

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique



Département Génie Industriel

Option Management de l'innovation

Entreprise Tchin-lait

Mémoire de projet de fin d'études

Pour l'obtention du diplôme d'ingénieur d'état en Management de l'Innovation

Contribution à l'élaboration d'un système basé sur les
réseaux de neurones artificiels pour la prédiction du
comportement des points de vente.

Cas Tchin-lait

Khaled BOUAZIZ

Mehdi MEDIOUNI

Sous la direction de Mlle Nacéra ABOUN

Présenté et soutenu publiquement le 20/06/2018

Composition du Jury :

Président	M. Ali BOUKABOUS	MAA	ENP
Promoteur	Mlle Nacéra ABOUN	MAA	ENP
Co Promoteur	M. Yacine OUZEGDOUH	Cadre	Tchin Lait
Examineur	Mme Nadjwa BOUKADOUM	MAA	ENP

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique



Département Génie Industriel

Option Management de l'innovation

Entreprise Tchik-lait

Mémoire de projet de fin d'études

Pour l'obtention du diplôme d'ingénieur d'état en Management de l'Innovation

Contribution à l'élaboration d'un système basé sur les
réseaux de neurones artificiels pour la prédiction du
comportement des points de vente.

Cas Tchik-lait

Khaled BOUAZIZ

Mehdi MEDIOUNI

Sous la direction de Mlle Nacéra ABOUN

Présenté et soutenu publiquement le 20/06/2018

Composition du Jury :

Président	M. Ali BOUKABOUS	MAA	ENP
Promoteur	Mlle Nacéra ABOUN	MAA	ENP
Co Promoteur	M. Yacine OUZEGDOUH	Cadre	Tchik Lait
Examineur	Mme Nadjwa BOUKADOUM	MAA	ENP

Dédicace

*A mes chers parents pour tous leurs sacrifices, leur amour,
leur dévouement et leur soutien tout au long de mes études*

A ma grand-mère qui n'a jamais cessé de prier pour moi

A mes deux frères et à ma petite sœur Besma

A toute ma famille

*A «camarades » et à mes amis qui ont fait de ces trois
dernières années une période fabuleuse et inoubliable*

A la chambre B113 de la cité Bouraoui

A tout être qui m'est cher

Je dédie ce travail

Mehdi

Dédicace

Ce projet de fin d'étude est dédié à mes chers parents qui m'ont toujours poussé et motivé dans mes études. Sans eux, je n'aurais certainement jamais été ce que je suis. Ce mémoire représente donc l'aboutissement du soutien et des engagements qu'ils m'ont prodiguée tout au long de ma scolarité. Qu'ils en soient remerciés par cette modeste dédicace.

C'est un moment de plaisir de dédier cet œuvre à mes sœurs : Sabiha, Chahinez, Hadjer et Marwa ainsi qu'à mon neveu et ma nièce : Mehdi et Ilhem.

A N.Y. pour avoir toujours été présente dans les moments les plus difficiles.

Enfin, à mes amis Mehdi et Nadir, au CAP et à toute personne ayant laissé une trace dans mon parcours.

Khaled

Remerciements

Nous tenons à remercier en premier lieu notre promoteur et chef de département Génie Industriel Mlle Nacera ABOUN pour ses précieux efforts et pour son infaillible implication tout au long du projet, merci énormément.

Nous remercions aussi les responsables de Tchir-lait en l'occurrence M. T.E. BERKATI et M. Y. OUZEGDOUH qui ont ouvert les portes de l'entreprise et nous ont permis de réaliser ce travail.

Notre sincère gratitude est adressée à tout le corps des enseignants du département Génie Industriel pour leurs efforts et leurs sacrifices.

Nous remercions M. R. TAHMI, Directeur des études à l'Ecole Nationale Polytechnique pour son dévouement et son soutien affiché quotidiennement aux étudiants

Merci à toute personne ayant contribué de près ou de loin à la réalisation de ce projet.

