

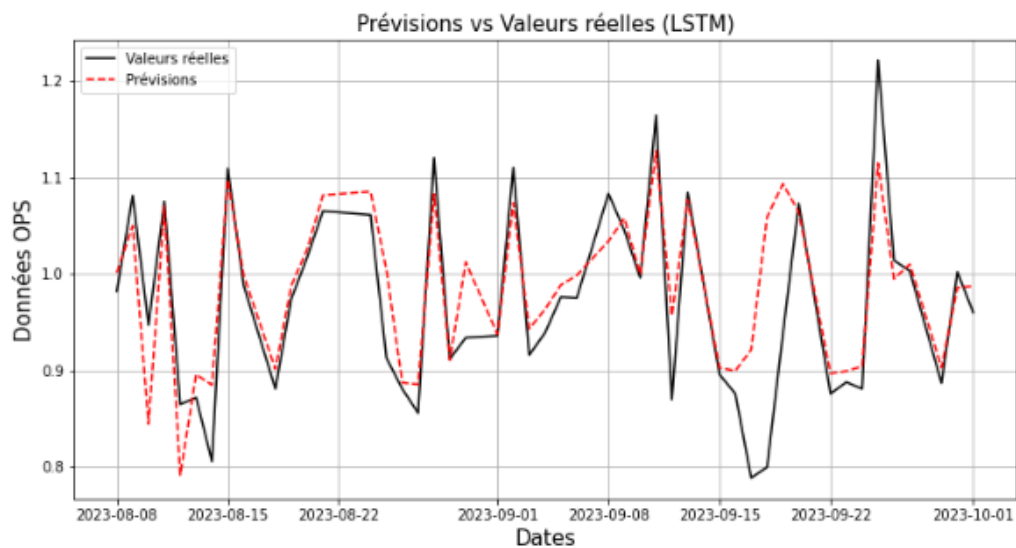


Figure 4.15: Comparaison des prévisions LSTM avec les valeurs réelles (générée avec un script Python développé dans cette étude).

Le graphique (Figure 4.15) compare les prévisions du modèle LSTM avec les valeurs réelles des données OPS. Dans l'ensemble, il parvient à saisir les tendances générales et les fluctuations importantes. Cependant, on observe des divergences, comme des sous-estimations, des surestimations et des retards dans l'ajustement face aux fluctuations soudaines. Bien que prometteur, le modèle pourrait être amélioré en enrichissant les données d'entraînement, en optimisant les hyperparamètres ou en explorant des variantes avancées du LSTM pour mieux gérer les irrégularités et affiner les prévisions.

En ce qui concerne le modèle ARIMA, le graphique ci-dessous offre une comparaison entre les prévisions du modèle et les valeurs réelles des données OPS. Cela permet d'évaluer sa capacité à saisir les dynamiques des séries temporelles observées.

