# ***In memoriam***

*A ma mère chérie KAZAYI PELO Géneviève et mon petit frère Philly KAZAYI (Day) qu’une mort précoce a arrachée à notre affection et qui de ce fait, n’a pas pu gouter à leurs saveurs du fruit qu’elle(Géneviève) a semées et entretenu avec beaucoup de sacrifices et d’encadrement spirituel.*

*Je ne vous oublierai jamais ma mère et mon petit frère pour tant de sacrifices et des motivations consentis à mon égard. Je t’aime Maman et je t’aimerai toujours ma meilleure amie.*

**«Hakuna matata »**

**MBUKU KAZONZI Arnauld**

*Big Naulde*

# ***Epigraphe***

***« L’homme est seul devant sa douleur et son chagrin et ne peut sortir tout seul dans une Puit profonde d’épreuves, mais en saisissant une main secourable il y arrivera… »***

Arnauld MBUKUKAZONZI

**Big Naulde**

# ***Remerciements***

A toi Souverain Tout Puissant **Jéhovah**, Dieu d’Amour, Créateur de l’Univers, le Donateur de l’Esprits Saint ; sagesse, intelligence, bénédiction, esprit de savoir faire les choses dont tu m'as doté depuis l'enfance à ce jour malgré mes défauts mais ton Amour m’accompagne toujours.

A toi le fils Unique engendré **Jésus-Christ**, mon Seigneur et Sauveur, tu m'as donné la Connaissance de connaître Le Père et grâce à toi j’ai la priorité de la vie Eternel au monde nouveau, et je fais partie aujourd’hui de la Grande Famille Internationale.

Malgré les situations pathétiques et lugubres que traverse le monde en particulier notre état, nos sincères remerciement vont droit à toutes les autorités Académique de l’Université de Kinshasa(UNIKIN), pour le cadre intellectuel de grande envergure qu’elles nous ont mis à notre disposition durant notre cursus et celles du Département des Mathématiques et Informatique en particulier, qui ont constitués pour nous une bonne école de formation pour la fin du cycle.

Nos distinguées et sincères gratitudes au Professeur **SAINT-JEAN DJUNGU** qui en dépit de ses multiples occupations a accepté de diriger avec amour d’un parent envers son enfant ce travail qui sanctionne la fin de notre premier cycle académique. Aussi une mention particulière à mon Encadreur **MBULAYI ONESIME**.

A mes très chers parents, mon Père **Michel Mbuku Bashiakabua** et maman **Hortance kazayi waku**, pour l’amour et les soucis qu’ils ont toujours manifesté à mon égard et en veillant toujours pour mon avenir.

A ceux qui sont du sang, Frères et Sœurs : Laurette Kazayi , Christian Mbuku, Japhet Mbuku, Kalusia Mbuku, Adrien Bangu, Eloge, Divine, Grace Tshituka, Splaodie, Feston, Jacques Kazayi, Mariella, Consolata, Edwige, José, Briante, Gloire, Esperant, Deo et Prosper Kazayi.

A mes Oncles et Tantes : La famille KAZAYI : Molo, Philly, Aimé, Adrien , Nadine Kiaku, Clémentine, Blanchard. Et la famille MBUKU : Ledoux Kazonzi, Anderson Nzimbu,pour vos conseils et encouragements.

Dans la symphonie de la vie, il y a des ami(e)s dont le cœur et l’esprit ne peuvent oublier à cause de leurs sacrifices, souvenirs et difficultés que nous avons passés ensemble mes collègues de lutte : Deogracias Ngongo, Givny Mumpoto, Gradi Mananga, Kadiebwe, Gustave Muteba, Huguette Kuzengamena, Fortuné Kwabo, Henock Ngalamulume, Japhet Sita, David Mbaya, Khalivanda, Boyongo et Syntiche Ngueze. Sans oublier mon président Blackson GILOLO

Nous ne pouvons clôturer cette page sans remercier Prescillia Gala, mon ami de cœur Pathy Mubiala, et mes chers brothers : Henock Kazadi, Russell Kamwanya, Ladack Kikwiti, et toutes les connaissances, pour vos conseils et encouragements.

Sans oublier toute personne qui nous a apporté de près ou de loin son soutien de quelque forme que c’est soit, trouve ici l'expression de notre reconnaissance.

# ***Répertoire des figures***

[Figure 1 : Représentation graphique de types du Machine Learning 5](#_Toc117032814)

[Figure 2 : Présentation et explication détaillée d’un Dataset pour la location des appartements 7](#_Toc117032815)

[Figure 3 : Dernière convention : pour désigner une cellule de notre tableau, c’est-à-dire que pour désigner la qualité du 3ième appartement que vous avez visité[8]. 7](#_Toc117032816)

[Figure 4 : les étapes de l’apprentissage supervisé sur Machine Learning [3] 9](#_Toc117032817)

[Figure 5: Regroupement et différenciation des données[10] 11](#_Toc117032818)

[Figure 6: Représentation d’un réseau de neurone 15](#_Toc117032819)

[Figure 7: Rôle principale de Deep learning 16](#_Toc117032820)

[Figure 8: Présentation du palmarès de résultat d’examen d’Etat édition (2014) [19] 21](#_Toc117032821)

[Figure 9: Représentation tabulaire du Dataset 24](#_Toc117032822)

[Figure 10. Schéma de recherche 27](#_Toc117032823)

[Figure 11: Représentation graphique d’une série chronologique 30](#_Toc117032824)

[Figure 12: La structure de LSTM classique[39] 47](#_Toc117032825)

[Figure 13: les couches de déclenchement d’une Connexions judas 48](#_Toc117032826)

[Figure 14: schéma d’une Unité récurrente fermée 49](#_Toc117032827)

[Figure 15: La tangente hyperbolique « Tanh » 51](#_Toc117032828)

[Figure 16. Représentation graphique des participants par rapport à l’année et mois 56](#_Toc117032829)

[Figure 17. La seconde prévision pour FacebookProphet 57](#_Toc117032830)

[Figure 18. Evaluation de l'autocorrélation et Représentation des valeurs sur une période 58](#_Toc117032831)

[Figure 19. Schéma des paramètres de prévision et le taux d’erreur(perte) 58](#_Toc117032832)

[Figure 20. La prévision du modèle LSTM sur la colonne « Participants » 59](#_Toc117032833)

[Figure 21. Visualisation de données par différenciation et autocorrélation 59](#_Toc117032834)

[Figure 22. Evolution de la densité 60](#_Toc117032835)

[Figure 23. Disposition de données actuelles et futures 60](#_Toc117032836)

[Figure 24. Différents opération sur les données 61](#_Toc117032837)

# ***Répertoire des tableaux***

[Tableau 1: Code complet d’un participant d’examen d’Etat [19] 23](#_Toc116929163)

[Tableau 2. Métrique d'évaluation sur FacebookProphet/Participants 61](#_Toc116929164)

[Tableau 3. Métrique d'évaluation sur FacebookProphet/Réussite\_Garçon 61](#_Toc116929165)

[Tableau 4. Métrique d'évaluation sur FacebookProphet/Réussite\_Fille 61](#_Toc116929166)

[Tableau 5. Métrique d'évaluation pour LSTM/Participants 62](#_Toc116929167)

[Tableau 6. Métrique d'évaluation pour LSTM/Réussite\_Fille 62](#_Toc116929168)

[Tableau 7. Métrique d'évaluation pour LSTM/Réussite\_Garçon 62](#_Toc116929169)

[Tableau 8. Métrique d'évaluation ARIMA/Participants 62](#_Toc116929170)

[Tableau 9. Métrique d'évaluation ARIMA/Réussite\_Fille 63](#_Toc116929171)

[Tableau 10.Métrique d'évaluation ARIMA/Réussite\_Garçon 63](#_Toc116929172)

# ***Répertoire des Acronymes***

|  |  |
| --- | --- |
| **A** |  |
| **ARIMA** | : Auto-Regressive Integrated Moving Average |
| **ADF** | : Augmented Dickey-Fuller |
| **AR** | : Auto-Régression |
| **ARMA** | : Auto-Regressive Moving Average |
| **E** |  |
| **EPST** | : Enseignement primaire secondaire et technique |
| **G** |  |
| **GRU** | : Gated Recurrent Unit |
| **I** |  |
| **IA** | : Intelligence artificielle |
| **K** |  |
| **KPSS** | : Kwiatkowski-Phillips-Schmidt-Shin |
| **KDE** | : Kernel Density Estimate |
| **L** |  |
| **LSTM** | : Long short terme memory |
| **M** |  |
| **MSE** | : Erreur quadratique moyenne |
| **MCO** | : Moindres Carrés Ordinaires |
| **MA** | : Moving Average |
| **MAE** | : Erreur Absolue Moyenne |
| **MLR** | : Régression Linéaire Multiple |
| **N** |  |
| **NME** | : Erreur Moyenne Normalisée |
| **NRMSE** | : Erreur quadratique moyenne normalisée |
| **R** |  |
| **RMSE** | : Erreur quadratique moyenne racine |
| **RNN** | : Réseau de Neurones récurrents |
| **RDC** | : République Démocratique du Congo |
| **S** |  |
| **SVM** | : machine à vecteur de support |
| **SARIMA** | : Seasonal AutoRegressive Integrated Moving Average |
| **SMA** | : Simple Moving Average |
| **V** |  |
| **VARIMA** | : Vector AutoRegressive Integrated Moving Average |
| **VECM** | : Vector Error Correction Model |
| **VMA** | : Vector Moving Average |
| **VAR** | : Vector Autoregression |
| **VARMA** | : Vector AutoRegressive Moving Average |
| **W** |  |
| **WMA** | : Weighted Moving Average |

# ***INTRODUCTION GENERALE***

L’éducation est une valeur intrinsèque qui participe des droits, des libertés et des capabilités des personnes qui sont supérieurs à toute conception utilitariste ou de logiques marchandes. Un système éducatif est équitable quand les contractants peuvent y souscrire sans savoir à l’avance quel bénéfice personnel ils pourront en tirer. L’inégalité qui peut se justifier pour des raisons d’efficacité n’est justifiable que sous deux conditions : elle doit préserver l’égalité des chances et prendre en compte la situation des plus défavorisés.

Le système éducatif en République Démocratique du Congo est placé sous la tutelle de trois ministères : le ministère de l’Enseignement primaire, secondaire et technique ; le ministère de l’Enseignement supérieur et Universitaire ; et le ministère des Affaires sociales, action humanitaire et solidarité nationale. La République Démocratique du Congo a réalisé des avancées significatives vers l’accès universel à l’enseignement primaire au cours de ces dernières décennies.

Dans le domaine de l’éducation, il est possible d’automatiser certaines activités, notamment en matière des corrections d’évaluations. L’intelligence artificielle (IA) permet en outre de personnaliser les apprentissages selon les besoins et les capacités d’assimilation de chacun[1].

En se basant sur un triangle pédagogique (l’enseignant, l’apprenant et le savoir), il est possible de catégoriser les impacts de l’IA sur l’éducation. L’IA assure le développement de plusieurs technologies qui sont susceptibles de remplacer des tâches prévisibles et répétitives des professeurs. Il s’agit par exemple d’automatiser certaines tâches avec peu de valeur ajoutée, comme les corrections d’examens qui consomment beaucoup du temps des enseignants. Néanmoins, le travail de l’enseignant dépasse largement le spectre des tâches automatisables[1].

Le système pédagogique en république Démocratique du Congo a perdu son potentialité en particulier l’Enseignement primaire, secondaire et technique. Quatre problèmes importants ont été identifiés : (i) une couverture relativement faible au niveau primaire, avec de grandes inégalités dans l’accès, et une expansion incontrôlée aux niveaux secondaire et supérieur (ii) une grave détérioration de la qualité de l’éducation à tous les niveaux (iii) un système d’administration scolaire lourd et désuet et (iv) un très bas niveau de dépenses et un système de financement inefficace et inéquitable[2]. La ville province de Kinshasa compte environ 25 000 écoles et elle ne dispose pas un outil informatique pour garder plus haut son système pédagogique. Sur ce, il manque les différents mesures d’évaluation informatique au ministère de l’Enseignement primaire, secondaire et technique de la ville province de Kinshasa. Quel mécanisme informatique allons-nous utilisés ? et quel sera le résultat attendu ?

Idée est de mettre en place une application auto-régression de la série chronologique dans le domaine de Machine Learning enfin de résoudre ce problème. Nous allons faire cette étude partant de résultat d’examen d’Etat que le ministère de l’Enseignement primaire, secondaire et technique publie chaque année. Allons faire une étude comparative sur les modèles auto-régression, pour une période donnée, nous allons prédire le résultat et définir la performance d’une école. Alors comment allons-nous parvenir ?

Comme il est question de conception d’une application d’Intelligence Artificielle, où le Machine Learning (apprentissage automatique) est un domaine scientifique, et plus particulièrementune sous-catégorie de l’intelligence artificielle. Elle consiste à laisser des algorithmes découvrir des ” patterns “, à savoir des motifs récurrents, dans les ensembles de données. Ces données peuvent être des chiffres, des mots, des images, des statistiques…

Le machine learning permet de développer, de tester et d’appliquer des algorithmes d’analyse prédictive sur différents types de données afin de prédire le futur. En automatisant le développement de modèle analytique, le Machine Learning permet aussi d‘accélérer l’analyse de données et de la rendre plus précise.

En ce qui concerne notre étude, nous allons utiliser différents algorithmes d’auto régression d’une séries chronologique (Time Series). Partant de données chronologique (ou périodique) nous pouvons prédire la courbe pour chaque école avec ces 3 algorithmes : ARIMA, FACEBOOKPROPHET et LSTM ainsi que la comparaison des algorithmes pour que nous ayons un algorithme optimal. Excepté l’introduit et la conclusion, ce travail est subdivisé en trois (3) chapitres :

* Chapitre 1. Contexte théorique
* Chapitre 2 : prétraitement et expérimentation
* Chapitre 3 : Résultat et Discussion

# ***CHAPITRE I. CONTEXTE THEORIQUE***

## ***I.1 Introduction***

L’étude scientifique des algorithmes et des modèles statistiques que les ordinateurs utilisent pour accomplir une tâche sans instruction explicite, mais plutôt en s’appuyant sur des motifs et de l’inférence[3]. Il est massivement utilisé pour la [Data Science](https://datascientest.com/data-science-definition) et l’analyse de données. Le machine Learning permet de développer, de tester et d’appliquer des algorithmes d’analyse prédictive sur différents types de données afin de prédire le futur. En automatisant le développement de modèle analytique, le Machine Learning permet d‘accélérer l’analyse de données et de la rendre plus précise. Il permet d’assigner aux machines des tâches au cœur de l’analyse de données comme la classification, le Clustering ou la détection d’anomalie.

Pour l’heure, un humain reste toutefois nécessaire pour passer en revue les résultats des analyses produites par les algorithmes de Machine Learning. Son rôle est de donner du sens à ces résultats, ou encore de s’assurer que les données traitées par l’algorithme ne soient ni biaisées ni altérées.

De manière générale, les [algorithmes de Machine Learning](https://datascientest.com/machine-learning-clustering-focus-sur-algorithme-cah)apprennent de manière autonome à effectuer une tâche ou à réaliser des prédictions à partir de données et améliorent leurs performances au fil du temps. Une fois entraîné, l’algorithme pourra retrouver les patterns dans de nouvelles données. Nous pouvons maintenant expliciter et interpréter ce qui est implicite. Tout au long du chapitre nous allons parler du machine Learning et tous qui fait son existence en informatique.

## ***I.2. Définitions :***

Def 1 : Le Machine Learning ou apprentissage automatique est un domaine scientifique, et plus particulièrementune sous-catégorie de l’intelligence artificielle. Elle consiste à laisser des algorithmes découvrir des ” patterns “, à savoir des motifs récurrents, dans les ensembles de données[4].

Def 2 : Le machine Learning est un modèle d’intelligence artificielle qui apprend seul grâce à l’analyse d’un jeu de données[5].

Def 3 : Le machine Learning, ou apprentissage automatique, est la principale technique de l'IA. Il consiste à entrainer des algorithmes à partir de base d'apprentissage pour leur permettre de faire des prédictions ou automatiser des tâches[6].

Def 4 : Le machine learning est une technique de programmation informatique qui utilise des probabilités statistiques pour donner aux ordinateurs la capacité d'apprendre par eux-mêmes sans programmation explicite[7].

## ***I.3. Différents types de machine learning***

On distinguetrois techniques de Machine Learning :

1. L’apprentissage supervisé,
2. L’apprentissage non-supervisé, et
3. L’apprentissage par renforcement.

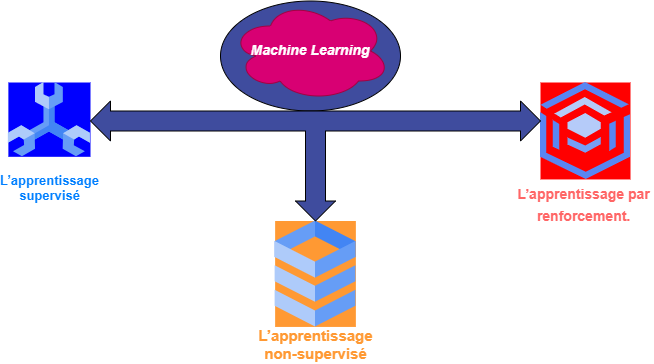


Figure 1 : Représentation graphique de types du Machine Learning

### ***I.3.1 L’apprentissage supervisé***

L’apprentissage supervisé (en anglais : Supervised Learning) est le paradigme d’apprentissage le plus populaire en Machine Learning et en Deep Learning. Comme son nom l’indique, cela consiste à superviser l’apprentissage de la machine en lui montrant des exemples (des données) de la tâche qu’elle doit réaliser[8]. Les applications sont nombreuses : Reconnaissance vocale, vision par ordinateur, régressions, classifications… La grande majorité des problèmes de Machine Learning et de Deep Learning utilisent l’apprentissage supervisé. Dans le cas de l’apprentissage supervisé, le plus courant, les données sont étiquetées afin d’indiquer à la machine quels patterns elle doit rechercher. L’apprentissage supervisé fonctionne en 4 étapes :

1. Importer un Dataset (x, y) qui contient nos exemples ;
2. Développer un Modèle aux paramètres aléatoires ;
3. Développer une Fonction Coût qui mesure les erreurs entre le modèle et le Dataset ;
4. Développer un Algorithme d’apprentissage pour trouver les paramètres du modèle qui minimisent la Fonction Coût.
5. *Le Dataset*

Est un ensemble cohérent de données produites dans le cadre d’un même projet, sur un même objet d’étude et/ou recueillies sur un même lieu. Le dataset peut être représenté sous différents types, que ce soient des tableaux, des graphes, des arbres ou autres. On travaille souvent avec des structures de tableaux dans les algorithmes de machine learning[8].