Résumé et mots clés

ملخص:

الهدف من هذا العمل هو المساهمة في إرساء نظام للتنبؤ بتصرف زبائن تشين لاي. هذا النظام الذي يركز على الشبكات العصبونية الاصطناعية بشكل بالنسبة للمسير أداة لتسهيل القرار من أجل تعزيز قوة البيع.

بعد تشخيص الشركة ومحيطها، تمت دراسة عديد التقنيات التي تركز على التعلم الآلي وخاصة الشبكات العصبونية، ثم تم اختيار الشبكة الملائمة للحالة المدروسة بالإضافة إلى اختبارها وتقديم بعض الإرشادات فيما يخص إرساء النظام والتدابير اللازمة من أجل ضمان عمله الحسن.

كلمات مفتاحية: تنبؤ، نقاط البيع، تعلم آلي، الشبكات العصبونية.

Abstract:

The aim of this work is to contribute to the implementation of a dynamic system which predicts the behavior of Tchinq-lait company's customers. This system based on artificial neural networks constitutes for the manager a decision-making tool allowing to strengthen the sales force.

After a diagnosis of the company and its environment, the various techniques based on machine learning and mainly the artificial neural networks were reviewed. The most suitable network for this problem was chosen and tested before making recommendations regarding the implementation of the system and the necessary actions for its proper functioning.

Keywords: forecasting, outlets, machine learning, artificial neural networks, NARX neural network.

Résumé :

L'objectif de ce travail est de contribuer à la mise en place d'un système dynamique qui prédit le comportement de clients de l'entreprise Tchinq-lait. Ce système basé sur les réseaux de neurones artificiels constitue pour le manager un outil d'aide à la décision permettant de renforcer la force de vente

Après un diagnostic de l'entreprise et de son environnement, les différentes techniques basées sur l'apprentissage automatique et principalement les réseaux de neurones artificiels sont été passés en revue. Il s'agira par la suite de choisir le réseau le plus adéquat pour le cas étudié et de le tester avant de formuler des recommandations concernant l'implémentation du système et les actions nécessaires pour son bon fonctionnement.

Mots clés : prédiction, points de vente, apprentissage automatique, réseaux de neurones, réseau NARX.

Table des matières

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABREVIATIONS

INTRODUCTION GENERALE	11
CHAPITRE I : ETUDE DE L'EXISTANT	13
1. Présentation de l'entreprise	14
1.1.Introduction	14
1.2.Organigramme	15
1.3.Contrat de franchise	16
1.4.Evolution de l'entreprise	17
1.5.Les unités de production	17
1.6.Produits	18
2. Diagnostic	18
2.1.Diagnostic externe	18
2.2.Diagnostic interne	20
Conclusion	28
CHAPITRE II : ETAT DE L'ART : MACHINE LEARNING ET RESEAUX DE NEURONES	29
Introduction	30
1. L'apprentissage automatique	31
1.1.Introduction	31
1.2.Types d'apprentissage automatique	31
1.2.Intérêt de l'apprentissage automatique	33
1.4.Comparaison entre programmation classique et apprentissage automatique	34
2. Les réseaux de neurones artificiels	34
2.1.Introduction	34
2.2.Propriétés fondamentales des réseaux de neurones	35
2.3.Fonctionnement d'un réseau de neurones	36
2.4.Le neurone formel	37
2.5.Fonctions de transfert	37

2.6.Types de réseaux de neurones	39
2.7.L'apprentissage	43
Conclusion	48
CHAPITRE III : CHOIX DE LA METHODE ET SIMULATION	47
Introduction	50
1. Réseau NARX	41
1.1.Définition	41
1.2.Architecture	41
2. Sélection des variables exogènes	51
2.1.Introduction	51
2.2.ANOVA	52
3. Simulation	61
3.1.Les étapes de construction du modèle	62
3.2.Exécution	65
3.3.Discussion	66
3.4.Analyse de la performance et validation	67
3.5.Amélioration de la performance	70
3.6.Importance des variables exogènes	71
4. Recommandations	72
4.1.Analyse de l'environnement	73
4.2.Organisation	73
4.3.Technique	76
4.4.Système informatique	76
Conclusion	77
CONCLUSION GENERALE	77
BIBLIOGRAPHIE	78
ANNEXES	80