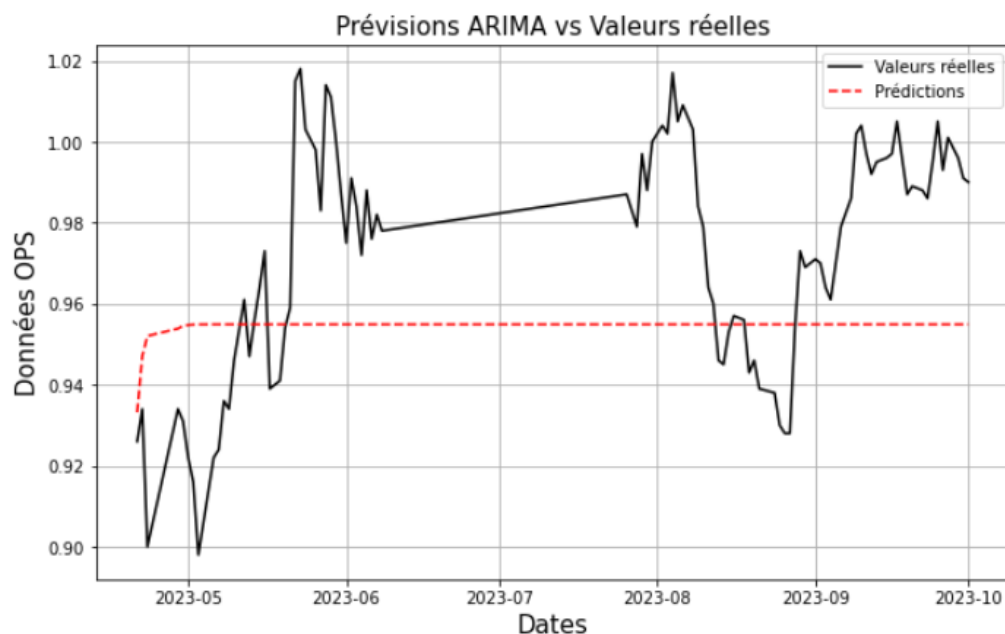


Figure 4.16: Comparaison des prévisions ARIMA avec les valeurs réelles (générée avec un script Python développé dans cette étude).

Les prévisions ARIMA (Figure 4.16) montrent une évolution régulière, sans beaucoup de variations, ce qui indique que le modèle n'a pas réussi à saisir les fluctuations et les tendances observées dans les données réelles. Le modèle ARIMA peut s'avérer approprié pour des séries temporelles à tendance stable. Cependant, son incapacité à s'adapter aux changements soudains ou aux comportements complexes des données OPS limite son efficacité dans ce contexte spécifique. Cela pourrait être dû à des limites dans les paramètres ou au fait que les données ne suivent pas une structure linéaire ou stationnaire compatible avec ARIMA.

Après analyse, le modèle QLSTM se distingue par sa capacité à suivre de près les variations des données réelles, en capturant à la fois les tendances générales et les fluctuations. Le LSTM offre aussi de bons résultats, mais avec une légère perte de précision sur les oscillations rapides. Toutefois, le modèle ARIMA, qui fonctionne bien avec les séries stationnaires, montre ici des limites notables en termes de réactivité face aux variations dynamiques. En bref, le QLSTM s'avère être la méthode la plus efficace pour traiter ces données complexes, devançant légèrement le LSTM, tandis que l'ARIMA se révèle moins approprié dans cette situation.

4.3 Conclusion partielle

Cette section était dédiée à la création et à l'examen du système de prévision. Le processus a commencé par une analyse des données, dans le but d'assurer leur qualité et leur adaptabilité pour le modèle. En effet, les données recueillies sur un site web présentaient une certaine complexité d'utilisation, principalement en raison de

préoccupations quant à leur fiabilité globale. L'analyse préalable a permis de structurer les données, de déterminer une variable clé à prédire pour l'athlète et de les préparer pour leur intégration dans le modèle QLSTM.

Après avoir posé les bases, la modélisation a été abordée par la création du modèle, l'apprentissage, l'optimisation et la détermination des hyperparamètres, puis par les prévisions. Ces phases s'avéraient cruciales, car elles décomposaient le procédé de conception et d'utilisation du modèle. Après l'analyse des résultats, des suggestions ont été formulées pour améliorer les méthodes. Il s'agit notamment d'explorer des méthodes d'optimisation plus avancées, de peaufiner la recherche des hyperparamètres et d'envisager l'utilisation de simulateurs quantiques pour renforcer la fiabilité du modèle.

Finalement, cette section s'est terminée par une comparaison des modèles LSTM, ARIMA et QLSTM. Il en est ressorti que les modèles classiques, comme le LSTM, offrent des avantages, notamment en termes de simplicité d'utilisation et de résultats satisfaisants dans des contextes tels que le sport, mais qu'ils présentent aussi certaines limites. En dépit de son intérêt théorique, le modèle ARIMA s'est avéré moins adapté à cette situation particulière.