La première étape d’un algorithme de Machine Learning consiste donc à importer un Dataset qui contient les exemples que la machine doit étudier. Ce Dataset inclut toujours 2 types de variables : Une variable objectif (target) y, et une ou plusieurs variables caractéristiques (features) x.

Par exemple, imaginez que vous visitez une série d’appartements un samedi après-midi. Pour chaque appartement, vous notez dans un tableau Excel le prix y et les caractéristiques x de l’appartement (la surface, la qualité, ville, etc.)

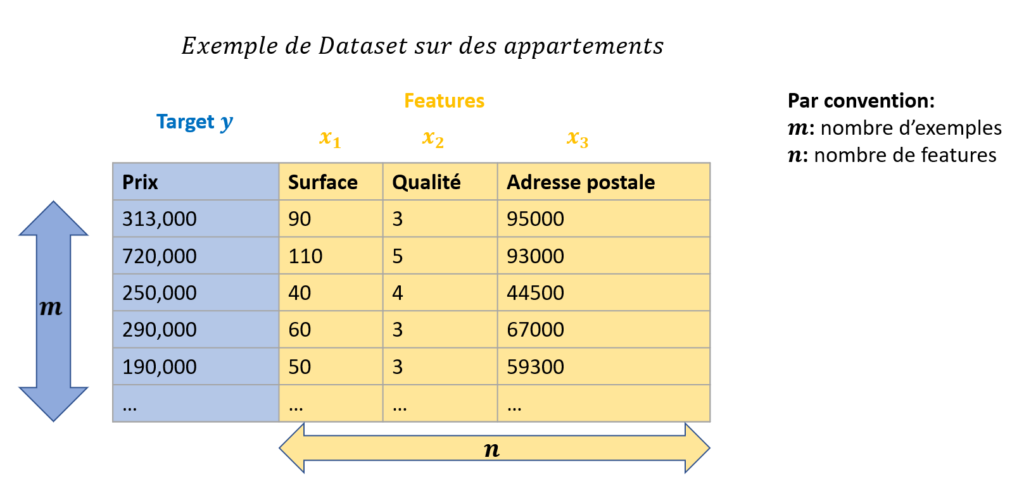


Figure 2 : Présentation et explication détaillée d’un Dataset pour la location des appartements

Par convention, on dit que notre Dataset contient m exemples, c’est-à-dire m lignes. Si vous avez visité 6 appartements, alors m=6. Par convention, on note également n le nombre de future dans notre Dataset, c’est-à-dire le nombre de colonnes (hormis la colonne y). Si vous avez noté 3 caractéristiques pour vos appartement (Surface, qualité, ville), alors n=3

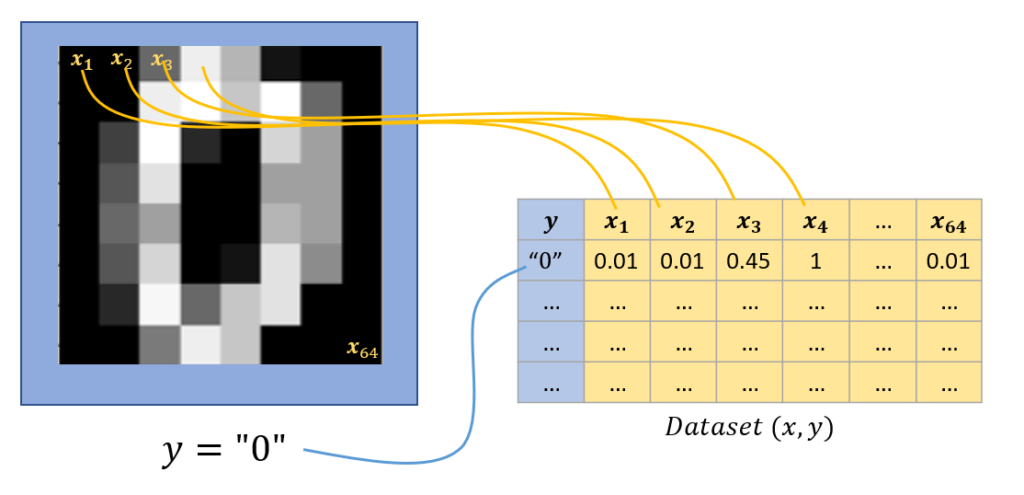


Figure 3 : Dernière convention : pour désigner une cellule de notre tableau, c’est-à-dire que pour désigner la qualité du 3ième appartement que vous avez visité[8].

Avec un tel Dataset, il devient possible de prédire de nouvelles valeurs y à partir de valeurs de x en développant un modèle, et c’est là notre 2ième étape dans la résolution d’un problème de Supervised Learning (ou apprentissage supervisé).

1. *Le modèle*

Le cœur de notre programme. Un modèle est une représentation simplifiée de la réalité, que l’on peut utiliser pour prédire ce qui se passerait dans certaines conditions. Ça peut être un dessin, une équation physique, une fonction mathématique, une courbe… bref, n’importe quelle représentation. Le modèle est en quelque sorte le cœur de notre programme, c’est lui qui va effectuer la tâche que nous cherchons à accomplir, par exemple reconnaître un animal sur une photo ou prédire le prix d’un appartement[8]***.***

Par exemple, si je lâche une pomme de 100 g depuis une hauteur de 4 mètres, en combien de temps tombera t’elle sur la tête de ce bon vieux Newton ? On peut prédire cela avec les équations de Newton. Dans ce cas, notre modèle est déterministe : il donnera toujours la même réponse sous les mêmes conditions.

Les modèles de Machine Learning, à la différence du modèle illustré plus haut, un modèle de Machine Learning ne repose pas sur une démonstration mathématique ou une équation physique[8]. A la place, il est construit à partir de données, comme un modèle statistique. Maintenant, la question est : comment faire pour que la machine trouve le meilleur modèle ? Autrement dit, comment faire pour que la machine apprenne ?

1. ***La Fonction coût*** ***ou Mesure de la performance***

Pour que la machine trouve le meilleur modèle, il faut déjà qu’elle puisse mesurer la performance d’un modèle donné. Il existe beaucoup de métriques pour mesurer nos erreurs (MSE, MAE, RMSE, etc.) et nous verrons l’un d’entre eux dans le prochain chapitre de ce travail, ce qui nous permettra de passer au concret avec cette histoire de Fonction Coût.

1. ***L’Algorithme d’apprentissage***

Sur ce point, la machine cherche les paramètres de modèle qui minimisent la Fonction Coût. C’est ça qu’on appelle l’apprentissage. Cette phase est très importante pour trouver les paramètres qui minimisent la fonction Coût, et un paquet de stratégies. Une autre stratégie, très populaire en Machine Learning, est de considérer la Fonction Coût comme une **fonction convexe**, c’est-à-dire une fonction qui n’a qu’un seul minimum, et de chercher ce minimum avec un algorithme de minimisation appelé Gradient Descent. Cette stratégie apprend de façon graduelle, et assure de converger vers le minimum de la fonction Coût (si convexe).

Le système s’entraîne sur un ensemble de données étiquetées, avec les informations qu’il est censé déterminer. Les données peuvent même être déjà classifiées de la manière dont le système est supposé le faire. Cette méthode nécessite moins de données d’entraînement que les autres, et facilite le processus d’entraînement puisque les résultats du modèle peuvent être comparés avec les données déjà étiquetées. Cependant, l’étiquetage des données peut se révéler onéreux. Un modèle peut aussi être biaisé à cause des données d’entraînement, ce qui impactera ses performances par la suite lors du traitement de nouvelles données.

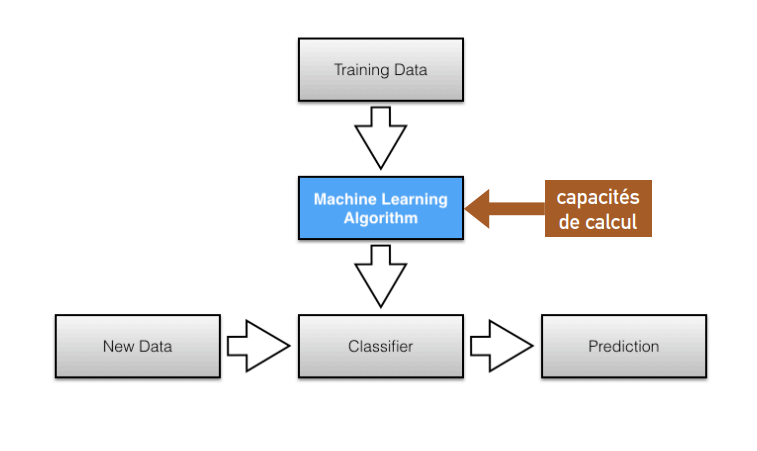
******

Figure 4 : les étapes de l’apprentissage supervisé sur Machine Learning [3]

### ***I.3.2 L’apprentissage non-supervisé***

L'objectif de l'apprentissage non supervisé est de modéliser la structure ou la distribution sous-jacente dans les données afin d'en apprendre davantage sur les données. On l'appelle apprentissage non supervisé car, contrairement à l'apprentissage supervisé ci-dessus, il n'y a pas de réponse correcte ni d'enseignant. Les algorithmes sont laissés à leurs propres mécanismes pour découvrir et présenter la structure intéressante des données.

L'apprentissage non supervisé détecte des données ou individus présentant des caractéristiques ou des structures communes. Typiquement, l'unsupervised learning peut servir à développer un moteur de recommandation de produits, qui est conçu pour proposer à un visiteur des produits au regard des appétences des clients présentant les mêmes caractéristiques[9]. Dans le cas del’apprentissage non supervisé, les données n’ont pas d’étiquettes. La machine se contente d’explorer les données à la recherche d’éventuelles patterns. Elle ingère de vastes quantités de données, et utilise des algorithmes pour en extraire des caractéristiques pertinentes requises pour étiqueter, trier et classifier les données en temps réel sans intervention humaine.

Plutôt que d’automatiser les décisions et les prédictions, cette approche permet d’identifier les patterns et les relations que les humains risquent de ne pas identifier dans les données. L'apprentissage non supervisé comprend deux catégories d’algorithmes :

* Regroupement ou Clustering :

La mise en cluster consiste à séparer ou à diviser un ensemble de données en un certain nombre de groupes, de sorte que les ensembles de données appartenant aux mêmes groupes se ressemblent davantage que ceux d'autres groupes. En termes simples, l’objectif est de séparer les groupes ayant des traits similaires et de les assigner en grappes. Cette technique peut sembler proche de celle de la classification dans l’apprentissage supervisé, mais à la différence de cette dernière, les classes ne sont pas pré-remplies par un humain, c’est la machine qui “invente” ses propres classes, à un niveau de finesse pas toujours évident pour un humain[5].

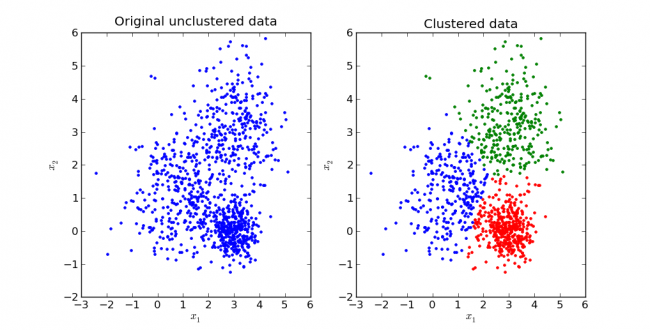


Figure 5: Regroupement et différenciation des données[10]

*.*

Le clustering est une technique d’apprentissage non supervisé, qui groupe des données non-étiquetées en fonction de leurs similarités and de leurs différences. Ainsi, les points sont rassemblés dans des groupes (ou des clusters) de telle manière à ce que les points au sein d’un même groupe soient le plus similaire possible, pendant que les points dans des groupes différents soient les plus différents possible. Pour cela, une analyse de cluster identifie les caractéristiques, au sein des données, et groupe les points en fonction de la présence ou de l’absence de ces caractéristiques. Parmi les méthodes de clustering, on trouve la méthode des k-moyennes (ou k-means), la classification hiérarchique (ou hierarchical clustering) ou encore la classification probabilistique (ou probabilistic clustering)[10].

* Association :

L'association consiste à découvrir des relations intéressantes entre des variables dans de grandes bases de données. Par exemple, les personnes qui achètent une nouvelle maison ont aussi tendance à acheter de nouveaux meubles. Il découvre la probabilité de cooccurrence d'éléments dans une collection.

En résumé, le clustering consiste à grouper des points de données en fonction de leurs similitudes, tandis que l'association consiste à découvrir des relations entre les attributs de ces points de données. Voici certains algorithmes d'apprentissage automatique non supervisés :

* K-means clustering ;
* Dimensionality Reduction (Réduction de la dimensionnalité) ;
* Neural networks / Deep Learning ;
* Principal Component Analysis (Analyse des composants principaux) ;
* Singular Value Decomposition (Décomposition en valeur singulière) ;
* Independent Component Analysis (Analyse en composantes indépendantes) ;
* Etc...

### ***I.3.3 L’apprentissage par renforcement***

L’[apprentissage par renforcement](https://it.toolbox.com/article/openais-robot-learns-to-solve-a-rubiks-cube-with-one-hand-peter-welinder-research-lead-openai-shares-insights) s'inspire directement de la façon dont les êtres humains apprennent des données dans leur vie. Il comporte un algorithme qui s'améliore et apprend de nouvelles situations en utilisant une méthode d'essais et d'erreurs. Basé sur le concept psychologique de conditionnement, l'apprentissage par renforcement fonctionne en plaçant l'algorithme dans un environnement de travail avec un interprète et un système de récompense. À chaque itération de l'algorithme, le résultat de sortie est donné à l'interpréteur, qui décide si le résultat est favorable ou non. Cette technique est notamment utilisée pour permettre à une IA de surpasser les humains dans les jeux.

Enfin,l’apprentissage par renforcement consiste à laisser un algorithme apprendre de ses erreurs pour atteindre un objectif. L’algorithme essayera de nombreuses approches différentes pour tenter d’atteindre son but.

## ***I.4. Fonctionnement du machine learning***

Le développement d’un modèle de Machine Learning repose sur quatre étapes principales. En règle générale, [c’est un Data Scientist](https://datascientest.com/formation-data-scientist) qui gère et supervise ce procédé[4].

La première étape consiste à sélectionner et à préparer un ensemble de données d’entraînement. Ces données seront utilisées pour nourrir le modèle de Machine Learning pour apprendre à résoudre le problème pour lequel il est conçu. Les données peuvent être étiquetées, afin d’indiquer au modèle les caractéristiques qu’il devra identifier. Elles peuvent aussi être non étiquetées, et le modèle devra repérer et extraire les caractéristiques récurrentes de lui-même. Dans les deux cas, les données doivent être soigneusement préparées, organisées et nettoyées. Dans le cas contraire, l’entraînement du modèle de Machine Learning risque d’être biaisé. Les résultats de ses futures prédictions seront directement impactés.

La deuxième étape consiste à sélectionner un algorithme à exécuter sur l’ensemble de données d’entraînement. Le type d’algorithme à utiliser dépend du type et du volume de données d’entraînement et du type de problème à résoudre.

La troisième étape est l’entraînement de l’algorithme. Il s’agit d’un processus itératif. Des variables sont exécutées à travers l’algorithme, et les résultats sont comparés avec ceux qu’il aurait dû produire. Les ” poids ” et le biais peuvent ensuite être ajustés pour accroître la précision du résultat. On exécute ensuite de nouveau les variables jusqu’à ce que l’algorithme produise le résultat correct la plupart du temps. L’algorithme, ainsi entraîné, est le modèle de Machine Learning.

La quatrième et dernière étape est l’utilisation et l’amélioration du modèle. On utilise le modèle sur de nouvelles données, dont la provenance dépend du problème à résoudre. Par exemple, un modèle de Machine Learning conçu pour détecter les spams sera utilisé sur des emails.

De son côté, le modèle de Machine Learning d’un aspirateur robot ingère des données résultant de l’interaction avec le monde réel comme le déplacement de meubles ou l’ajout de nouveaux objets dans la pièce**[4]**. L’efficacité et la précision peuvent également s’accroître au fil du temps.

## ***I.5. Les principaux algorithmes de machine Learning***

Les [algorithmes de Machine Learning](https://datascientest.com/machine-learning-clustering-focus-sur-algorithme-cah)apprennent de manière autonome à effectuer une tâche ou à réaliser des prédictions à partir de données et améliorent leurs performances au fil du temps[4]. Une fois entraîné, l’algorithme pourra retrouver les patterns dans de nouvelles données. Il existe une large variété d’algorithmes de Machine Learning. Certains sont toutefois plus couramment utilisés que d’autres. Tout d’abord, différents algorithmes sont utilisés pour les données étiquetées.