Liste des tableaux

Tableau 1 : Type de réseau et leur application (Heaton, 2015).....	50
Tableau 2 : Travaux réalisés par des réseaux récurrents	51
Tableau 3 : Total des ventes Viva au mois de Ramadhan et hors Ramadhan.....	57
Tableau 4:Rapport détaillé	57
Tableau 5:Résultat du test ANOVA sur le produit Viva sous l'influence du facteur «Ramadhan ».	58
Tableau 6:Résultats de l'analyse de la variance sur les produits Candia à $\alpha = 0,05$	59
Tableau 7:Liste des produits impactés par le facteur « Ramadhan »	60
Tableau 8 : Liste des produits impactés par le facteur «été».....	60
Tableau 9: Résultat simulation NARX.....	66
Tableau 10: Comparaison de la performance du réseau.....	70
Tableau 11:Comparaison de la prédiction à l'échelle macro et micro	74

Liste des figures

Figure 1: Organigramme de l'entreprise (Document interne, 2018).....	15
Figure 2 : Organigramme du département Marketing et Force de vente	16
Figure 3 : Réseau de distribution chez Candia	21
Figure 4 : Schéma du processus de vente	24
Figure 5. Evolution des approvisionnements de 4 produits différents chez un même point de vente	25
Figure 6.Evolution des approvisionnements du produit DE 11 chez 4 points de vente différents	26
Figure 7. Fréquence des achats.....	27
Figure 8. Les ruptures dans certains produits	28
Figure 9. Classification des algorithmes du machine learning.....	33
Figure 10. Différence entre la programmation classique et l'apprentissage automatique (Shukla, 2018).	34
Figure 11. Architecture d'un RNA (Zhang, 2004).	36
Figure 12. Schéma d'un neurone artificiel (Aouedad, 2018).	37
Figure 13. Tracé de la fonction logistique	38
Figure 14. Tracé de la fonction Tanh	38
Figure 15. Tracé de la fonction ReLU	39
Figure 16. Réseau de neurones non bouclé(Dreyfus, 2004).....	40
Figure 17. Réseau de neurones récurrent (Touzet, 1992).....	41
Figure 18. Architecture NARX, Séries-Parallèles (Topus, 2010).	42
Figure 19. Architecture NARX en mode parallèle (Topus,2010).	42
Figure 20. Cycle de l'apprentissage	47
Figure 21. Distribution de l'échantillon VIVA	55
Figure 22. Diagramme Q-Q de l'échantillon VIVA.....	56
Figure 23 : Etapes de la simulation	61
Figure 24. Les 7 premières lignes de la matrice qui alimente le réseau pour la simulation.....	62
Figure 25. Illustration de l'interface pour les RNA sur Matlab	63
Figure 26 :Répartition des données	64
Figure 27. Structure du modèle retenu	67
Figure 28. Corrélation entre les valeurs de prédiction et les valeurs réelles (NARX)	67
Figure 29. Diagramme de l'erreur	68
Figure 30. Histogramme illustrant le cumul de l'erreur	68
Figure 31. Diagramme des valeurs réelles	69
Figure 32. Diagramme des valeurs prédites.	69
Figure 33. Diagramme des valeurs réelles et prédites	69
Figure 34: Comparaison de série périodique (gauche) et série quelconque (droite).....	71
Figure 35 : Comparaison des ventes Viva en 2016 et 2017	71
Figure 36 : Corrélation entre les valeurs de prédiction et les valeurs réelles (NAR).....	72
Figure 37. Processus de prévision de vente (Mentzer, 2004).....	73
Figure 38 : Processus de veille	76