Chapitre 5 - CONCLUSION

La prévision à long terme du développement des athlètes de baseball à l'aide de l'intelligence artificielle dans un environnement de l'informatique quantique, a été notre sujet d'étude.

Cette étude a commencé par mettre l'accent sur les spécificités du baseball. Il examine également des études visant à mieux comprendre les athlètes et à identifier les facteurs clés de leur performance. L'utilisation de l'intelligence artificielle dans le domaine du sport fait l'objet d'une étude approfondie.

Ensuite, une revue des systèmes de prévision a été réalisée, en s'appuyant sur une architecture de référence qui guide le développement du système proposé. Les principes et performances des modèles ARIMA, LSTM et QLSTM ont été décrits, de même que leur mise en œuvre.

L'étude s'est conclue par la mise en place du système de prévision, l'analyse des résultats et une comparaison. On a constaté que le QLSTM se distingue par sa capacité à anticiper des séries chronologiques compliquées, mais qu'il nécessite quelques ajustements pour accroître sa fiabilité.

Le système mis au point permettra d'effectuer des prévisions plus précises dans différents domaines, notamment le sportif, grâce à l'exploitation rigoureuse de données soigneusement sélectionnées. En assurant une sélection optimale des données et en minimisant les biais, le modèle pourrait atteindre un niveau de fiabilité élevé, offrant ainsi des applications transversales dans divers domaines qui ont besoin de prévisions complexes.

BIBLIOGRAPHIE

- [1] G. Presto, "Tous les postes de baseball expliqués," in *Nike.com* vol. 2024, N. R. B.V., Ed., ed, 2023. [En ligne], [Disponible]: <https://www.nike.com/ca/fr/a/postes-baseball>.
- [2] MLB. stats [En ligne]. Disponible: <https://www.mlb.com/stats/>
- [3] Tsutsui, T., Maemichi, T., Saeki, J., Torii, S, "Characteristics of upper limb mass, muscle CSA and stiffness in adolescent baseball players with and without elbow injury," *J Biomech.*, 2023 Aug. Epub 2023 Aug 2. [En ligne]. Disponible :PMID <https://pubmed.ncbi.nlm.nih.gov/37549463/>. [Consulté le 04 novembre 2023].
- [4] V. J. Hoffman JR, Pichardo N, Tenenbaum G., "Anthropometric and performance comparisons in professional baseball players," *J Strength Cond Res.*, 2009 Nov;23(8):2173-8., [En ligne]. Disponible: PMID, <https://pubmed.ncbi.nlm.nih.gov/19826310/> . [Consulté le 05 novembre 2023].
- [5] H. J. Mangine GT, Fragala MS, Vazquez J, Krause MC, Gillett J, Pichardo N, "Effect of age on anthropometric and physical performance measures in professional baseball players.," 2013 Feb. [En ligne]. Disponible: PMID, <https://pubmed.ncbi.nlm.nih.gov/21912291/>. [Consulté le 06 novembre 2024]
- [6] C. C. Crotin RL, Szymanski DJ, Oseguera J., "Anthropometric Evaluation of First Round Draft Selections in Major League Baseball," *J Strength Cond Res.*, 2023 Aug. [En ligne]. Disponible: PMID, <https://pubmed.ncbi.nlm.nih.gov/31343601/> [Consulté le 06 novembre 2024].