* **Les algorithmes de régression**
* **Linéaire**: Permettent de comprendre les relations entre les données. La régression linéaire est utilisée pour prédire la valeur d’une variable dépendante base sur la valeur d’une variable indépendante. Il s’agirait par exemple de prédire les ventes annuelles d’un commercial en fonction de son niveau d’études ou de son expérience.
* **Logistique**: La régression logistique est quant à elle utilisée quand les variables dépendantes sont binaires. Un autre type d’algorithme de régression appelé machine à vecteur de support est pertinent quand les variables dépendantes sont plus difficiles à classifier**[4]**.
* **L’**[**arbre de décision**](https://datascientest.com/random-forest-definition)

Cet algorithme permet d’établir des recommandations basées sur un ensemble de règles de décisions en se basant sur des données classifiées. Par exemple, il est possible de recommander sur quelle équipe de football parier en se basant sur des données telles que l’âge des joueurs ou le pourcentage de victoire de l’équipe[4].

* **Algorithmes d’association**

Les algorithmes d’association permettent quant à eux de découvrir des patterns et des relations dans les données, et à identifier les relations ” si / alors ” appelées ” règles d’association “. Ces règles sont similaires à celles utilisées dans le domaine du Data Mining ou forage de données[4].

* **Réseaux de neurones**

Les réseaux de neurones sont des algorithmes se présentant sous la forme d’un réseau à plusieurs couches. La première couche permet l’ingestion des données, une ou plusieurs couches cachées tirent des conclusions à partir des données ingérées, et la dernière couche assigne une probabilité à chaque conclusion.

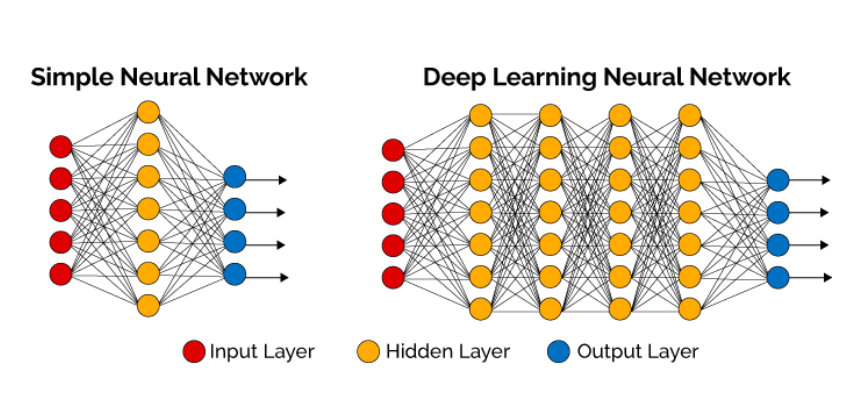


Figure 6: Représentation d’un réseau de neurone

* La couche rouge : permet l’ingestion des données,
* La couche jaune : une ou plusieurs couches cachées tirent des conclusions à partir des données ingérées, et
* La dernière couche Blue : assigne une probabilité à chaque conclusion [3]

## ***I.6. Deep learning***

Le [Deep Learning](https://www.youtube.com/watch?v=G7WxP0dcPC8&list=PLbH8UGHWFlsTo6cG8Fgbgz0MWT7X4TGTW&index=8)est une branche du Machine Learning, mais il s’agit aujourd’hui de la plus couramment utilisée. Il s’agit d’une invention de Geoffrey Hinton, datée de 1986[4]. Cette technique est appelée réseau de neurones profond. Cette profondeur correspond au large nombre de couches de nœuds de calcul qui constituent ces réseaux et travaillent en collaboration pour traiter les données et délivrer des prédictions. Ces réseaux de neurones s’inspirent directement du fonctionnement du cerveau humain. Les nœuds de calcul sont comparables aux neurones, et le réseau en lui-même s’apparente au cerveau.

Le Deep learning, le système identifie lui-même les caractéristiques discriminantes des données, sans avoir besoin d’une catégorisation préalable. Le système n’a pas besoin d’être entraîné par un développeur. Il évalue lui-même le besoin de modifier le classement ou de créer des catégories inédites en fonction des nouvelles données.

***I.6.1 Définitions***

Le *deep learning* ou apprentissage profond est un type d'intelligence artificielle dérivé du *machine learning* (apprentissage automatique) où la machine est capable d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter à la lettre des règles prédéterminées[11].

### ***I.6.2 Fonctionnement***

Le *Deep Learning* s'appuie sur un réseau de [neurones artificiels](https://www.futura-sciences.com/tech/actualites/intelligence-artificielle-synapse-artificielle-reproduire-fonctionnement-neurones-34674/) s'inspirant du [cerveau](https://www.futura-sciences.com/sante/definitions/biologie-cerveau-3125/) humain. Ce réseau est composé de dizaines voire de centaines de « couches » de [neurones](https://www.futura-sciences.com/sante/definitions/biologie-neurone-209/), chacune recevant et interprétant les informations de la couche précédente. Le système apprendra par exemple à reconnaître les lettres avant de s'attaquer aux mots dans un texte, ou détermine s'il y a un visage sur une photo avant de découvrir de quelle personne il s'agit[11].

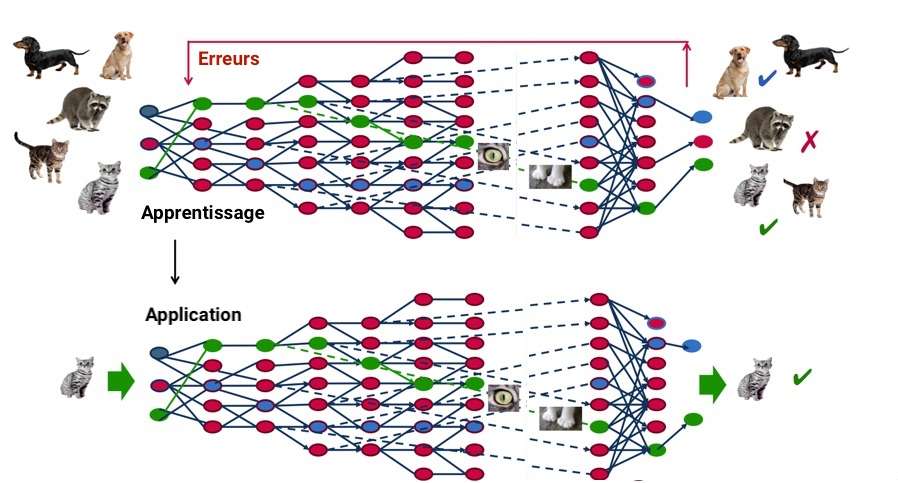
******

Figure 7: Rôle principale de Deep learning

À travers un processus d’autoapprentissage, le deep Learning est capable d’identifier un chat sur une photo. À chaque couche du réseau neuronal correspond un aspect particulier de l’image [11].

***I.6.4 Les différences entre le Deep learning et le Machine learning***

Le deep learning constituant un sous-domaine du machine learning, il convient de connaître les différences entre ces deux programmes d'intelligence artificielle. Le machine learning ou l'apprentissage automatique est une technologie permettant à un système d'évoluer à partir d'un retour humain. Cela implique donc de fournir des données organisées. À partir de celles-ci, l'ordinateur est capable de comprendre ces données, de les catégoriser et d'exécuter des actions programmées. Les algorithmes sont optimisés grâce aux développeurs et les bases de données sont contrôlables.

Le deep learning, quant à lui, fonctionne à partir de données non structurées. En effet, ces algorithmes s'appuient sur un réseau neuronal de plusieurs couches, similaire au cerveau humain. Un système avec un apprentissage profond est capable d'effectuer des tâches complexes alors qu'avec le machine learning il est plutôt question d'actions de routine[12]. Le Machine learning (apprentissage automatique) et le Deep learning (apprentissage profond) sont les deux concepts les plus importants qui rendent l’intelligence artificielle possible. On confond bien souvent ces deux termes, alors qu’ils désignent deux méthodes bien distinctes employées dans des champs d’application différents[11].

## ***I.7. Séries chronologiques (ou Time series)***

Une série chronologique est une séquence d’observations sur une certaine période. Une série chronologique univariées se compose des valeurs prises par une seule variable à des instances temporelles périodiques sur une période, et une série chronologique multivariée se compose des valeurs prises par plusieurs variables dans les mêmes instances temporelles périodiques sur une période. L’exemple le plus simple d’une série chronologique que nous rencontrons tous au jour le jour est le changement de température tout au long de la journée, de la semaine, du mois ou de l’année.

L’analyse des données temporelles est capable de nous donner des informations utiles sur la façon dont une variable change au fil du temps, ou comment elle dépend de la variation des valeurs d’autres variables. Les séries chronologiques sont également utilisées dans plusieurs contextes non financiers, comme la mesure de l’évolution de la population au fil du temps.

### ***I.7.1 Définition***

* Les time séries (que l'on traduit par séries chronologiques ou séries temporelles) se présentent sous la forme d'une suite de valeurs numériques correspondant à l'évolution d'une variable dans le temps[14].
* En [mathématiques](https://en.wikipedia.org/wiki/Mathematics) , une série chronologique est une série de [points de données](https://en.wikipedia.org/wiki/Data_point) indexés (ou répertoriés ou représentés graphiquement) dans l'ordre du temps[15].

### ***I.7.2 Modélisation Séries chronologiques***

Eléments constitutifs d’une série temporelle Une série temporelle est généralement constituée de plusieurs éléments[16] :

* **Tendance** : représente l’évolution à long terme de la série (échelle interannuelle). Ou le comportement croissant ou décroissant d’une variable avec le temps. Exemples : croissance économique, évolution climatologique à long terme (cyclique ou non)
* **Saisonnalité** : évolution se répétant régulièrement tous les ans. Exemples : En météorologie, température plus faibles en hiver qu’en été. En économie, saisonnalité induite par les périodes de vacances, les périodes de fêtes, le climat...
* **Composante stationnaire (ou résiduelle)** : ce qui reste lorsque l’on a enlevé les autres composantes. Décrit l’évolution à court terme de la série (échelle journalière).

Pour capturer ces composants, il existe un certain nombre d’algorithmes de modélisation de séries chronologiques populaires. Le chapitre 3 explique d’une manière détaillée les 3 algorithmes de séries chronologiques utiliser durant notre étude comparative.

## ***I.8 conclusion***

Les modèles structurels de séries chronologiques sont formulés en termes de composants, tels que les tendances, les saisons et les cycles, qui ont une interprétation directe. En plus de fournir un cadre pour la décomposition de séries chronologiques par extraction de signaux, ils peuvent être utilisés pour la prévision et pour la « prévision immédiate ». L'interprétation structurelle permet des extensions à des classes de modèles capables de traiter divers problèmes dans des séries multivariées et de faire face à des observations non gaussiennes et à des modèles non linéaires. Le traitement statistique se fait par la forme de l'espace d'état et, par conséquent, les irrégularités des données telles que les observations manquantes sont facilement traitées. Les modèles à temps continu offrent une plus grande flexibilité dans la mesure où ils peuvent gérer des espacements irréguliers[17]. Enfin nous venons de découvrir les notions importantes pour aboutir à un bon résultat. Durant ce chapitre nous avons vu l’importance de la série chronologique sur le Machine Learning et deep learning pour la prévision des modèles structurels de séries chronologiques. Le chapitre qui suive, parlera d’une manière claire la disposition de notre Dataset.

# ***CHAPITRE 2 : PRETRAITEMENT ET EXPERIMENTATION***

## ***II.1 Introduction***

Le système éducatif de la République Démocratique du Congo (RDC) continue de se développer à tous les niveaux, bien que lentement et avec des hauts et des bas. Bien que cela, la qualité est faible à tous les niveaux du système éducatif, le contenu et les normes de l’enseignement secondaire et de l’enseignement supérieur n’ayant pas été, au demeurant, réformés depuis plus de vingt ans. L’enseignement primaire et secondaire est délivré dans des écoles d’Etat (écoles non conventionnées), dans des écoles gérées par des Eglises (écoles conventionnées) et dans des écoles privées. Les deux premières catégories d’écoles reçoivent le même financement public et sont considérées comme des écoles publiques.

Le système éducatif en RDC joue un rôle important dans la reconstruction des institutions politiques du pays, dans la stimulation de la croissance et dans la correction des inégalités. L’articulation d’une politique éducative cohérente avec les objectifs politiques et économiques généraux est essentielle pour l’élaboration d’une politique globale en faveur du développement du pays. L’objectif principal de ce chapitre est d’aider la RDC à identifier les priorités de la politique éducative et à développer une stratégie éducative qui soit à la fois techniquement et administrativement réalisable, mais aussi financièrement acceptable[2].

## ***II.2 Présentation de Données***

Une données est une collection d’éléments de données connexes et discrets qui peuvent être consultés individuellement ou en combinaison ou gérés comme une entité entière[18].

Un ensemble de données est organisé en un certain type de structure de données. Dans une base de données, par exemple, un ensemble de données peut contenir des données commerciales (noms, salaires, coordonnées, chiffres d’affaire, etc.). [18] La base de données elle-même peut être considérée comme un ensemble de données, tout comme les dataset qu’elle contient et qui sont liés à un type particulier d’informations, comme les données de vente d’un service particulier de l’entreprise.

La République Démocratique du Congo sous l’ordonnancement du Président de la République chef du gouvernement et du ministre de l’Enseignement primaire, secondaire et technique, l’Inspection générale publie les résultats d’examen d’Etat dans un site officiel du ministère de l’Enseignement primaire, secondaire et technique chaque année. Une représentation de palmarès nous est définie ci-dessous :



Figure 8: Présentation du palmarès de résultat d’examen d’Etat édition (2014) [19]

Notre Dataset regroupe les informations pertinentes ci-après :

* Nom de l’école ;
* Code ;
* Participants ;
* Participants (Filles) ;
* Réussites ;
* Réussites filles.

1. **Nom de l’école** : Définie comme « l’établissement où l’on enseigne les savoirs fondamentaux », comme expliqué ci-dessus nous parvenons à étudier la performance d’une ou plusieurs école(s) partant de différents paramètres que celle-ci possède.
2. **Code**:

* Code\_Province(District) : Une bonne division administrative est très capitale en l’enseignement, pour la ville province de Kinshasa, il existe 5 districts éducatifs (11 : KINSHASA-MONT AMBA, 12 : KINSHASA-FUNA, 13 : KINSHASA-LUKUNGA, 14 : KINSHASA-TSHANGU et 15 : KINSHASA-PLATEAU). Nous pouvons aussi prédire la performance d’un district quelconque enfin de sortir un milieu éducatif optimal.
* Code\_Centre : ce l’identification de l’environnement où passe chaque année les épreuves d’examen d’état.
* Code\_Option : la République Démocratique du Congo compte plusieurs cycles d’orientation et chacun comporte un identifiant dont nous citons :
* 101 : Latin-Philo
* 103 : Biologie-Chimie ;
* 201 : Pédagogie ;
* Etc.
* Num\_Ecole\_Centre (Numéro d’ordre de l’école par rapport au centre) : cette option est pertinente. Pourquoi ? parce que nous parvenons à connaitre le nombre que regorge un centre d’évaluation pour chaque District.
* Code\_gestion d’établissement : c’est un identifiant permettant de catégoriser des écoles par rapport à une politique religieuse.

**Exemple** : 0 : Autodidacte ; 1 : Non conventionné ; 2 : Catholique

1. **Participants (Filles et Garçon)** : comme tout autres sociétés, chaque année l’inspection générale de la ville province de Kinshasa recense pour chaque école le nombre de participants filles ou garçons et publie dans son site officiel pour une bonne traçabilité de résultat.
2. **Réussite (Filles et Garçons)** : comme toute sorte d’évaluation, une école peut faire cent pourcent de réussite ou non. D’où nous avons une colonne « Réussite » pour savoir le résultat que présente une école par rapport à la participation.

Exemple : Code complet d’un participant :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Code\_district** | **Code\_centre** | **Code-Option** | **Num\_Ecole\_Centre** | **Numéro du candidat** | **Code\_gestion d’établissement** |
| 11 | 078 | 103 | 04 | 026 | 2 |

Tableau 1: Code complet d’un participant d’examen d’Etat [19]

## ***II. 3 Traitement et Normalisation de données***

### ***II.3.1 Traitement de données***

Lorsque l’on parle de traitement des données, on entend le recours à différent procédées : colleté, enregistrement, organisation, conservation, utilisation, adaptation, modification, extraction, consultation, communication par transmission ou diffusion, etc.[20]. En informatique, le terme traitement de données ou traitement électronique des données renvoi à une série de processus qui permettent d’extraire de l’information ou de produire du savoir à partir de données brutes[21].

Contrairement à nous, nous avons eu des difficultés pour reconstituer notre Dataset. Pourquoi ? Parce que L’inspection générale de la ville province de Kinshasa ne nous a pas permis aux données sous format Excel. Nous étions obligés de télécharger les palmarès de la ville province de Kinshasa de 2019 à 2021 et les extraient en format Excel.