Liste des abréviations

AR : Auto Regressive
ARIMA : Auto Regressive Integrated Moving Average
ARMA : Auto Regressive Mobile Average
DE : Demi écrémé
FDV : Force de Vente
FFNN : Feedforward neural network
IA : Intelligence artificielle
MA : Mobile Average
NAR : Non linear autoregressive
NARX : Non linear autoregressive with exogene
NN : Neural network
NNC : Nombre de neurones cachés
PDA : Personal Digital Assistant
RNA : Réseau de neurones artificiels
RNN : Recurent neural network
UHT : Upérisation à haute température

Introduction générale

Dans un environnement dynamique, volatile et très concurrentiel, la possession de l'information au bon moment est devenue un élément crucial pour le bon pilotage des entreprises. De nos jours, les managers cherchent de plus en plus à acquérir des outils d'aide à la décision qui leur permettent d'anticiper le futur et d'adapter leurs plans d'action. Le propos de E. Von Kuenheim, Président du Conseil d'Administration BMW qui a dit « aujourd'hui, ce n'est plus les gros qui mangent les petits mais les plus rapides qui mangent les plus lents » en est une preuve.

Dans un même contexte, le développement des outils technologiques et principalement les algorithmes de l'intelligence artificielle combinés aux *Big data* ont offert de nouvelles perspectives pour les entreprises et risquent de redéfinir l'équilibre des forces.

Il est donc évident pour une entreprise comme Tchiv-lait, qui évolue dans un domaine très particulier, caractérisé par une forte concurrence, de chercher à exploiter de tels outils. Ces derniers permettraient d'aider les dirigeants à prendre les meilleures décisions stratégiques et opérationnelles dans le but d'avoir un avantage concurrentiel et consolider la position de l'entreprise sur son marché.

Aujourd'hui, l'entreprise souhaiterait avoir une meilleure visibilité sur son marché, ses points de vente ainsi que sur les facteurs qui influent sur la demande des consommateurs. En effet, la demande des produits proposés par Tchiv-lait est caractérisée par de fortes fluctuations, ce qui peut causer des ruptures au niveau des points de vente mais aussi des problèmes dans la distribution.

Pour faire face à de telles situations, les dirigeants de l'entreprise envisagent de mettre en place un outil intelligent qui prédirait le comportement des points de vente; un système qui se veut dynamique et réactif aux changements qui surviennent dans l'environnement.

Outre les avantages que présente un tel outil, le système d'information de Tchiv-lait est déjà préétabli et est en train d'accumuler de bonnes quantités de données; ceci est un facteur supplémentaire ayant motivé les responsables à soumettre cette problématique.

C'est pour répondre à cette attente que le présent projet a été initié et dont l'objectif est de poser les bases d'un système intelligent qui permettrait de prédire le comportement des points de vente et qui peut s'adapter avec les données qui lui sont fournies.

Aussi et pour atteindre cet objectif, le travail a été structuré en 3 chapitres comme suit :

Le premier chapitre est dédié à l'étude de l'existant. Ainsi, et après une présentation de l'entreprise, un diagnostic général comportant une analyse interne et externe est réalisé. Ce diagnostic permet de mieux comprendre le marché de l'entreprise. L'accent est mis sur le système de distribution actuel, son organisation et son fonctionnement.

Dans le deuxième chapitre, un état de l'art sur les algorithmes d'apprentissage automatique ou *Machine Learning* et plus particulièrement sur les réseaux de neurones artificiels (RNA) est présenté. Le problème étudié relevant du domaine des prédictions, c'est vers les RNA, qui font partie des algorithmes d'apprentissage supervisé, que s'est porté notre choix. En effet, grâce à leurs qualités de prédiction ils sont les mieux indiqués pour le cas d'étude.