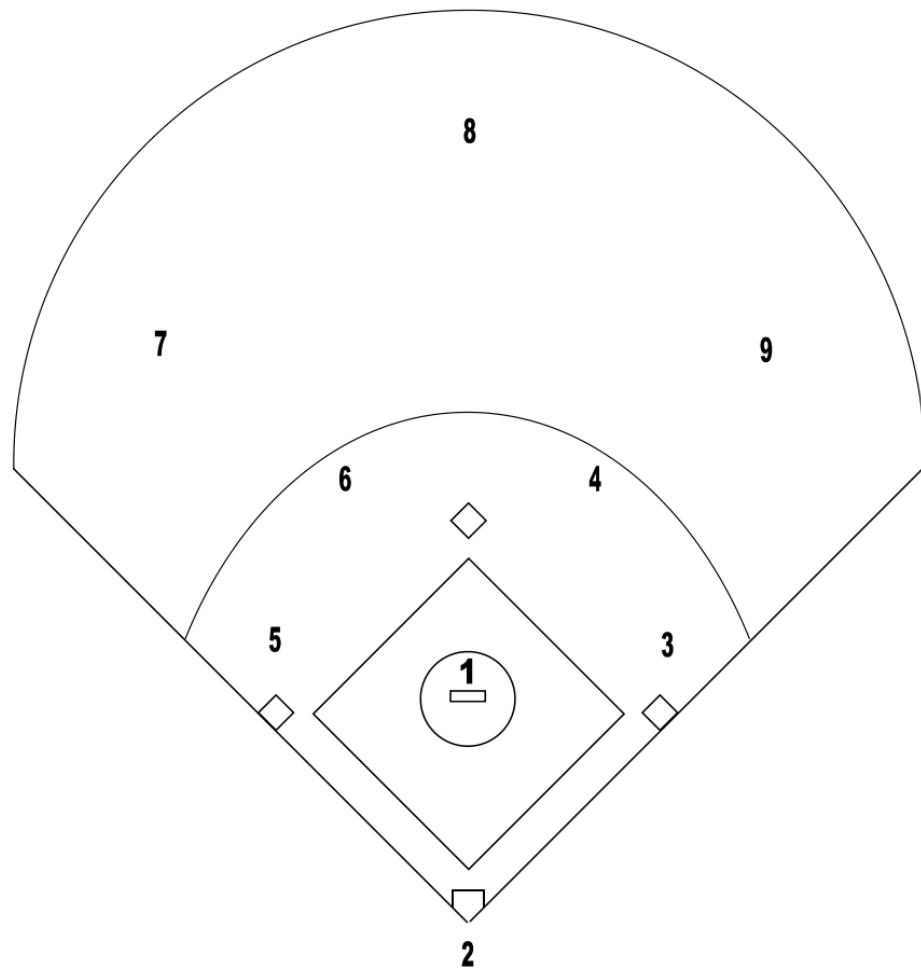
- [7] S. J. Szymanski DJ, Schade RL, Bradford TJ, McIntyre JS, DeRenne C, Madsen NH., "The relation between anthropometric and physiological variables and bat velocity of high-school baseball players before and after 12 weeks of training," *J Strength Cond Res.*, 2010 Nov. [En ligne]. Disponible: PMID, <https://pubmed.ncbi.nlm.nih.gov/20881505/>. [Consulté le 11 novembre 2024]
- [8] L. Guang, "Prediction model and data simulation of sports performance based on the artificial intelligence algorithm," *Computational Intelligence and Neuroscience*, 2022. [EN ligne]. Disponible: PMID, <https://pubmed.ncbi.nlm.nih.gov/36248943/>. [Conculté le 13 novembre 2023]
- [9] H.-C. Sun, T.-Y. Lin, and Y.-L. Tsai, "Performance prediction in major league baseball by long short-term memory networks," *International Journal of Data Science and Analytics*, vol. 15, no. 1, pp. 93-104, 2023/01/01. [En ligne]. Disponible: Springer, <https://link-springer-com/>. [Consulté le 20 Mars 2023]
- [10] A. Aitor, & Gorka, A., "Predicting human behaviour with recurrent neural networks," *Applied Sciences*, 8(2), 305. 2018. [En ligne]. Disponible: EBSCO, <https://web-p-ebscohost-com/>. [Consulté le 11 mars 2024].
- [11] J. Noble, "ARIMA model," in *IBM* vol. 2024, IBM, Ed., ed, 2024. [En ligne]. Disponible: IBM, <https://www.ibm.com/fr-fr/topics/arima-model>. [Consulté le 11 septembre 2024].
- [12] R. Duan, Sun, P., Satapathy, S. C., Agrawal, R., & García Díaz, V, "Basketball sports neural network model based on nonlinear classification," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no, pp. 5917–5926, 2021. [En ligne].