Bien qu’il existe des logiciels d’extraction en ligne mais celui-ci n’aboutissait pas aux résultats attendu. Nous avons fait l’extraction manuel pour obtenir le résultat ci-dessous :

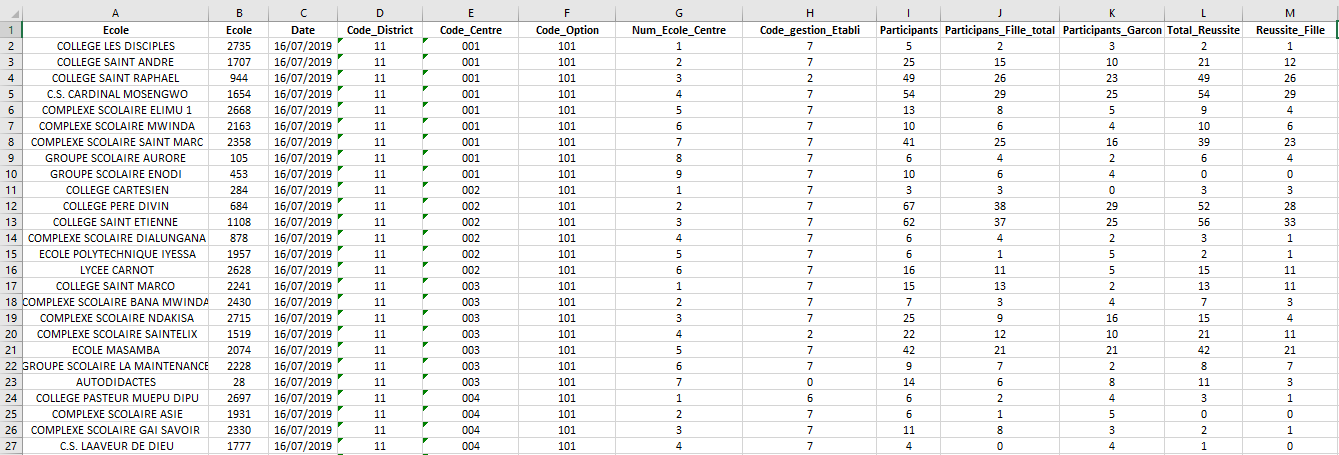


Figure 9: Représentation tabulaire du Dataset

### ***II.3.2 Normalisation de données***

La normalisation est une méthode de prétraitement des données qui permet de réduire la complexité des modèles. C’est également un préalable à l’application de certains algorithmes[22]. Une pratique courante en machine learning qui consiste à transformer colonnes numériques à une échelle commune, sans que les différences de plages de valeurs ne soient faussées et sans perte d’informations. Toutes les variables ont une influence similaire sur le modèle, améliorant la stabilité et les performances de l’algorithme d’apprentissage.

La normalisation permet également de créer de nouvelles valeurs qui conservent la même distribution générale et les mêmes ratios que les données sources tout en appliquant la même échelle aux valeurs des différentes colonnes numériques utilisées dans le modèle.

Une chose importante qui nous concerne par rapport aux algorithmes de prédiction que nous allons utiliser, il est question d’examiner bien ses exigences en termes de données avant d’appliquer la normalisation aux données d’apprentissage. Lorsque nous créons un modèle à des fins de prévision dans l'analyse de séries chronologiques, nous avons besoin d'une [série chronologique](https://analyticsindiamag.com/general-overview-of-time-series-data-analysis/) stationnaire pour une meilleure prédiction. Donc la première étape pour travailler sur la modélisation est de rendre une série temporelle stationnaire (pour plus de détail confer chapitre 3).

## ***II.4 conclusion***

Enfin, une grande partie de notre travail vient d’être effectué dès le traitement de données jusqu’à la stationnarité de la Dataframe. Deux tests de vérification de la stationnarité d'une série temporelle sont utilisés, à savoir le test ADF et le test KPSS. La suppression des tendances est effectuée en utilisant la différenciation. La série chronologique stationnaire de tendance est converties en série chronologique stationnaire strictes. Le modèle de prévision requis peut maintenant être appliqué sur des données de séries chronologiques stationnaires[23].

Nous allons mettre en présent dans le chapitre qui suit, des modèles de prévision sur le nombre de participants (Filles et Garçons), ainsi que la réussite fille et garçon chaque année de l’évaluation. En suit nous allons faire une étude comparative pour sélectionner le meilleur modèle de prévision pour le déploiement. Cela va permettre l’EPST de bien examiner chaque établissement sur son niveau d’enseignement.

# ***CHAPITRE 3 : RESULTAT ET DISCUSSION***

## ***III.1 Introduction***

L’objectif de tout le monde est de réussir. Tel a été notre détermination dès le début du ce travail. Compte tenu des problèmes soulèvent à l’introduction générale, nous portons maintenant une solution idéale pour permettre aux décideurs (c’est-à-dire l’inspection générale) a la bonne prise de(s) décision(s) partant des comportements de chaque école durant une période donnée. Nous allons parler tout au long du chapitre d’une manière générale des étapes parcourus, et particulièrement des algorithmes utilisent. Nous allons aussi faire une étude comparative pour chaque algorithme partant de leurs métriques d’évaluation. La figure 10 suivante explique le déroulement de notre travail :

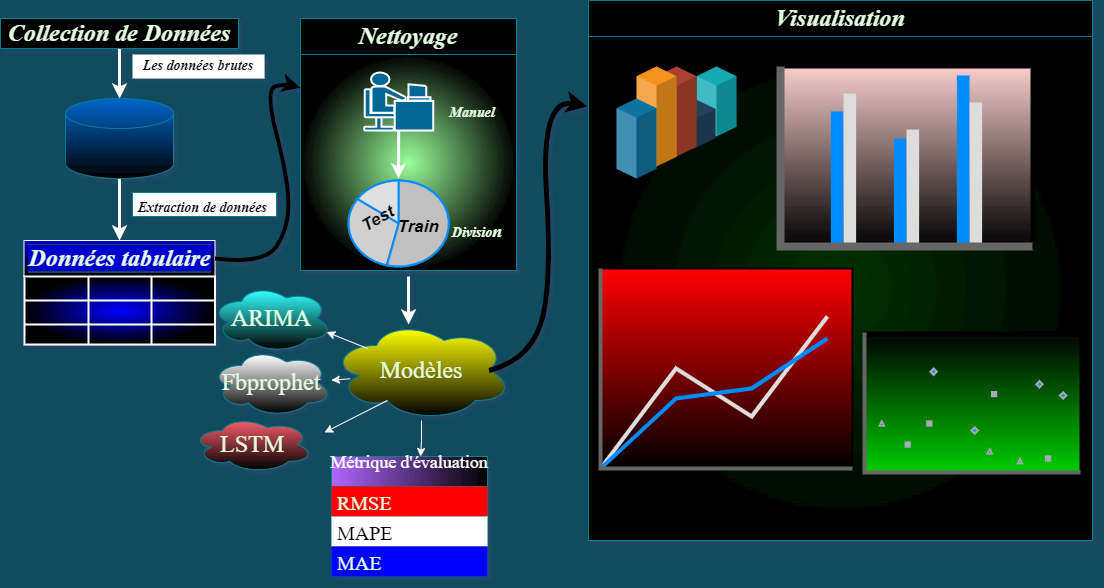
****

Figure 10. Schéma de recherche

## ***III.2 Collection de données***

La récolte de données est un point pertinent de notre travail de fin d’étude, notre Dataset regroupe des informations importantes (cfr chapitre 2). Ce n’était pas facile d’avoir ces informations. D’où nous demandions à l’Etat Congolais une mise à jour de sa politique aux informations publique dont il est responsable enfin de permettre aux chercheurs d’approfondir leur expérience sur un domaine.

## ***III.3 Les données tabulaires***

Il existe différentes formes de représentation de données, pour notre cas nous avons utilisés la représentation tabulaire. Partant de cette forme nous avons pu mettre en ordre nos informations par rapport à l’étude de séries chronologiques sa représentation est au chapitre 2 (II.3.1 Traitement de données) de ce travail.

## ***III.4 Nettoyage***

En ce qui concerne le nettoyage, nous avons utilisés comme outil Excel 2016. Lors du nettoyage, il y a eu des écoles mal orthographié, le changement du numéro d’ordre de l’école chaque année, et une même école peut s’écrire par différente forme. Malgré cela nous avons eu à identifier 2 822 écoles sur 5 districts de la ville province de Kinshasa durant une période de 3 ans (2019, 2020 et 2021). Nous profitons sur ce point une demande au ministère de l’Enseignement primaire, secondaire et technique de mettre à jour le palmarès d’examen d’Etat pour éliminer les différentes erreurs trouvées. La visualisation de notre dataset en format Excel est sur l’adresse suivante : « https://github.com/ArnauldMbuku/Time-series-Naulde/blob/main/Data\_Pal.xlsx»

## ***III.5 Modèles de prévisions***

La prévision de séries chronologiques consiste à utiliser des données passées pour prédire des événements futurs. La prévision de séries chronologiques est couramment utilisée dans les affaires et la finance pour prédire les ventes ou les cours des actions, et en science pour prédire les conditions météorologiques. Les modèles de prévision de séries chronologiques constituent une classe spéciale de modélisation prédictive utilisée pour prévoir les événements futurs. Les modèles de prévision de séries chronologiques sont différents des autres techniques de modélisation prédictive car ils tiennent compte de la nature temporelle des données.

### ***III.5.1 Modèle ARIMA***

Acronyme pour désigner l’autoregressive integrated moving average, l’ARIMA est un modèle statistique conçu pour l’analyse et la prédiction de données d'une série temporelle ou chronologique. Il détermine les valeurs intégrées à cette dernière en fonction des précédentes valeurs observées. Un système autorégressif suggère que les fonctions prédictives sont applicables dans un cas particulier, tout en minimisant le risque d’erreur inhérent à un tel exercice. Concrètement, cela permet d’anticiper l’évolution d’un phénomène. Cette notion mathématique demeure essentielle dans l'analyse de statistiques et le domaine de l’économétrie. Elle est aussi utilisée dans les processus d’action liés à l’intelligence artificielle (IA)[24].

Personne ne peut vraiment prédire le futur. Mais les méthodes statistiques modernes, les modèles économétriques, et les logiciels de Business Intelligence peuvent, dans une certaine mesure, aider les entreprises à prévoir et à estimer ce qui va se produire dans le futur[25].

* ***Définition***

ARIMA est un modèle statistique pour analyser ou prédire les données de séries temporelles ou time series. Le principe est également connu sous l'appellation moyenne mobile autorégressive intégrée[26].

* ***Historique***

L'Analyse de Séries Temporelles du Modèle ARIMA, emploie des retards et analyse les données historiques pour découvrir des modèles (par exemple des moyennes mobiles, un caractère saisonnier) et prévoir le futur. Le Modèle ARIMA a été développé pour la première fois vers la fin des années 60 mais il a été systématisé par Box and Jenkins en 1976. Le modèle ARIMA peut être plus complexe à employer que d'autres techniques de prévisions statistiques, bien qu'une fois mis en œuvre correctement le Modèle ARIMA puisse être tout à fait puissant et souple[25].

* ***Différentes classes de données de séries chronologiques***

Il existe deux classes différentes de données de séries chronologiques : les données stationnaires et non stationnaires. Les données de séries chronologiques stationnaires sont un ensemble de données de séries chronologiques qui ne présente pas de tendance ou d'effet saisonnier. L'erreur aléatoire est la seule source de variabilité dans l'ensemble de données. L'image ci-dessous représente des données stationnaires de séries chronologiques sans signe de tendance ou de saisonnalité :

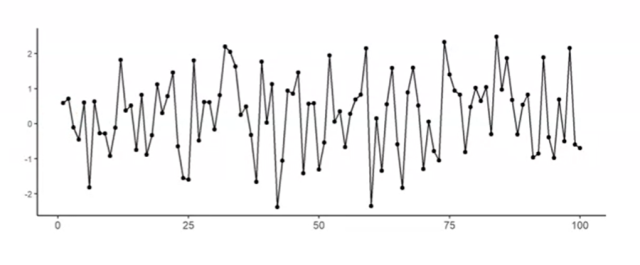
[](https://vitalflux.com/wp-content/uploads/2022/08/time-series-stationary-data.png)

Figure 11: Représentation graphique d’une série chronologique

Les données de séries chronologiques non stationnaires sont un ensemble de données de séries chronologiques qui présentent une tendance ou un effet saisonnier. L'erreur aléatoire n'est plus la seule source de variabilité dans l'ensemble de données[27]. Le test de [stationnarité](https://analyticsindiamag.com/how-to-check-time-series-stationarity-beginners-guide-in-python/) est une activité fréquemment utilisée dans la modélisation autorégressive. Nous pouvons effectuer divers tests comme le KPSS, Phillips-Perron et Augmented Dickey-Fuller[28]. Pour notre Dataset, nous utilisons deux test de stationnarité :

* **ADF (Augmented Dickey-Fuller)**

Le test ADF (Augmented Dickey-Fuller) est un test de signification statistique, ce qui signifie que le test donnera des résultats dans des [tests d'hypothèses](https://analyticsindiamag.com/importance-of-hypothesis-testing-in-data-science/) avec des hypothèses nulles et alternatives. En conséquence, nous aurons une valeur p à partir de laquelle nous devrons faire des inférences sur la série chronologique, qu'elle soit stationnaire ou non[28]. Avant d'entrer dans le test ADF, nous devons connaître le test de racine unitaire car le test ADF appartient au test de racine unitaire.

* Test de racine unitaire

Un test de racine unitaire prouve si une série chronologique n'est pas stationnaire et consiste en une racine unitaire dans l'analyse des séries chronologiques. La présence d'une racine unitaire dans les séries chronologiques définit l'hypothèse nulle, et l'hypothèse alternative définit les séries chronologiques comme stationnaires. Mathématiquement, le test de racine unitaire peut être représenté par :

Où

Est la composante déterministe.

Est la composante stochastique.

Est le processus d'erreur stationnaire.

Le concept de base du test de racine unitaire consiste à déterminer si le (composante stochastique) consiste en une racine unitaire ou non[28]. Un modèle AR(Auto-Régression) simple peut être représenté comme suit :

Où

Est la variable d'intérêt au temps t

Est un coefficient qui définit la racine unitaire

Est du bruit ou peut être considéré comme un terme d'erreur.

Si = ​​1, la racine unitaire est présente dans une série temporelle et la série temporelle est non stationnaire. Si un modèle de régression peut être représenté comme

Où

Est un opérateur de différence.

=

Donc ici, si = ​​1, ce qui signifie que nous obtiendrons la différence comme terme d'erreur et si le coefficient a des valeurs inférieures à un ou supérieures à un, nous verrons les changements en fonction de l'observation passée. Ainsi, si une série chronologique n'est pas stationnaire, elle aura tendance à renvoyer un terme d'erreur ou une tendance déterministe avec les valeurs temporelles. Si la série est stationnaire, elle aura tendance à ne renvoyer qu'un terme d'erreur ou une [tendance déterministe.](https://analyticsindiamag.com/guide-to-detrending-using-scipy-signal/)

Dans une série chronologique stationnaire, une grande valeur à tendance à être suivie d'une petite valeur, et une petite valeur a tendance à être suivie d'une grande valeur. Et dans une série chronologique non stationnaire, la grande et la petite valeur s'accumuleront avec des probabilités qui ne dépendent pas de la valeur actuelle de la série chronologique[28].

Nous appliquons ADF sur un modèle, et il peut être représenté mathématiquement comme :

Où

Est une constante

Est le coefficient au temps.

Est l'ordre de retard du processus autorégressif.

Ici, dans la représentation mathématique de l'ADF, nous avons ajouté les termes de différenciation qui font des changements entre l'ADF et le test de Dickey-Fuller. Le test de racine unitaire est alors effectué sous l'hypothèse nulle = 0 contre l'hypothèse alternative de < 0. Une fois une valeur pour la statistique de test.

Une série chronologique est dite « stationnaire » si elle n'a pas de tendance, présente une variance constante dans le temps et a une structure d'autocorrélation constante dans le temps. Une façon de tester si une série temporelle est stationnaire consiste à effectuer un test de Dickey-Fuller augmenté, qui utilise les hypothèses nulles et alternatives suivantes[29] :

**H 0 :** La série temporelle est non stationnaire. En d'autres termes, il a une structure dépendante du temps et n'a pas de variance constante dans le temps.

**H A :** La série temporelle est stationnaire.

Si la [valeur de p](https://www.statology.org/p-values-statistical-significance/) du test est inférieure à un certain niveau de signification (par exemple, α = 0,05), nous pouvons alors rejeter l'hypothèse nulle et conclure que la série chronologique est stationnaire. Exemple : Voici comment interpréter les valeurs les plus importantes dans la sortie :

* Statistique de test :  -0,97538
* Valeur P :  0,7621

Puisque la valeur de p n'est pas inférieure à 0,05, nous ne rejetons pas l'hypothèse nulle. Cela signifie que la série chronologique n'est pas stationnaire. En d'autres termes, il a une structure dépendante du temps et n'a pas de variance constante dans le temps[29].

* **KPSS (Kwiatkowski-Phillips-Schmidt-Shin)**

KPSS est un autre test pour vérifier la stationnarité d'une série temporelle. Les hypothèses nulle et alternative du test KPSS sont opposées à celles du test ADF[23]. Le test KPSS est également couramment utilisé pour analyser la stationnarité d'une série. Cependant, il présente quelques différences essentielles par rapport au test ADF en termes de fonction et d'utilisation pratique. Par conséquent, il n'est pas sûr de les utiliser de manière interchangeable[30].

***Comment mettre en œuvre le test KPSS ?***

En python, le statsmodel package fournit une implémentation pratique du test KPSS, une différence essentielle par rapport au test ADF est l'hypothèse nulle du test KPSS, c'est-à-dire que la série est stationnaire. Donc, pratiquement, l'interprétation de [la valeur p](https://www.staging33machinelearningplus.com/statistics/p-value/) est juste l'opposée l'une de l'autre.