Le troisième chapitre est consacré l'application. Dans un premier temps il s'agira d'identifier le RNA optimal. Suite à cela, et afin de tester la performance de l'outil sélectionné, une simulation qui prédit la demande globale des points de vente d'un produit test est réalisée. Ce chapitre se termine par un ensemble de recommandations en vue d'aider l'entreprise dans l'implémentation du système mais aussi pour son bon fonctionnement.

Enfin, une conclusion générale viendra clôturer ce travail.

Chapitre I : ETUDE DE L'EXISTANT

Introduction

La première étape de ce projet, qui est en effet indispensable dans toute approche scientifique, consiste à analyser l'état actuel de l'entreprise. Cette étape permettra de mieux comprendre le fonctionnement de l'entreprise et de son environnement, de cerner les problèmes, d'identifier les voies de solutions et de décider des actions à entreprendre pour améliorer le système étudié.

Mais avant de passer à cette analyse, nous allons en premier lieu faire une présentation de l'entreprise.

1. Présentation de l'entreprise

1.1. Introduction

L'entreprise Tchîn-lait est une société à responsabilité limitée de droit algérien spécialisée dans la production des produits laitiers sous le label de Candia. Fondée en 1999 par Mr. Fawzi BERKATI et implantée sur l'ancien site de la limonaderie Tchîn-tchîn situé à l'entrée de la ville de Bejaia.

Historiquement, l'entreprise Tchîn-tchîn, fonctionnelle depuis 1952, était spécialisée dans la production des boissons gazeuses. Dans les années 1990, l'entrée du pays dans l'économie du marché et l'arrivée des entreprises multinationales disposant de grandes ressources financières et d'une forte image publicitaire a beaucoup compliqué la mission des entreprises locales dont Tchîn-tchîn à cause d'une concurrence insoutenable. Pour réagir à cette situation, la direction de l'entreprise a décidé de réviser sa stratégie et d'envisager un futur dans d'autres marchés. Un accord a été signé avec le géant français Candia pour produire et commercialiser les produits de cette marque, transformant ainsi la limonaderie en une nouvelle laiterie.

Avant d'obtenir l'accord, plusieurs industriels algériens se sont adressés à Candia afin de pouvoir produire sous le nom de cette marque fortement appréciée par le consommateur algérien. Finalement c'est le dossier présenté par Tchîn-tchîn qui a été retenu et un contrat de franchise a été signé afin que l'entreprise –dénommée désormais Tchîn-lait- soit le représentant exclusif de Candia en Algérie.

Il faut rappeler que les produits de Candia étaient déjà disponibles sur le marché algérien depuis plusieurs années. En 1998, l'entreprise a suspendu ses exportations à cause d'une hausse des tarifs douaniers en Algérie. Le lancement en 2001 de la production chez Tchîn-lait a permis aux produits de cette marque de réapparaître sur les étalages des épiceries algériennes.

En 2015, l'entreprise a acquis un nouveau site de production à Baraki en rachetant l'usine de Générale Laitière Jugurtha d'une superficie de 30 000 m².

En 2017, une restructuration de l'entreprise a donné naissance à la SPA Tchîn-lait, avec un capital social de 1.000.000.000,00 DA.

Aujourd'hui, Tchinq-lait est leader sur le marché du lait de consommation en Algérie grâce à sa spécialisation en production du lait UHT, un lait de qualité et qui peut se conserver à température ambiante sans tourner.

1.2. Organigramme

1.2.1. Structure globale

L'organigramme de l'entreprise a été retracé pour rassembler l'ensemble des usines sous une même direction générale. Il est présenté dans la figure 1.