- Disponible: EBSCO, <https://web.p.ebscohost.com/>. [Consulté le 12 décembre 2024].
- [13] D. G. T. a. I. M. Murwantara, " Comparative Study of ARIMA and LSTM in Forecasting Time Series Data," presented at the 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2022. [En ligne]. Disponible: IEEE Xplore, <https://ieeexplore.ieee.org/>. [Consulté le 12 décembre 2024].
- [14] A. Huerga, U. Aguilera, A. Almeida et AB Lago, "A Quantum Computing Approach to Human Behavior Prediction," presented at the 7th International Conference on Smart and Sustainable Technologies (SpliTech), 2022. [En ligne]. Disponible: IEEE Xplore, <https://ieeexplore.ieee.org/>. [Consulté le 20 mars 2024].
- [15] A. Ceschini, A. Rosato and M. Panella, "Design of an LSTM Cell on a Quantum Hardware," pp. 1822 - 1826, 2022,. [En ligne]. Disponible : IEEE Xplore, <https://ieeexplore.ieee.org/>. [Consulté le 15 mars 2024].
- [16] S. Sai, N. Département de génie chimique, Rourkela, Odisha, Inde, and A. S. C. M. A. M. O. S. M. Paul, "QGAPHnet : Quantum Genetic Algorithm Based Hybrid QLSTM Model for Soil Moisture Estimation," presented at the IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 2024. [En ligne], Disponible: IEEE Xplore, <https://ieeexplore.ieee.org/>. [Consulté le 11 decembre 2024]
- [17] L. Franck, *Comprenons-Nous Vraiment La Mécanique Quantique?* (Savoirs Actuels. Physique.). EDP Sciences: CNRS ÉDITIONS, 2018.

- [18] M. Nakahara, Y. Sasaki, I. & Symposium on Quantum, and C. Quantum, *Quantum information and quantum computing*, Singapore: World Scientific, 2013. [Livre électronique]. Disponible: <http://site.ebrary.com/id/10622795>. [Consulté le 01 octobre 2023].
- [19] A. BODIN. *Quantum - Un peu de mathématiques pour l'informatique quantique*, 2024. [En ligne]. Disponible : Exo7, <http://exo7.emath.fr/>. [Consulté le 02 juillet 2024].
- [20] W. P. ELEANOR RIEFFEL, "An Introduction to Quantum Computing for Non-Physicists," vol. 32, no. 3, p. 45, September 2000. [En ligne]. Disponible: Semantics, <https://www.semanticscholar.org/>. [Consulté le 24 février 2023].
- [21] A. G. Josh Patterson, *Deep learning en action* O'REILLEY: Edition First, 2018.[Livre électronique]. Disponible: O'REILLY, <https://www.oreilly.com/>. [Consulté le 12 décembre 2023].
- [22] M. V. S. S. K. K. Agrawal, "Investigation of multiple models of artificial neural networks," presented at the 2017 International Conference on Intelligent Sustainable Systems (ICISS), 2017. [En ligne]. Disponible: IEEE Xplore, <https://ieeexplore.ieee.org/>. [Consulté le 09 septembre 2023].
- [23] J. J. Daudin, Duby, C., Robin, S., & Trécourt, P *Analyse de séries chronologiques* (INA-PG, Mathématiques.). .1996. [Livre électronique]. Disponible: HAL, <https://hal.inrae.fr/>. [Consulté le 04 mars 2023]
- [24] Y. Z. Y. Wang, K. Yan, Y. Liu, R. Huang et T. Wu, "Software Usage Prediction Based on Hybrid LSTM-ARIMA Algorithm," presented at the 2023 IEEE 5th