Autrement dit, si la valeur p est < niveau significatif (disons 0,05), alors la série est non stationnaire. Alors que dans le test ADF, cela signifierait que la série testée est stationnaire. La sortie du test KPSS contient 4 éléments :

1. La statistique KPSS
2. Valeur p
3. Nombre de retards utilisés par le test
4. Valeurs critiques

La valeur p rapportée par le test est le score de probabilité sur la base duquel vous pouvez décider de rejeter ou non l'hypothèse nulle. Si la valeur de p est inférieure à un niveau alpha prédéfini (généralement 0,05), nous rejetons l'hypothèse nulle. Afin de rejeter l'hypothèse nulle, la statistique de test doit être supérieure aux valeurs critiques fournies. Si elle est en fait supérieure à la valeur critique cible, cela devrait automatiquement se refléter dans une valeur p faible. Autrement dit, si la valeur p est inférieure à 0,05, la statistique kpss sera supérieure à la valeur critique de 5 %.

Une différence majeure entre les tests KPSS et ADF est la capacité du test KPSS à vérifier la stationnarité en « présence d'une tendance déterministe ». Mais qu'est-ce qu'une « tendance déterministe » ? Le mot « déterministe » implique que la pente de la tendance dans la série ne change pas de façon permanente. C'est-à-dire que même si la série traverse un choc, elle a tendance à retrouver son chemin d'origine[30]. Il est toujours préférable d'appliquer les deux tests, afin de s'assurer que la série est bien stationnaire[31].

#### **Paramètres ARIMA**

Chaque composant d’ARIMA fonctionne comme un paramètre avec une notation standard. Pour les modèles ARIMA, une notation standard serait ARIMA avec p, d et q, où les valeurs entières remplacent les paramètres pour indiquer le type de modèle ARIMA utilisé. Les paramètres peuvent être définis comme suit :

* p : le nombre d’observations de décalage dans le modèle; également connu sous le nom d’ordre de décalage.
* d : le nombre de fois que les observations brutes sont différentes; également connu sous le nom de degré de différenciation.
* q : la taille de la fenêtre de la moyenne mobile; également connu sous le nom d’ordre de la moyenne mobile.

Dans un modèle [de régression linéaire](https://www.investopedia.com/terms/n/nonlinear-regression.asp), par exemple, le nombre et le type de termes sont inclus. Une valeur 0, qui peut être utilisée comme paramètre, signifierait que ce composant particulier ne doit pas être utilisé dans le modèle. De cette façon, le modèle ARIMA peut être construit pour remplir la fonction d’un modèle ARMA, ou même de simples modèles AR, I ou MA.

1. Différenciation[32].

L'estimation des modèles ARIMA suppose que l'on travaille sur une série stationnaire. Ceci signifie que la moyenne de la série est constante dans le temps, ainsi que la variance. La meilleure méthode pour éliminer toute tendance est de différencier, c'est-à-dire de remplacer la série originale par la série des différences adjacentes. Une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une version intégrée d'une série stationnaire (d'où le terme Integrated).

La correction d'une non-stationnarité en termes de variance peut être réalisée par des transformations de type logarithmique (si la variance croît avec le temps) ou à l'inverse exponentielle. Ces transformations doivent être réalisées avant la différenciation. Une différenciation d'ordre 1 suppose que la différence entre deux valeurs successives de y est constante.

µ est la constante du modèle, et représente la différence moyenne en y. Un tel modèle est un ARIMA (0, 1,0). Il peut être représenté comme un accroissement linéaire en fonction du temps. Si µ est égal à 0, la série est stationnaire.

Les modèles d'ordre 2 travaillent non plus sur les différences brutes, mais sur les différences de différence. La seconde différence de y au moment t est égale à () - (- ), c'est-à-dire à – 2+ . Un modèle ARIMA (0, 2,0) obéira à l’équation de prédiction suivante :

Ou encore :

1. Auto-régression

Les modèles auto-régressifs supposent que est une fonction linéaire des valeurs précédentes.

Littérairement, chaque observation est constituée d'une composante aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des observations précédentes. , et dans cette équation sont les coefficients d'auto-régression. A noter que cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA(1,1,0) on aura :

Ce qui peut également être écrit :

Notez qu'un processus auto-régressif ne sera stable que si les paramètres sont compris dans un certain intervalle ; par exemple, s'il n'y a qu'un paramètre auto-régressif, il doit se trouver dans l'intervalle -1<φ1. Dans les autres cas, les effets passés s'accumuleraient et les valeurs 3 successives des xt se déplaceraient infiniment vers l'avant, ce qui signifie que la série ne serait pas stationnaire.

1. Moyenne mobile

Les modèles à moyenne mobile suggèrent que la série présente des fluctuations autour d'une valeur moyenne. On considère alors que la meilleure estimation est représentée par la moyenne pondérée d'un certain nombre de valeurs antérieures (ce qui est le principe des procédures de moyennes mobiles utilisées pour le lissage des données).[32] Ceci revient en fait à considérer que l’estimation est égale à la moyenne vraie, auquel on ajoute une somme pondérée des erreurs ayant entaché les valeurs précédentes :

Littérairement, chaque observation est composée d'une composante d'erreur aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des erreurs aléatoires passées. θ1, θ2 et θ3 sont les coefficients de moyenne mobile du modèle. Comme précédemment cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA (0,1,1) on aura :

Ce qui peut également être écrit :

(14)

Un modèle de moyenne mobile correspond à des séries exhibant des fluctuations aléatoires autour d'une moyenne variant lentement. Plutôt que de prendre comme précédemment la valeur précédente comme prédicteur, on utilise une moyenne de quelques observations précédentes, de manière à éliminer le bruit, et estimer plus précisément la moyenne locale.

Cette logique correspond au lissage exponentiel simple, qui considère chaque observation comme la résultante d'une constante (b) et d'un terme d'erreur ε, soit :

La constante b est relativement stable sur chaque segment de la série, mais peut se modifier lentement au cours du temps. Si ce modèle est approprié, l'une des manières d'isoler la réelle valeur de b, et donc la partie systématique ou prévisible de la série, consiste à calculer une sorte de moyenne mobile, ou les observations courantes et immédiatement précédentes ("les plus récentes") ont une pondération plus forte que les observations plus anciennes. On peut également envisager des modèles mixtes : par exemple un modèle ARIMA (1,1,1) aura l'équation de prédiction suivante :

1. Détermination de l'ordre de différenciation

Une série stationnaire fluctue autour d'une valeur moyenne et sa fonction d'autocorrélation décline rapidement vers zéro. Si une série présente des autocorrélations positives pour un grand nombre de décalages (par exemple 10 ou plus), alors elle nécessite d'être différenciée. La différenciation tend à introduire des autocorrélations négatives. Si l'autocorrélation de décalage 1 est égale à 0 ou négative, la série n'a pas besoin d'être différenciée. Si l'autocorrélation de décalage 1 est inférieure à –0.5, la série est sur-différenciée.

L'ordre optimal de différenciation est souvent celui pour lequel l'écart-type est minimal. Un accroissement de l'écart-type doit donc être considéré comme un symptôme de sur-différenciation. Un troisième symptôme de sur-différenciation est un changement systématique de signe d'une observation à l'autre.

* ***Types***

[La prévision de séries chronologiques](https://en.wikipedia.org/wiki/Time_series) est un type de prévision qui prédit des événements futurs en fonction de points de données horodatés. Il existe de nombreux types de modèles de prévision de séries chronologiques, chacun ayant ses propres forces et faiblesses[27].

* *Modèle autorégressif (AR)*

Les modèles autorégressifs (AR) sont définis comme des modèles de régression dans lesquels la variable dépendante ou de réponse est une fonction linéaire des valeurs passées de la variable dépendante/réponse[27]. L'ordre d'un modèle autorégressif est noté 'p', qui représente le nombre de retards utilisés pour prédire la valeur actuelle. Par exemple, si p=0, cela signifie que nous prédisons le pas de temps actuel (t) en fonction du pas de temps précédent (t-0). Si p=n, alors nous prédisons le pas de temps (t) sur la base de n pas de temps passés.

* *Modèle de moyenne mobile (MA)*

Une moyenne mobile (MA) est un type de modèle utilisé pour la prévision de séries chronologiques. Les modèles de moyenne mobile sont principalement utilisés pour les données stationnaires, les données pour lesquelles nous ne voyons pas de tendances ou de saisonnalité significatives. Il existe deux types différents de modèle de moyenne mobile. Il s'agit d'un modèle de moyenne mobile simple et d'un modèle de moyenne mobile pondérée.

Une moyenne mobile simple(SMA) est un type de modèle de moyenne mobile qui utilise un nombre fixe de points de données pour le calcul de la moyenne. Ce type de MA est souvent utilisé pour les prévisions de séries chronologiques, car il est facile à calculer et peut être mis en œuvre dans une variété de langages de programmation[27].

Une moyenne mobile pondérée (WMA) est un type de modèle de moyenne mobile qui utilise un schéma de pondération pour donner plus d'importance aux points de données les plus récents. Ce type de MA peut être utilisé pour la prévision de séries chronologiques et peut aider à réduire l'impact des points de données plus anciens sur la moyenne.

* *Modèle de moyenne mobile autorégressive (ARMA)*

Le modèle de moyenne mobile autorégressive (ARMA) est une combinaison des modèles de moyenne mobile et autorégressif. Le modèle ARMA est défini comme un modèle de régression dans lequel la variable dépendante/réponse est une fonction linéaire des valeurs passées de la variable dépendante/réponse et du terme d'erreur. L'ordre d'un modèle ARMA est représenté par 'p' pour la partie autorégressive et 'q' pour la partie moyenne mobile. Par exemple, si p = 0 et q = 0, cela signifie que nous prédisons le pas de temps (t) en fonction du pas de temps (t) uniquement. Si p=n et q=m, alors nous prédisons le pas de temps (t) sur la base des n pas de temps passés de la variable dépendante/réponse et des m pas de temps passés du terme d'erreur[27].

* *Modèle de moyenne mobile intégrée autorégressive saisonnière (SARIMA)*

Il s’agit de l’extension du modèle ARIMA pour traiter les données saisonnières. Il divise les données en composantes saisonnières et non saisonnières et les modélise de la même manière. Il est caractérisé par 7 paramètres, pour les paramètres de partie non saisonnière identiques à ceux du modèle ARIMA et pour les paramètres de partie saisonnière où 'm' est le nombre de périodes saisonnières et P,D,Q sont similaires aux paramètres du modèle ARIMA[27]. Ces paramètres peuvent être calibrés à l’aide d’une recherche en grille ou d’un algorithme génétique.

SARIMA est un type de modèle de prévision de séries chronologiques qui prend en compte à la fois la saisonnalité et l'autocorrélation. Les modèles SARIMA sont basés sur une combinaison de processus de différenciation, d'autorégression et de moyenne mobile. Ces modèles peuvent être utilisés pour prévoir les tendances à court ou à long terme des données. Les modèles SARIMA sont généralement considérés comme plus précis que d'autres types de modèles de prévision de séries chronologiques, tels que les modèles ARIMA. Les modèles SARIMA sont également relativement faciles à interpréter et à utiliser.

Le modèle SARIMA peut être utilisé pour prévoir la demande d'un produit ou d'un service au cours d'une année. Le modèle prend en compte les données de ventes des années précédentes ainsi que la saisonnalité (par exemple, les ventes de vacances). SARIMA peut également être utilisé pour prévoir d'autres données de séries chronologiques, telles que les cours des actions ou les conditions météorologiques[27]. Le modèle SARIMA est une généralisation du modèle ARIMA (qui ne tient compte que de l'autocorrélation), et il peut être utilisé pour prévoir des données avec une saisonnalité.

* *Modèle vectoriel autorégressif (VAR)*

Le modèle vectoriel autorégressif (VAR) peut être défini comme un modèle multivarié de séries chronologiques qui capture l'interdépendance linéaire entre plusieurs variables de séries chronologiques. En plus des variables dépendantes, le modèle VAR inclut également une ou plusieurs valeurs décalées de chaque variable dépendante en tant que variables indépendantes. Le modèle VAR peut être estimé à l'aide des moindres carrés ordinaires (OLS). Une fois le modèle estimé, il peut être utilisé pour faire des prévisions.

Le modèle VAR est une généralisation du modèle autorégressif univariées (AR) et du modèle de régression linéaire multivarié. Il peut être utilisé pour étudier la dynamique d'une seule variable de série chronologique ou de plusieurs variables de série chronologique. L'un des avantages du modèle VAR est qu'il peut être facilement estimé à l'aide des MCO. Un autre avantage est qu'il peut être utilisé pour faire des prévisions. Un inconvénient du modèle VAR est qu'il peut être difficile à interpréter. Par exemple, il peut être difficile de déterminer quelles variables de séries chronologiques s'influencent mutuellement[27].

* *Moyenne mobile de régression automatique vectorielle (VARMA)*

C’est la combinaison de VAR et VMA et une version généralisée du modèle ARMA pour les séries chronologiques stationnaires multivariées. Il est caractérisé par des paramètres 'p' et 'q'. Tout comme ARMA est capable d’agir comme un modèle AR en définissant le paramètre 'q' comme 0 et comme un modèle MA en définissant le paramètre 'p' comme 0, VARMA est également capable d’agir comme un modèle VAR en définissant le paramètre 'q' comme 0 et comme un modèle VMA en définissant le paramètre 'p' comme 0[32].

Dans la littérature économétrique et dans l’analyse des séries temporelles (chronologiques), la classe des modèles autorégressifs moyennes mobiles vectoriels (VARMA pour Vector AutoRegressive Moving Average) sont utilisées non seulement pour étudier les propriétés de chacune de ces séries, mais aussi pour décrire de possibles relations croisées entre les différentes séries chronologiques. Ces modèles VARMA occupent une place centrale pour la modélisation des séries temporelles multivariées. Ils sont une extension naturelle des modèles ARMA qui constituent la classe la plus utilisée de modèles de séries temporelles univariées. Cette extension pose néanmoins des problèmes ardus, comme par exemple, l’identification et l’estimation des paramètres du modèle et suscite des axes de recherches spécifiques, comme la cointégration[33].

* *Modèle de correction d'erreur vectorielle (VECM)*

Le modèle vectoriel de correction d'erreurs (VECM) est un type de modèle de série chronologique souvent utilisé pour prévoir les données économiques. VECM est une extension du modèle traditionnel d'autorégression vectorielle (VAR). Le VECM a été introduit pour la première fois par Engle et Granger (1987). Les modèles VECM sont similaires aux modèles VAR, mais ils intègrent un terme de « correction » qui tient compte du fait que les erreurs d'une période sont susceptibles d'être corrigées au cours des périodes suivantes[27].

Cela rend les modèles VECM plus précis que les modèles VAR lors de la prévision de données avec une corrélation série significative. Les modèles VECM sont également relativement faciles à interpréter puisque les coefficients peuvent être directement interprétés comme l'impact d'une variable sur une autre. VECM est devenu un outil populaire en économétrie et en finance en raison de sa flexibilité et de sa facilité d'utilisation.

Le VECM présente plusieurs avantages par rapport au modèle VAR. Premièrement, il est plus facile d'interpréter les résultats d'un VECM que d'un VAR. Deuxièmement, le VECM peut être utilisé pour tester la co-intégration entre les variables de séries chronologiques, contrairement au VAR. Enfin, le VECM peut être appliqué à des données de séries chronologiques non stationnaires, alors que le VAR ne le peut pas. Les inconvénients du VECM incluent sa complexité et le fait qu'il nécessite une grande quantité de données pour produire des résultats fiables. De plus, le VECM est sensible aux valeurs aberrantes et aux changements dans les données de séries chronologiques[27].

* ***Fonctionnement la prévision ARIMA***

La prévision ARIMA est réalisée en connectant des données de séries chronologiques pour la variable d’intérêt. Un logiciel statistique identifiera le nombre approprié de décalages ou la quantité de différence à appliquer aux données et vérifiera la stationnarité. Il produira ensuite les résultats, qui sont souvent interprétés de la même manière que ceux d’un modèle de régression linéaire multiple.

### ***III.5.2 Modèle FacebookProphet***

* ***Définition***

Prophet est une librairie open source (R et Python) de prévision des données de séries temporelles basée sur un [modèle additif](https://en.wikipedia.org/wiki/Additive_model). Cette librairie permet d’analyser aisément des [séries temporelles](https://fr.wikipedia.org/wiki/S%C3%A9rie_temporelle) même pour des personnes n’ayant pas une grande expertise dans ce domaine[34].

* ***Historique***

En 2017, Facebook a ouvert le modèle prophet qui était capable de modéliser les séries chronologiques avec de fortes saisonnalités multiples au niveau du jour, de la semaine, de l’année, etc. et de la tendance. Il a des paramètres intuitifs qu’un scientifique de données si expert peut ajuster pour de meilleures prévisions. À la base, il s’agit d’un modèle régressif additif qui peut détecter les points de changement pour modéliser la série chronologique[35].

Facebook prophet a pour but de résoudre les problématiques suivantes : La difficulté de créer des modèles de prévision fiables : Car cette discipline requiert une expérience particulière et la rigidité et le manque de robustesse des techniques de prévisions automatiques.

* ***Fonctionnement Facebook Prophet***

Prévoir et analyser les données temporelles font partie des challenges en data science (par exemple prévoir la valeur du cours d’un indice boursier).  
C’est cette problématique que l’outil open source Facebook prophet tente de résoudre.

Comme présenté plus haut, Facebook prophet est basée sur un modèle additif. Ce qui veut dire qu’une série temporelle est modélisée comme la somme de plusieurs composantes. On parle de modèle de décomposition d’une série temporelle. Le Modèle additif se présente ainsi :

Où

 Correspond à la modélisation de la série temporelle,

La tendance,

 La composante saisonnière,

La composante aléatoire ou erreur.