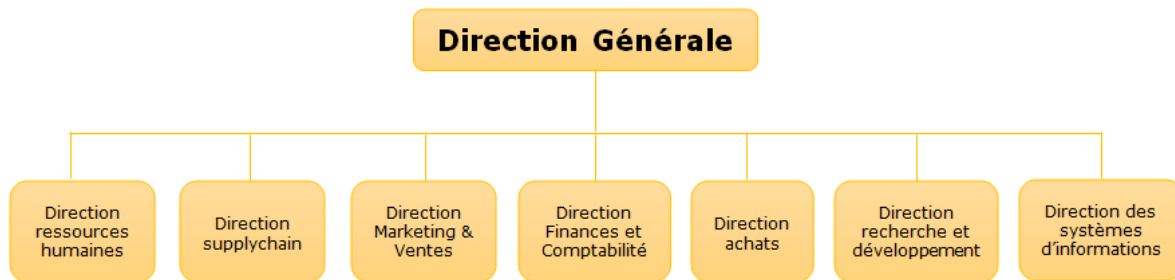


Figure 1: Organigramme de l'entreprise (Document interne, 2018)

1.2.2. Force de vente

Le département force de vente chez Tchinq-lait est le département responsable de l'action commerciale auprès des clients et des prospects; sa principale mission est de contribuer à l'évolution du chiffre d'affaire ainsi que d'augmenter les parts du marché de l'entreprise. Cela se fait par le levier de deux actions :

- La diffusion numérique (DN) : élargir quantitativement le portefeuille client et recruter de plus en plus de client.
- La diffusion valeur (DV) : développer les commandes faites par les clients déjà existants.

La force de vente est composée de deux types de personnel :

- Sédentaire : personnel recevant les clients dans les locaux de l'entreprise.
- Itinérant : personnel qui se déplace chez le client à l'extérieur des locaux de l'entreprise. C'est le type le plus présent dans le cas de Tchinq-lait. Ceci sera développé en détail dans la partie relative à la distribution.

La force de vente chez Tchinq-lait est responsable aussi de :

- Superviser, contrôler et promouvoir les ventes Tchinq-Lait auprès des distributeurs, grossistes, épiciers et supérettes.
- Coordonner et relayer les actions Marketing/Promotions auprès des intermédiaires du circuit de distribution Tchinq-Lait.

- Veille concurrentielle : faire remonter toutes les informations relatives au marché, aux concurrents et à la distribution.
- Prospecter en permanence de nouveaux clients (grossistes, supérettes et détaillants).

L'architecture et la hiérarchie du département force de vente de Tchinq-lait sont illustrées dans de la figure 2.

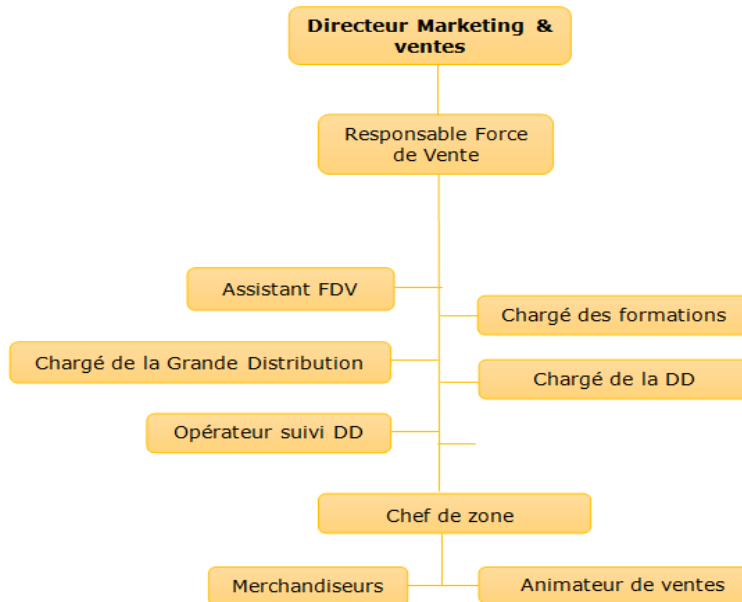


Figure 2 : Organigramme du département Marketing et Force de vente

1.3. Contrat de franchise

Ce contrat de franchise est un partenariat qui lie l'entreprise Tchinq-Lait avec Candia où chacune des parties trouve son intérêt : Candia peut, grâce à ce contrat, élargir son marché et renforcer la notoriété de ses produits à l'échelle internationale ; Tchinq-Lait, quant à elle, peut bénéficier des 40 ans d'expérience de Candia pour réussir dans sa nouvelle activité.