- International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, 14-16 july 2023, 2023, 699-703. [En ligne]. Disponible : IEEE Xplore, <http://ieeexplore.ieee.org/>. [Consulté le 12 décembre 2024].
- [25] H. W. a. W. Yahui, "Load Forecast of Gas Region Based on ARIMA Algorithm," presented at the 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 2020. [En ligne]. Disponible : IEEE Xplore, <http://ieeexplore.ieee.org/>. [Consulté le 10 juin 2024].
- [26] I. Quantum. Variational Algorithm Design. [En ligne] Disponible: IBM, <https://learning.quantum.ibm.com/>. [Consulté le 05 avril 2024].
- [27] S. Y. -C. Chen, S. Yoo and Y. -L. L. Fang, "Quantum Long Short-Term Memory," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 8622-8626. [En ligne]. Disponible : IEEE Xplore, <http://ieeexplore.ieee.org/>. [Consulté le 28 février 2025].

ANNEXE A – DIAGRAMME DE TERRAIN IMPRIMABLE



ANNEXE B : EXIGENCE CHAMPIONNATS NATIONAUX 2024

Exigences pour les championnats nationaux 2024

Division & Catégorie	Exigences pour 2024	
13U	1 entraîneur est certifié 13U, les autres entraîneurs sont formés 13U	
Coupe Ray Carter 16U Filles	1 entraîneur est certifié 15U, les autres entraîneurs sont formés 15U Lors du championnat national des filles 16U, chaque équipe doit avoir au moins une femme dans son personnel. Si une femme ne fait pas partie de l'équipe d'entraîneurs, au moins un des chefs de mission ou des soigneurs doit être une femme. En 2025, chaque équipe doit avoir au moins une femme dans son équipe d'entraîneurs.	
18U	1 entraîneur est certifié 16+, les autres entraîneurs sont formés 16+	
21U Féminin 22U Masculin	1 entraîneur est certifié 16+, les autres entraîneurs sont formés 16+	
Féminin	1 entraîneur est certifié 16+, les autres entraîneurs sont formés 16+ Lors du championnat national féminin, chaque équipe doit avoir au moins une femme dans son personnel. Si une femme ne fait pas partie de l'équipe d'entraîneurs, au moins un des chefs de mission ou des soigneurs doit être une femme. En 2025, chaque équipe doit avoir au moins une femme dans son équipe d'entraîneurs.	
Masculin	TOUS les entraîneurs ont le module de formation en ligne "Initiation à l'entraînement".	
Coupe Canada *	3 ENTRAÎNEURS SUR L'ÉQUIPE 1 certifié Comp-Dev 1 Entraîneur Comp-Dev 1 entraîneur 16+ certifié (éligible pour une année seulement)	4 ENTRAÎNEURS SUR L'ÉQUIPE 1 certifié Comp-Dev 2 Entraîneur Comp-Dev 1 entraîneur 16+ certifié (éligible pour une année seulement)
Jeux du Canada	Tous les entraîneurs sont certifiés Compétition - Développement	

TOUS les **membres du personnel** qui assistent à un championnat national devront suivre la formation en ligne sur le sport sécuritaire, faire une vérification des antécédents, et signer un code de conduite avant de participer.

* Une province peut désigner un entraîneur comme "entraîneur de développement". Ces entraîneurs ne peuvent être des "entraîneurs de développement" que pendant un an. Un "entraîneur de développement" doit être un entraîneur 16+ certifié. Si l'entraîneur devait revenir pour une 2e année à la Coupe Canada, il devrait satisfaire aux exigences d'entraînement Comp-Dev



ANNEXE C : ENSEMBLES D'ÉTAPES INTUITIFS ET REPRODUCTIBLES POUR LES FLUX DE TRAVAIL QUANTIQUES