Prophet rajoute une nouvelle composante qui correspond à l’impact des congés/vacances sur le modèle. Ainsi, le modèle de décomposition de prophet est le suivant :

correspond à l’effet vacances (h comme Holiday)[34].

### ***III.5.3 Modèle LSTM***

Les réseaux LSTM sont bien adaptés à la classification, au traitement et à la prévision sur la base de données chronologiques, car il peut exister des décalages d'une durée inconnue entre les événements importants d'une série chronologique. Les LSTMs ont été développés pour faire face aux problèmes de gradients qui peuvent être rencontrées lors de l’entrainement des RNNs traditionnels[36].

Le modèle de mémoire à long terme (LSTM) est un réseau neuronal récurrent qui est utilisé pour les séries chronologiques afin de tenir compte des dépendances à long terme. Il peut être formé avec une grande quantité de données pour capturer les tendances dans des séries chronologiques multivariées. Les neurones de RNN ont un état cellulaire / mémoire, et l’entrée est traitée selon cet état interne, qui est obtenu à l’aide de boucles dans le réseau neuronal. Il existe des modules récurrents de couches « tanh » dans les RNN qui leur permettent de conserver des informations. Cependant, pas avant longtemps, c’est pourquoi nous avons besoin de modèles LSTM[37].

Un module LSTM a un état de cellule et trois portes qui leur donnent le pouvoir d’apprendre, de désapprendre ou de conserver sélectivement les informations de chacune des unités. L’état de la cellule dans LSTM aide l’information à circuler à travers les unités sans être modifiée en ne permettant que quelques interactions linéaires. Chaque unité dispose d’une entrée, d’une sortie et d’une porte d’oubli qui peut ajouter ou supprimer les informations à l’état de la cellule. La porte d’oubli décide quelles informations de l’état cellulaire précédent doivent être oubliées pour lesquelles elle utilise une fonction sigmoïde. La porte d’entrée contrôle le flux d’informations vers l’état actuel de la cellule à l’aide d’une opération de multiplication ponctuelle de « sigmoïde » et « tanh » respectivement.

* ***Définition***

Long short terme memory (LSTM) est une architecture de réseau de neurones récurrents (RNN) utilisé dans le domaine de l’apprentissage en profondeur (deep learning). À la différence des réseaux neuronaux à propagation avant, le LSTM a des connections de feedback[36].

* ***Types***
* Classique***LSTM***

L'architecture LSTM classique est caractérisée par un état de cellule linéaire persistant entouré de couches non linéaires alimentant l'entrée et analysant la sortie de celle-ci. Concrètement, l'état de la cellule fonctionne de concert avec 4 couches de déclenchement, celles-ci sont souvent appelées les portes d'oubli, d'entrée (2x) et de sortie.

La porte d'oubli choisit les valeurs de l'ancien état de cellule à supprimer, en fonction des données d'entrée actuelles. Les deux portes d'entrée (souvent notées i et j) fonctionnent ensemble pour décider quoi ajouter à l'état de la cellule en fonction de l'entrée. i et j ont généralement des fonctions d'activation différentes, que nous nous attendons intuitivement à utiliser pour suggérer un vecteur de mise à l'échelle et des valeurs candidates à ajouter à l'état de la cellule.

Enfin, la porte de sortie détermine quelles parties de l'état de la cellule doivent être transmises à la sortie. Notez que dans le cas des LSTM classiques, la sortie h se compose d'activations de couches cachées (celles-ci peuvent être soumises à d'autres couches pour la classification, par exemple) et l'entrée se compose de la sortie d'état précédemment cachée et de toute nouvelle donnée x fournie au courant. Pas de temps[38].

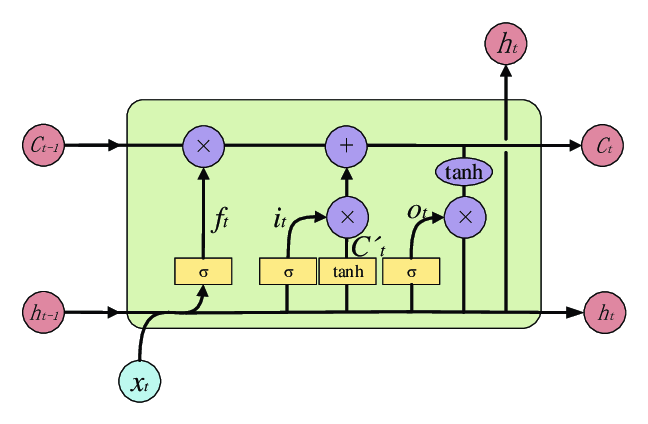


Figure 12: La structure de LSTM classique[39]

Une porte d'entrée détermine quelles informations doivent être mises à jour à l'état de la cellule. La fonction d'une porte d'entrée peut être formulée comme dans les équations 20 et 21, où C´t représente l'état de cellule provisoire calculé par la porte d'entrée, et il détermine quelles parties de C´t doivent être ajoutées à la finale CT d'état. L'équation 22 décrit comment la porte d'entrée se met à jour l'état cellulaire Ct[39].

Une porte de sortie détermine la valeur de sortie d'une cellule LSTM. La fonction d'une porte de sortie peut être formulée comme dans les équations 23 et 24, où ot représente un rapport provisoire qui détermine quelles parties de la cellule doivent être exportées et ht est la sortie finale de cette cellule[38].

#### **Connexions judas**

Le LSTM classique surmonte le problème de la disparition des gradients dans un réseau neuronal récurrent déroulé dans le temps en connectant des points absolus via un état cellulaire persistant (souvent appelé « carrousel d'erreurs constantes » dans les premiers articles décrivant les LSTM).

Cependant, les couches de déclenchement qui déterminent ce qu'il faut oublier, ce qu'il faut ajouter et même ce qu'il faut retirer de l'état de la cellule en sortie ne tiennent pas compte du contenu de la cellule elle-même.

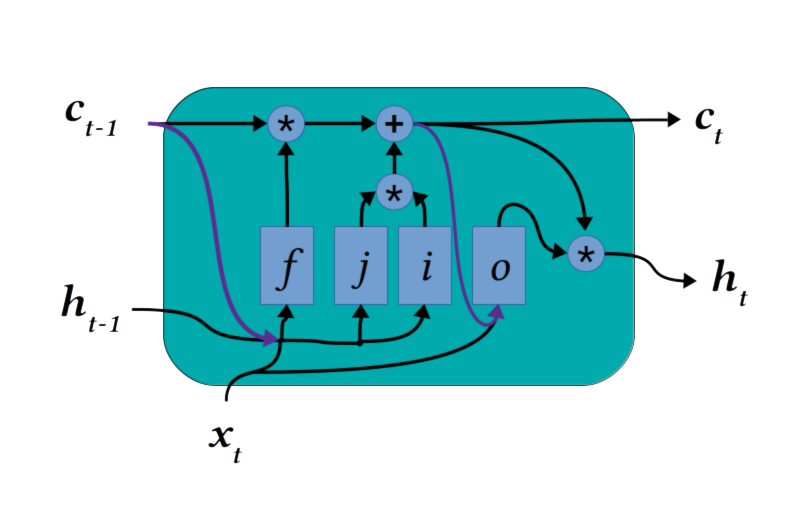


Figure 13: les couches de déclenchement d’une Connexions judas

Intuitivement, il est logique qu'un agent ou un modèle veuille connaître les souvenirs qu'il a déjà en place avant de les remplacer par de nouveaux. Entrez les connexions judas LSTM. Cette modification (représentée en violet foncé dans la figure ci-dessus) concatène simplement le contenu de l'état de la cellule aux entrées de la couche de déclenchement. En particulier, il a été démontré que cette configuration offrait une capacité améliorée de compter et de chronométrer les distances entre les événements rares lorsque cette variante a [été introduite à l'origine](https://www.researchgate.net/publication/3857862_Recurrent_nets_that_time_and_count) . Fournir certaines connexions d'état cellulaire aux couches dans un LSTM reste une pratique courante, bien que des variantes spécifiques diffèrent exactement sur les couches auxquelles l'accès est fourni.

#### **Unité récurrente fermée**

Schématiquement, une Gated Recurrent Unit (GRU) semble plus compliquée qu'un LSTM classique. En fait, c'est un peu plus simple et, en raison de sa relative simplicité, il s'entraîne un peu plus vite que le LSTM traditionnel. Les GRU combinent les fonctions de déclenchement de la porte d'entrée *j* et de la porte d'oubli *f* en une seule porte de mise à jour *z*.

Concrètement, cela signifie que les positions d'état de cellule réservées à l'oubli correspondront à des points d'entrée pour de nouvelles données. Une autre différence clé du GRU est que l'état de la cellule et la sortie cachée *h* ont été combinés en une seule couche d'état caché, tandis que l'unité contient également un état caché interne intermédiaire.

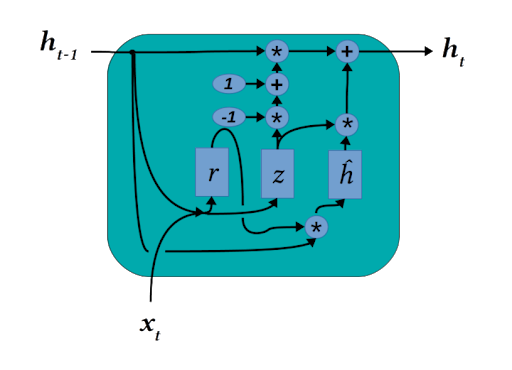


Figure 14: schéma d’une Unité récurrente fermée

Les unités récurrentes fermées (GRU) ont été utilisées comme base pour démontrer des concepts exotiques tels que [les GPU neuronaux](https://arxiv.org/abs/1511.08228) ainsi qu'un modèle plus simple pour l'apprentissage séquence à séquence en général, comme la traduction automatique. Les GRU sont une variante LSTM capable et ils sont assez populaires depuis leur création. Bien qu'ils puissent apprendre rapidement sur des tâches telles que la musique ou la génération de texte, ils ont été décrits comme finalement [moins puissants](https://arxiv.org/abs/1805.04908) que les LSTM classiques en raison de leurs limitations en matière de comptage[38].

* ***LSTM multiplicatif***

Le LSTM multiplicatif (mLSTM), une nouvelle architecture de réseau neuronal récurrent pour la modélisation de séquences qui combine la mémoire à long court terme (LSTM) et les architectures de réseau neuronal récurrent multiplicatif. mLSTM se caractérise par sa capacité à avoir différentes fonctions de transition récurrentes pour chaque entrée possible, ce qui, selon nous, le rend plus expressif pour l'estimation de densité autorégressive. Nous démontrons empiriquement que mLSTM surpasse le LSTM standard et ses variantes profondes pour une gamme de tâches de modélisation au niveau des caractères, et que cette amélioration augmente avec la complexité de la tâche. Ce modèle atteint une erreur de test de 1,19 bits/caractère sur les 4 derniers millions de caractères de l'ensemble de données du prix Hutter lorsqu'il est combiné à une évaluation dynamique[40].

Les architectures LSTM et mRNN étant complémentaires, nous proposons le LSTM multiplicatif (mLSTM), une architecture hybride qui combine la transition factorisée caché-caché des mRNN avec le cadre de déclenchement des LSTM. Les architectures mRNN et LSTM peuvent être combinées en ajoutant des connexions à partir de l'état intermédiaire mRNN du mRNN (qui est redéfini ci-dessous pour convenance) à chaque unité de déclenchement dans le LSTM, ce qui donne le système suivant :

Nous avons fixé la dimensionnalité de et égale pour toutes nos expériences. Nous avons également choisi de partager dans tous les types d'unités LSTM, résultant en un modèle avec 1,25 fois le nombre de poids récurrents comme LSTM pour le même nombre d'unités cachées.

Le but de cette architecture est de combiner les transitions flexibles dépendant de l'entrée des mRNN avec le long décalage temporel et le contrôle des informations des LSTM. Les unités fermées des LSTM pourraient faciliter pour contrôler (ou contourner) les transitions complexes qui résultent de la matrice factorisée des poids cachés. L'entrée sigmoïde supplémentaire et les portes oubliées des unités LSTM permettent encore plus de flexibilité fonctions de transition dépendantes de l'entrée que dans les mRNN ordinaires[41].

* ***Fonctionnement du LSTM***

Pour en revenir au fonctionnement d’un RNN, abordé dans l’introduction du sous-titre, chacune des couches cachées contient une seule fonction d’activation pour l’ensemble des neurones de celle-ci. Cette fonction va permettre d’affecter un seuil d’activation à chacun des neurones et celle-ci va affecter ensuite le poids synaptique des connexions à la couche suivante. La figure suivante montre un réseau de neurones récurrent avec une seule fonction d’activation « Tanh » par couche[42] :

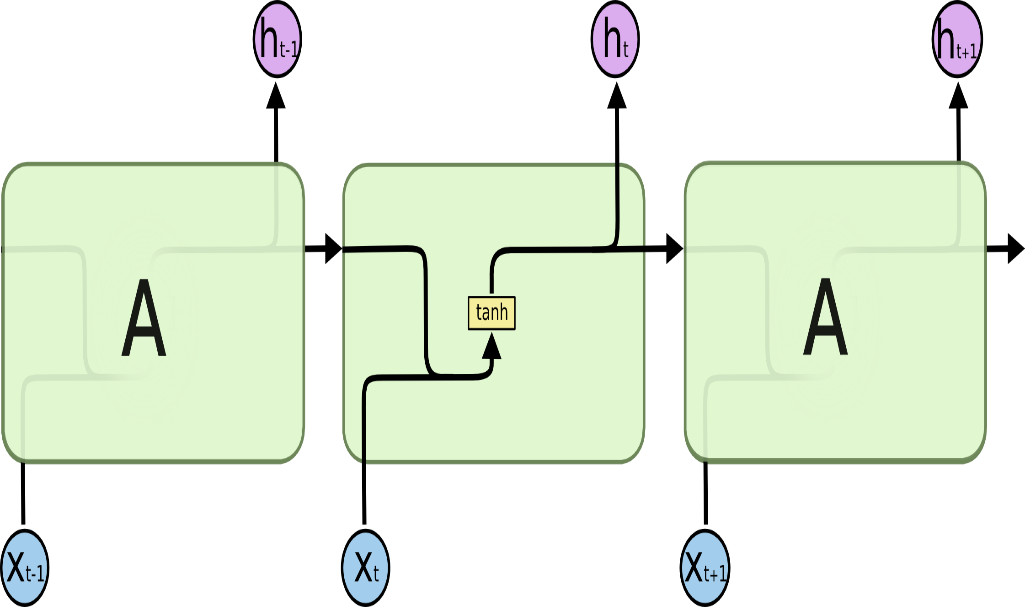


Figure 15: La tangente hyperbolique « Tanh »

La tangente hyperbolique « Tanh» permet d’affecter un seuil d’activation à chaque neurone d’une couche et de modifier les poids qui vont découler de la sortie de ce neurone.[42]

## ***III.6 Différentes métriques d'évaluation de prévision***

Mesurer les performances de tout modèle d'apprentissage automatique est très important, non seulement du point de vue technique, mais également du point de vue commercial. Pourquoi il est nécessaire de mesurer les performances d'un modèle de prévision de séries chronologiques[43] ?

Le fait que l'avenir est totalement inconnu et ne peut être prédit qu'à partir de ce qui s'est déjà produit est une distinction importante dans la prévision. La capacité d'un modèle de prévision de séries chronologiques à prédire l'avenir est défini par ses performances. Cela se fait souvent au détriment de la capacité d'expliquer pourquoi une prédiction particulière a été faite, des intervalles de confiance et même une meilleure compréhension des causes sous-jacentes du problème.

### ***III.6.1 Catégorie de métrique d’évaluation***

Sur la base de l'analyse de la structure de nombreux indicateurs de performance présentés dans la littérature, nous proposons un cadre de métriques : métriques primaires, métriques étendues, composites métriques et ensembles hybrides de métriques. Un aperçu et des exemples de chaque catégorie suivent.

**Les métriques primaires** sont sans doute la catégorie la plus nombreuse et incluent des métriques couramment utilisées tels que MAE, MSE, sMAPE, etc. La structure du primaire métrique implique trois par étapes : calculer la distance entre les points, effectuer la normalisation et agréger des résultats ponctuels sur un ensemble de données. Les métriques étendues sont généralement basées sur les métriques principales avec une normalisation supplémentaire. La délimitation avec les métriques principales est que la normalisation est effectuée après l'agrégation.

**Les ensembles hybrides** de métriques sont représentés par plusieurs métriques (deux ou plus) qui sont utilisées dans le même expérience avec plusieurs résultats de sortie. Ces ensembles ne sont pas destinés à être combinés dans une structure mathématique unique pour fournir une sortie à un seul nombre. Aucune liste de mesures ne peut constituent un ensemble hybride. Dans un ensemble hybride, les métriques proposées doivent être utilisées pour fournir mutuellement propriétés complémentaires permettant une meilleure compréhension des erreurs de performance, par exemple mesure partialité et précision. L'utilisation d'ensembles hybrides est conforme aux conseils de Fildes et Goodwin (2007) concernant l'utilisation plusieurs mesures de précision des prévisions.

Il existe différentes mesures d’évaluation utilisée pour mesurer les performances d'un modèle de prévision de séries chronologiques :

* R au carré
* Erreur absolue moyenne
* Erreur absolue moyenne en pourcentage
* Erreur quadratique moyenne
* Erreur quadratique moyenne
* Erreur quadratique moyenne normalisée
* Etc.

Voyons maintenant les métriques d'évaluation que nous avons utilisées pour mesurer les performances de nos modèles de prévision de séries chronologiques :

* **Erreur absolue moyenne (MAE)** : La MAE est définie comme la moyenne de la différence absolue entre les valeurs prévues et réelles.

Où

y i est la valeur attendue

x i est la valeur réelle (indiquée sous la formule).

La lettre n représente le nombre total de valeurs dans l'ensemble de test.

Le MAE nous montre le degré d'imprécision auquel nous devons nous attendre de la prévision en moyenne. MAE = 0 signifie que les valeurs anticipées sont correctes et que les statistiques d'erreur sont dans les unités d'origine des valeurs prévues. Plus la valeur MAE est faible, meilleur est le modèle ; une valeur de zéro indique que la prévision est sans erreur. En d'autres termes, le modèle avec le MAE le plus bas est jugé supérieur lors de la comparaison de nombreux modèles[43].

* **Erreur absolue moyenne en pourcentage (MAPE)** : MAPE est la proportion de la différence absolue moyenne entre les valeurs projetées et réelles divisée par la valeur réelle.

Où

Est la valeur anticipée

Est la vraie valeur

Le nombre n fait référence au nombre total de valeurs dans l'ensemble de test. Cela fonctionne mieux avec des données exemptes de zéros et de valeurs extrêmes en raison du dénominateur[43]. La valeur MAPE prend également une valeur extrême si cette valeur est extrêmement petite ou énorme. Le modèle est meilleur si le MAPE est faible.

* **Erreur quadratique moyenne (MSE)** : MSE est défini comme la moyenne des carrés d'erreur. Il est également connu comme la métrique qui évalue la qualité d'un modèle de prévision ou d'un prédicteur. MSE prend également en compte la variance (la différence entre les valeurs anticipées) et le biais (la distance entre la valeur prédite et sa valeur réelle)[43].

Où

Désigne la valeur prédite et

Désigne la valeur réelle.

Le nombre n fait référence au nombre total de valeurs dans l'ensemble de test.

La MSE est presque toujours positive et des valeurs inférieures sont préférables. Plus le MSE est proche de zéro, mieux qu'il surmonte les valeurs extrêmes MAE et MAPE et les problèmes zéro, il peut être nocif dans certains cas. Lorsqu'il s'agit d'un faible volume de données, cette statistique peut ignorer les problèmes ; pour résoudre ce problème, voir Erreur de pourcentage absolu pondéré et Erreur de pourcentage absolu moyen pondéré.

* **Erreur quadratique moyenne racine (RMSE)** : Cette mesure est définie comme la racine carrée de l'erreur quadratique moyenne et est une extension de MSE. Où y' désigne la valeur prédite et y désigne la valeur réelle. Le nombre n fait référence au nombre total de valeurs dans l'ensemble de test. Cette statistique, comme MSE, pénalise davantage les erreurs les plus importantes.

Cette statistique est également toujours positive, les valeurs inférieures indiquant des performances supérieures. Le nombre RMSE est dans la même unité que la valeur projetée, ce qui est un avantage de cette technique. Par rapport à MSE, cela facilite la compréhension. Plus l'écart entre RMSE et MAE est grand, plus la taille de l'erreur est erratique. Cette statistique peut masquer des problèmes avec un faible volume de données[43].

## ***III.7 Résultat de l’expérience***

### ***III.7.1 Visualisation de FacebookProphet***

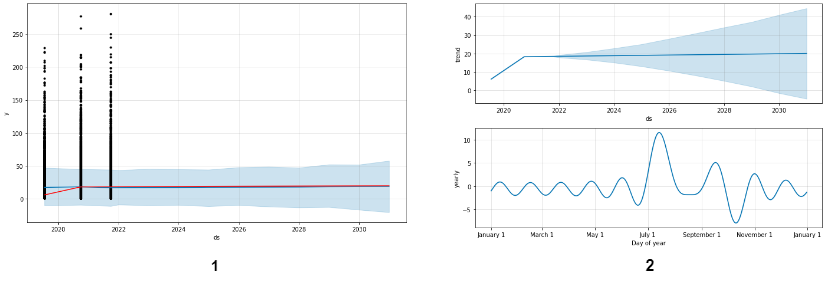


Figure 16. Représentation graphique des participants par rapport à l’année et mois

Sur la figure 16.1, nous trouvons les trois lignes noire verticales qui représente les 3 dernières années de notre dataset et la prévision périodique de 10 ans. La ligne bleue est les valeurs actuelles, la ligne rouge est des nouvelles valeurs de la prévision et le bleu-ciel est la tendance de nos deux valeurs (actuelles et futures) durant cette période. Contrairement à la figure 16.2, la première figure nous montre l’orientation de la tendance sur les nouvelles valeurs au futur. Et la seconde explique le comportement ou l’évolution d’une manière détaille (c’est-à-dire par Mois) pour les 10 ans. La courbe entre le mois de juillet et septembre augmente au futur mais arriver entre septembre et novembre on remarque une chute sur le taux de participants de plus de 5ans.

Alors cela interviens les décideurs pour connaitre le pourquoi et quoi faire pour ce genre de chute pendant une longue période.

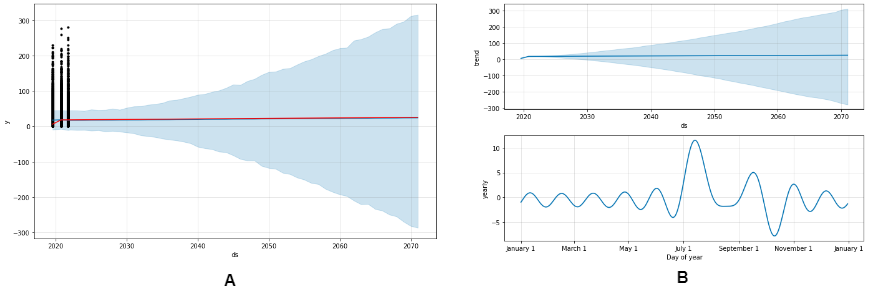


Figure 17. La seconde prévision pour FacebookProphet

La figure 17, nous mets à l’avance sur la prévision de participants à l’épreuve d’examen d’Etat sur la ville province de Kinshasa. Partant d’une période de 3ans, nous avons pu présenter la prévision pour 50ans (c’est-à-dire de 2030 à 2070). Nous constatons sur la figure 17.A que les nouvelles valeurs restent constantes alors que la tendance augmente au fur et à mesure pour chaque année. La figure 17.B reflète la figure 17.A, entre les années 2060 et 2070, les décideurs doivent faire très attention car la tendance peut diminuer ou augmenter sur cette période. Tout comme au second graphe de la figure 17.B nous constatons un taux de participants faible début.

### ***III.7.2. Visualisation de LSTM***

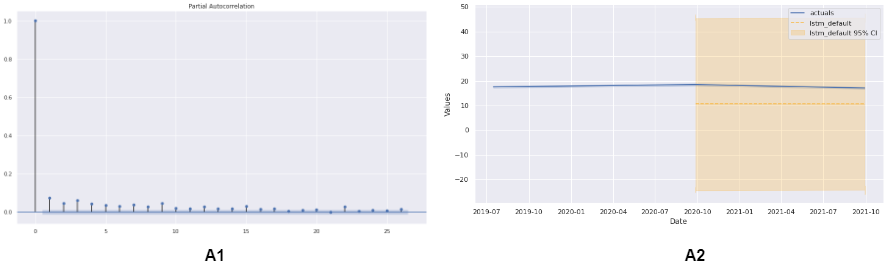


Figure 18. Evaluation de l'autocorrélation et Représentation des valeurs sur une période

L’autocorrélation permet de vérifier le caractère aléatoire d’un ensemble de données. Sur la figure 18. A1, nous avons effectués le calcul de l’autocorrélation pour les valeurs actuelles par rapport à un ensemble de valeurs passées. Les lignes bleue verticale montre l’autocorrélation qui existe entre nos différentes valeurs sur la colonne et cela promet un très bon modèle de prévision. Comme nous la voyons, la variation est de +1 c’est-à-dire la série temporelle 1 augmente en valeur et la série temporelle 2 augmente également. Et la figure 18. A2 détermine l’amélioration du modèle LSTM par l’un des paramètres qui lstm\_defaut, la ligne bleue horizontale est la représentation de valeur actuelle, les pointilles en orange définisse la zone que notre paramètre est effectué (2020 à 2021) et enfin l’estimation du modèle est de 95% pour les deux dernières années.

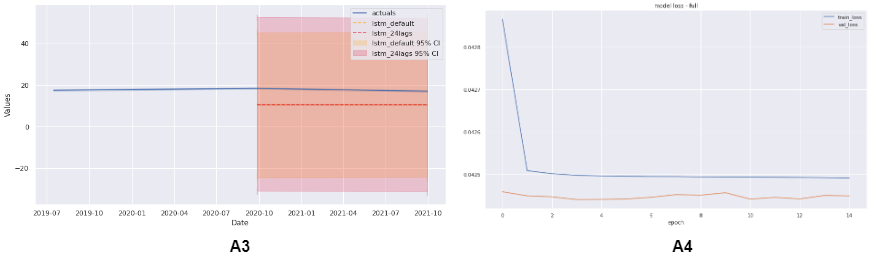


Figure 19. Schéma des paramètres de prévision et le taux d’erreur(perte)

Notre figure 19. A3 améliore la figure 18. A2 grâce décalage de 24 mois pour chaque couche du modèle LSTM. Comme constant : les pointilles de deux modèle reste stable c’est-à-dire sur une même longueur et la différence est sur le taux d’estimation que chaque modèle a par rapport aux données. Nous pouvons conclure que le décalage du 24 mois par couche est meilleur que le modèle lstm\_defaut. Et enfin Le modèle nous présente la courbe de perte(erreur) sur la figure 19. A4 de formation et de la validation du modèle LSTM à l’aide du métrique d’évaluation MAPE par rapport à l’époque.

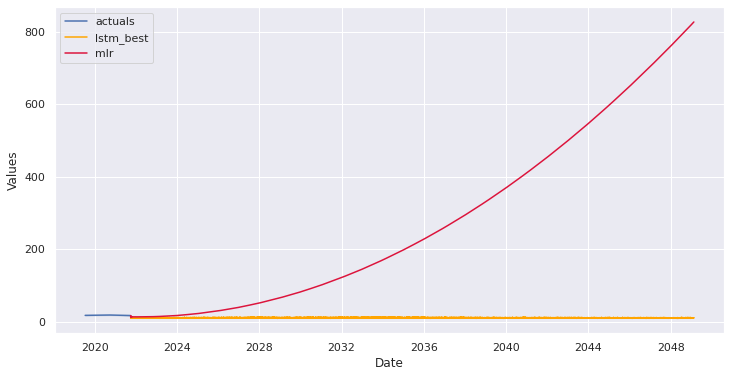


Figure 20. La prévision du modèle LSTM sur la colonne « Participants »

Sur cette figure, une comparaison majeure est faite entre le meilleur modèle LSTM et la régression linéaire multiple(MLR). Notre meilleur modèle LSTM garde une bonne tendance alors le modèle MLR sur les données différenciées, le package n’affiche par défaut que les données de niveau ce qui fait qu’il ait une suppression des intervalles de confiance.

### ***III.7.2. Visualisation de ARIMA***

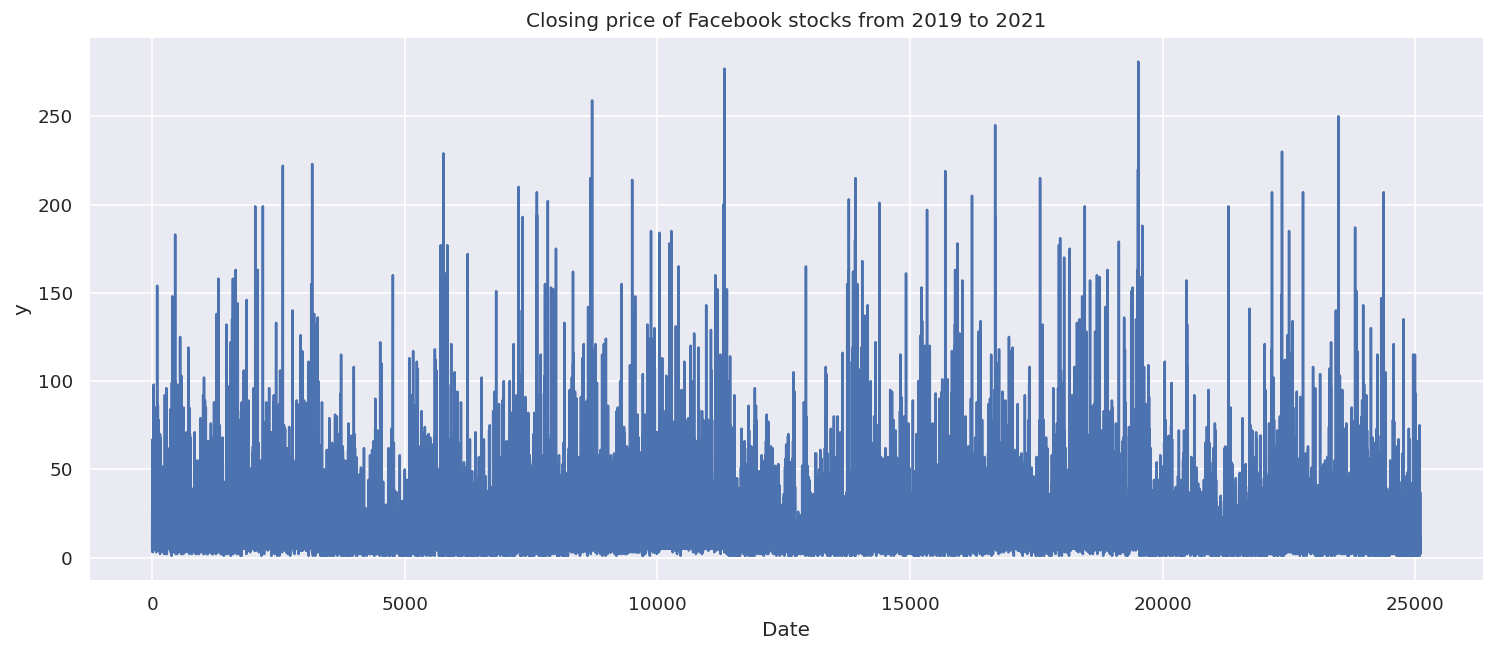


Figure 21. Visualisation de données par différenciation et autocorrélation

La représentation ci-dessus nous permets de voir le nombre de participants pour chaque école de 2019 à 2021. La majorité des écoles regorgent moins de 100 participants par différentes options alors qu’il y a des écoles qui ont plus de 250 participants pour une année.

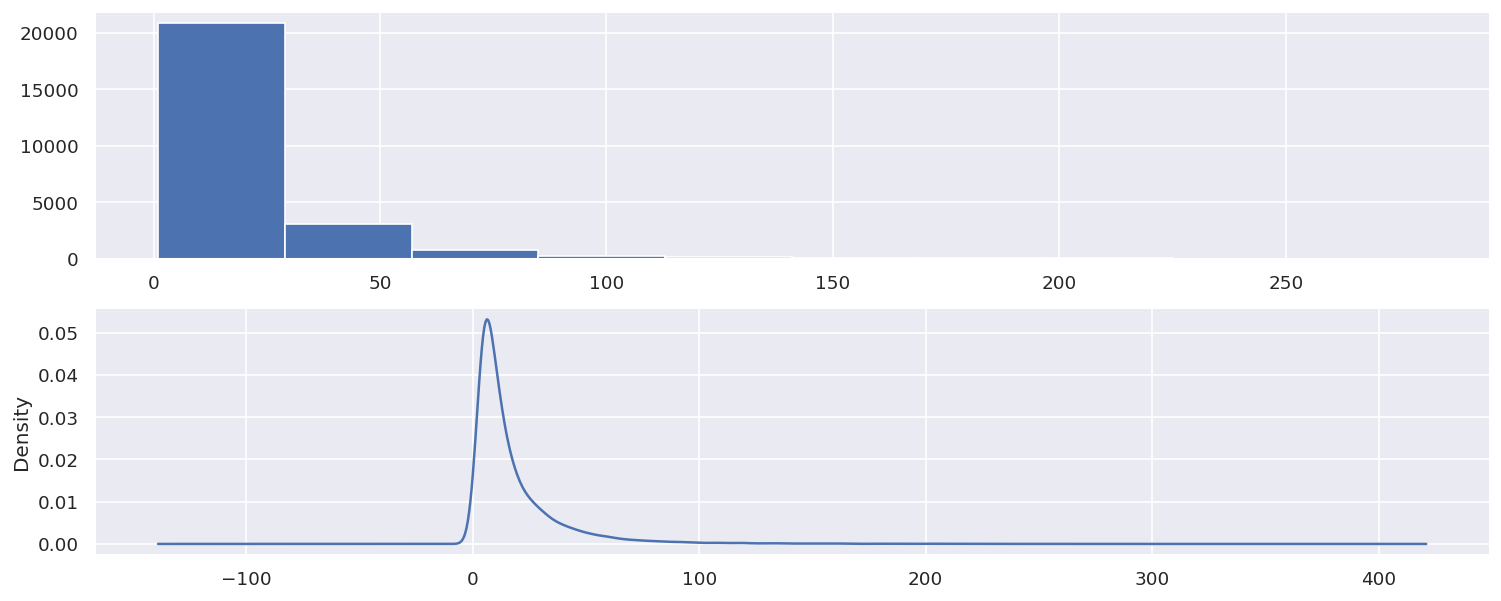


Figure 22. Evolution de la densité

La densité est un type d’outil de visualisation de données et elle permet de déterminer une grande quantité de données et sa fréquence. Sur la figure 22, Les données sur notre variable sont beaucoup plus regrouper entre 0 et 25 en plus nous trouvons sur la figure surnommer densité l’estimation est de -100 à 400.

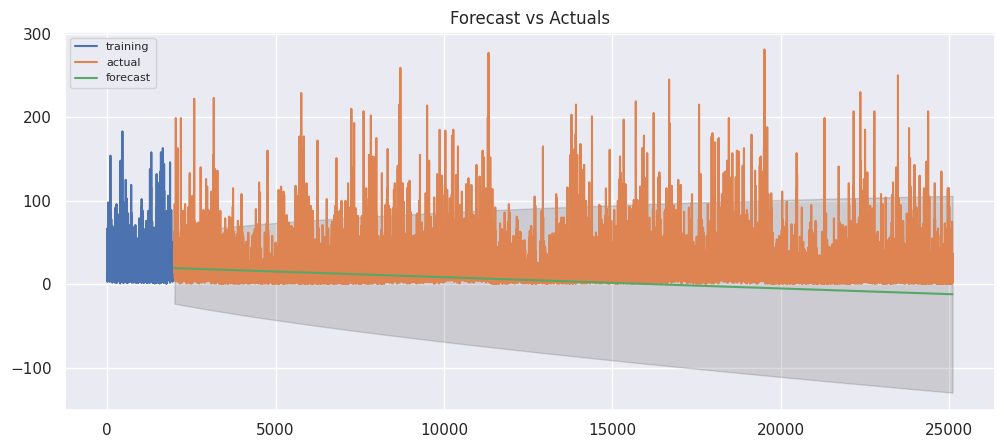


Figure 23. Disposition de données actuelles et futures

Cette figure est très importante pour notre étude parce qu’elle est composée des éléments de base pour une étude prédictive (les données de l’entrainement, données de test et la prévision). La couleur bleue représente les données de l’entrainement, en orange le test et en verte la courbe de la prévision pour les nouvelles valeurs de prédiction. Enfin l’espace en grise définisse la tendance que peut prendre notre modèle de prévision.

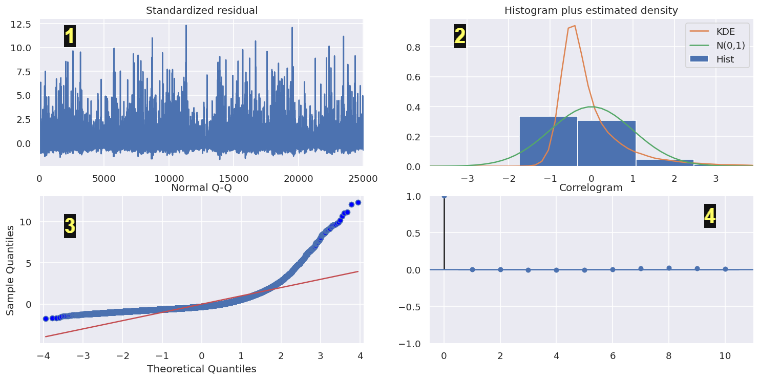


Figure 24. Différents présentation sur les données

Sur ce stade que nous étudions le comportement de nos données. Nous avons effectué 4 représentation et voici les explications pour chaque figure :

* Figure 24.1 : avec standardised residual, nous parvenons à représenter nous valeurs existantes pour chaque école enregistrée sur l’année
* Figure 24.2 : cette figure nous présentes un noyau qui permet à l’ordinateur d’exécuter différentes taches lorsqu’il est sollicité. Pour notre cas, le noyau est en couleur verte et son estimation varie selon le regroupement du point à l’aide d’un histogramme, KDE nous a permis de visualiser la densité de probabilité de notre variable (Participants) continue, tel est défini sur la figure, le KDE est trop élevé par rapport aux autres estimateurs.
* Figure 24.3 : notre variable a un problème de régression, c’est-à-dire les valeurs en bleue n’est pas bien aligner sur l’axe en rouge. Il est demande pour ce genre de cas l’application de cross validation (validation croise).
* Figure 24.4 : nous examinons la corrélation qui existe entre nos valeurs. Il y a une forte corrélation entre nos valeurs qui est +1 c’est-à-dire elle augmente entre eux.

## ***III.8 Etude comparative***

Notre étude comparative est basée sur les différents métrique utilisent à nos 3 algorithmes :

1. **Modèle FacebookProphet**

* **Participants**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **horizon** | **Rmse** | **Mae** | **Mape** | **Mdape** | **Coverage** |
| 365 days | 21.18258 | 14.191326 | 2.18386 | 0.93073 | 0.930678 |

Tableau 2. Métrique d'évaluation sur FacebookProphet/Participants

* **Réussite\_Garcon**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **horizon** | **Rmse** | **Mae** | **Mape** | **Mdape** | **Coverage** |
| 365 | 10.151261 | 7.925618 |  | 3.622943 | 0.964034 |

Tableau 3. Métrique d'évaluation sur FacebookProphet/Réussite\_Garçon

* **Réussite\_Fille**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **horizon** | **Rmse** | **Mae** | **Mape** | **Mdape** | **Coverage** |
| 365 | 21.18258 | 14.191326 | 2.18386 | 0.93073 | 0.929657 |

Tableau 4. Métrique d'évaluation sur FacebookProphet/Réussite\_Fille

1. **Modèle LSTM**

* **Participants**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ModelNickname** | **LevelTestSetMAPE** | **LevelTestSetRMSE** | **LevelTestSetMAE** | **best\_model** |
| lstm\_24lags | 1.091290 | 22.550051 | 11.855900 | False |
| lstm\_default | 1.095452 | 22.521301 | 11.853783 | True |
| lstm\_24lags\_5epochs | 1.104105 | 22.526322 | 11.870963 | False |
| lstm\_24lags\_earlystop\_3layers | 1.108253 | 22.524652 | 11.890491 | False |
| lstm\_best | 1.127219 | 22.511034 | 11.904010 | False |
| Mlr | 3.637123 | 41.000235 | 31.059848 | False |

Tableau 5. Métrique d'évaluation pour LSTM/Participants

* **Reussite\_Fille**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ModelNickname** | **LevelTestSetMAPE** | **LevelTestSetRMSE** | **LevelTestSetMAE** | **best\_model** |
| lstm\_24lags | NaN | 1.093998e+01 | -7.320212e-02 | False |
| lstm\_default | NaN | 1.090733e+01 | -6.680535e-02 | True |
| lstm\_24lags\_5epochs | NaN | 1.096532e+01 | -7.817914e-02 | False |
| lstm\_24lags\_earlystop\_3layers | NaN | 1.098356e+01 | -8.176945e-02 | False |
| lstm\_best | NaN | 1.096867e+01 | -7.883820e-02 | False |
| Mlr | NaN | 7.826340e+16 | -5.492450e+31 | False |

Tableau 6. Métrique d'évaluation pour LSTM/Réussite\_Fille

* **Reussite\_Garcon**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ModelNickname** | **LevelTestSetMAPE** | **LevelTestSetRMSE** | **LevelTestSetMAE** | **best\_model** |
| lstm\_24lags | NaN | 10.736774 | -0.057335 | False |
| lstm\_default | NaN | 10.733581 | -0.056707 F38 | True |
| lstm\_24lags\_5epochs | NaN | 10.764782 | -0.062859 | False |
| lstm\_24lags\_earlystop\_3layers | NaN | 10.789321 | -0.067710 | False |
| lstm\_best | NaN | 10.794446 | -0.068725 | False |
| Mlr | NaN | 10.592620 | -0.029134 | False |

Tableau 7. Métrique d'évaluation pour LSTM/Réussite\_Garçon

1. **Modèle ARIMA**

* **Participants**

|  |  |  |
| --- | --- | --- |
| **MAE** | **RMSE** | **MAPE** |
| 10.82005664537149 | 16.497062507467817 | 213.72221689125533 |

Tableau 8. Métrique d'évaluation ARIMA/Participants

* **Reussite\_Fille**

|  |  |  |
| --- | --- | --- |
| **MAE** | **RMSE** | **MAPE** |
| 0.47536960054332555 | 0.6532874751012795 | NaN |

Tableau 9. Métrique d'évaluation ARIMA/Réussite\_Fille

* **Reussite\_Garcon**

|  |  |  |
| --- | --- | --- |
| **MAE** | **RMSE** | **MAPE** |
| 7.05413706680666 | 9.824706256638361 | NaN |

Tableau 10.Métrique d'évaluation ARIMA/Réussite\_Garçon

## ***III.9 Discussion***

Sur 3 tableaux du modèle FacebookProphet, RMSE est faible par rapport aux deux autres modèles. La prévision comporte des erreurs compte tenu de l’analyse MAE et MAPE est aussi faible sur les trois tableaux du modèle enfin nous constatons la convergence proche de 100%.

Le modèle LSTM, nous avons mis au point 6 paramètres et partant de celle-ci est effectué nos différent métrique (MAPE, RMSE, MAE) sur la colonne MAPE du tableau 5 le résultat est favorable car chaque paramètre est proche de zéro tandis que aux tableaux 6 et 7 nous donne vide sur tous les paramètres. Sur la colonne RMSE nous trouvons les valeurs inférieures au tableau 6 et l’écart entre RMSE et MAE est plus grand c’est-à-dire la taille de l’erreur est erratique.

Par rapport aux résultats trouvent sur le modèle ARIMA, MAE nous donnes un meilleur résultat que le modèle FecebookProphet car il est proche de zéro ; RMSE est positive et il y a des valeurs inférieures qui indiquent une performance supérieure et l’écart entre RMSE et MAE n’est pas volumineux. Pendant que MAPE ne nous donne pas un bon résultat.

# 

# ***CONCLUSION GENERALE***

# ***BIBLIOGRAPHIE***

[1] “Rôle et impact de l’intelligence artificielle en éducation \_ Groupe IGS.”

[2] “Le système éducatif de la république démocratique du Congo : Priorités et alternatives Région Afrique Banque mondiale Région Afrique Département du développement humain Document de travail,” 2005.

[3] M. Learning, “COMMENT LES MACHINES APPRENNENT ?,” 2019.

[4] “Machine Learning \_ Définition, fonctionnement, utilisations.”

[5] “Apprentissage supervisé et non supervisé | Microsoft experiences.” https://experiences.microsoft.fr/articles/intelligence-artificielle/apprentissage-supervise-et-non-supervise-quelles-differences/ (accessed Sep. 14, 2022).

[6] “Machine learning \_ définition, modèle, algorithme et langage.”

[7] “Machine learning \_ définition et fonctionnement \_ Talend \_ Talend.”

[8] “Apprentissage Supervisé : Introduction - Machine Learnia.” https://machinelearnia.com/apprentissage-supervise-4-etapes/ (accessed Aug. 31, 2022).

[9] “Apprentissage non-supervisé : définition et algorithmes populaires.” https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501309-apprentissage-non-supervise/ (accessed Aug. 31, 2022).

[10] “Qu’est-ce que l’apprentissage non supervisé ? | Linedata.” https://fr.linedata.com/quest-ce-que-lapprentissage-non-supervise (accessed Aug. 31, 2022).

[11] C. Deluzarche, “Définition | Deep Learning - Apprentissage profond | Futura Tech.” 2019. [Online]. Available: https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-deep-learning-17262/

[12] “Deep learning \_ définition, fonctionnement et applications.”

[13] “Deep Learning vs Machine Learning - Copy.pdf.”

[14] “Time series \_ comprendre le rôle des séries temporelles.”

[15] “Séries chronologiques - Wikipédia.”

[16] R. M. Arma, “Table des matières,” 2011.

[17] “Chapitre 7 Prévision avec des modèles de séries temporelles à composantes non observées - ScienceDirect.”

[18] “Definition Data Set - Actualité Informatique.”

[19] C. C. Scolaire *et al.*, “Province : KINSHASA - EST - Code : 11 Province : KINSHASA - EST - Code : 11,” pp. 1–546, 2014.

[20] “Qu’est-ce que le traitement des données ? Définition.” https://donnees-rgpd.fr/traitement-donnees/ (accessed Aug. 27, 2022).

[21] “Traitement de données — Wikipédia.” https://fr.wikipedia.org/wiki/Traitement\_de\_données (accessed Aug. 27, 2022).

[22] “Normalisation - Data Analytics Post.” https://dataanalyticspost.com/Lexique/normalisation/ (accessed Aug. 27, 2022).

[23] “Stationnarité et suppression des tendances (ADF\_KPSS) — statsmodels.”

[24] “Autoregressive integrated moving average (ARIMA) \_ définition.”

[25] “Le Dico du Marketing.” [Online]. Available: http://www.ledicodumarketing.fr/

[26] Y. W. A. Nanlohy, “Autoregressive Integrated Moving Average (ARIMA),” *Tensor*, vol. 2, no. 1. pp. 25–32, 2021. [Online]. Available: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

[27] “Différents types de modèles de prévision de séries chronologiques - Analyse de données.”

[28] “Guide complet du test Dickey-Fuller dans l’analyse de séries chronologiques.”

[29] “Test de Dickey-Fuller augmenté en Python (avec exemple).” https://www.statology.org/dickey-fuller-test-python/ (accessed Aug. 27, 2022).

[30] “Test KPSS de stationnarité - Machine Learning Plus.” https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/ (accessed Aug. 27, 2022).

[31] “Stationnarité et décroissance (ADF/KPSS) – statsmodels | Jumbuck.” https://thejumbuckisalmostextinct.com/fr/stationnarité-et-décroissance-adfkpss/ (accessed Aug. 27, 2022).

[32] D. Delignières, “Séries temporelles-Modèles ARIMA,” 2000.

[33] Y. B. Mainassara, C. Francq, Y. B. Mainassara, C. Francq, Y. B. Mainassara, and C. Francq, “Estimation des modèles VARMA structurels avec innovations linéaires non corrélées mais non indépendantes To cite this version : HAL Id : inria-00494707 Estimation des modèles VARMA structurels avec innovations linéaires non corrélées mais non indépendante,” 2010.

[34] “Facebook prophet \_ La prévision à grande échelle \_ Le Data Scientist.”

[35] “Forecasting des séries temporelles avec la librairie fbprophet – Methodidacte.”

[36] “Long short term memory: définitions, exemple.” [Online]. Available: https://24pm.com/117-definitions/409-long-short-terme-memory

[37] “Series Forecasting with Recurrent Neural Networks (LSTM) - CYBERTEC.”

[38] “5 types de réseaux de neurones récurrents LSTM et que faire avec eux \_ Blog exact.”

[39] L. Hu, J. Zhang, Y. Xiang, and W. Wang, “Neural Networks-Based Aerodynamic Data Modeling: A Comprehensive Review,” *IEEE Access*, vol. 8, pp. 90805–90823, 2020, doi: 10.1109/ACCESS.2020.2993562.

[40] “LSTM multiplicatif pour la modélisation de séquences \_ OuvrirRévision.”

[41] B. Krause, I. Murray, S. Renals, and L. Lu, “Multiplicative LSTM for sequence modelling,” *5th Int. Conf. Learn. Represent. ICLR 2017 - Work. Track Proc.*, pp. 1–11, 2019.

[42] A. Kalousis, “A study of neural networks for natural language processing”.

[43] V. LENDAVE, “A Guide to Different Evaluation Metrics for Time Series Forecasting Models,” *Developers Corner*. p. 1, 2021. [Online]. Available: https://analyticsindiamag.com/a-guide-to-different-evaluation-metrics-for-time-series-forecasting-models/

# ***TABLES DE MATIERES***

*In memoriam* I

*Epigraphe* II

*Remerciements* III

*Répertoire des figures* IV

*Répertoire des tableaux* V

*Répertoire des Acronymes* VI

*INTRODUCTION GENERALE* 1

*CHAPITRE I. CONTEXTE THEORIQUE* 4

*I.1 Introduction* 4

*I.2. Définitions :* 4

*I.3. Différents types de machine learning* 5

*I.3.1 L’apprentissage supervisé* 6

*I.3.2 L’apprentissage non-supervisé* 9

*I.3.3 L’apprentissage par renforcement* 12

*I.4. Fonctionnement du machine learning* 12

*I.5. Les principaux algorithmes de machine Learning* 13

*I.6. Deep learning* 15

*I.6.1 Définitions* 16

*I.6.2 Fonctionnement* 16

*I.7. Séries chronologiques (ou Time series)* 17

*I.7.1 Définition* 18

*I.7.2 Modélisation Séries chronologiques* 18

*I.8 conclusion* 18

*CHAPITRE 2 : PRETRAITEMENT ET EXPERIMENTATION* 20

*II.1 Introduction* 20

*II.2 Présentation de Données* 20

*II. 3 Traitement et Normalisation de données* 23

*II.3.1 Traitement de données* 23

*II.3.2 Normalisation de données* 25

*II.4 conclusion* 25

*CHAPITRE 3 : RESULTAT ET DISCUSSION* 27

*III.1 Introduction* 27

*III.2 Collection de données* 27

*III.3 Les données tabulaires* 28

*III.4 Nettoyage* 28

*III.5 Modèles de prévisions* 28

*III.5.1 Modèle ARIMA* 29

*III.5.2 Modèle FacebookProphet* 43

*III.5.3 Modèle LSTM* 45

*III.6 Différentes métriques d'évaluation de prévision* 52

*III.6.1 Catégorie de métrique d’évaluation* 52

*III.7 Visualisation* 56

*III.7.1 FacebookProphet* 56

*III.7.2. LSTM* 57

*III.7.2. ARIMA* 59

*III.8 Etude comparative* 61

*III.9 Discussion* 64

*CONCLUSION GENERALE* 65

*BIBLIOGRAPHIE* 66

*TABLES DE MATIERES* 68