1.3.1. Engagement du franchiseur

L'engagement le plus important de Candia est, sans doute, le fait d'autoriser le franchiseur à produire et vendre ses produits et de porter la marque. De ce fait, Tchinq-lait pourra bénéficier de la notoriété internationale de cette marque. De plus, elle apporte le soutien technique et commercial que nécessite le franchiseur afin de réussir dans sa mission :

- Sur le plan Technique : Assistance au franchiseur pour le choix des procédés de fabrication le choix des équipements et la mise en œuvre du contrôle qualité.
- Sur le plan commercial : Assistance pour l'élaboration de plans marketing et campagnes de communication grâce à l'expérience internationale en commercialisation de Candia.

1.3.2. Engagement du franchisé

En plus des redevances annuelles (royalties représentant un certain pourcentage du chiffre d'affaires) et du droit d'entrée qu'il doit payer après la signature du contrat, le franchisé doit respecter d'autres engagements en matière d'hygiène et de qualité, qui ont pour but de préserver l'image de marque de Candia ou de se prémunir contre toute concurrence déloyale à ses produits. Le contrat entre les deux parties stipule que :

- Tous les ingrédients, les emballages et les accessoires utilisés dans la production doivent être conformes aux normes de qualité prescrites.
- Candia se réserve le droit de refuser l'intervention de tout fournisseur d'équipement ou de matière première si elle estime que celui-ci peut affecter les normes que suit l'entreprise.

1.4. Evolution de l'entreprise

1.4.1. Evolution de la production

Depuis l'entrée en exploitation en 2001, la production de l'entreprise a connu une croissance continue et très importante; cette évolution témoigne des ambitions de l'entreprise à élargir toujours plus ses parts de marché. De plus, la diversification et l'enrichissement de la gamme de produits a nettement contribué à ce résultat.

1.5. Les unités de production

L'usine de Bejaia est une laiterie moderne, construite sur une superficie totale de 6.000m², avec une capacité de production de 740.000 litres/jour. Cette unité de production comporte :

- Un atelier de production : reconstitution du lait, traitement thermique et conditionnement.
- Un laboratoire : pour les analyses microbiologique set physico-chimiques du lait.
- Les utilités : Chaudières, station de traitement des eaux, compresseurs, groupes électrogènes, onduleurs, station de froid.
- Administration Générale : Direction générale et administration, Direction marketing et vente, Direction qualité, Direction achats et approvisionnements, Direction finances et comptabilité.
- Dépôt de stockage des produits finis: pouvant contenir près de 3 millions de litres. Ce dépôt sert aussi comme plateforme d'expédition, pour la livraison des distributeurs, à travers tout le territoire national.

Cette usine emploie en 2018, plus de 540 personnes dont 29 femmes. 9% des employés sont des cadres, 40% sont des agents de maîtrise et le reste est composé d'agents d'exécution (Document interne, 2018).

L'usine implantée à Baraki s'étale sur une superficie de 30 000 m² et emploie dans ses chaînes de production plus de 310 personnes (Document interne, 2018).

1.6. Produits

La liste des produits de Tchîn-lait a évolué au fil des années, plusieurs produits ont été testés puis retirés pour finalement opter pour la gamme de produit actuel présentée en Annexe A. Une classification ABC de ces mêmes produits est développée en annexe B pour mesurer l'importance de chacun des produits et son poids par rapport au chiffre d'affaire de l'entreprise.

2. Diagnostic

Plusieurs éléments seront mis en lumière lors de cette analyse, principalement, il s'agira d'analyser le marché dans lequel évolue l'entreprise, ses canaux de distribution, la relation avec les points de vente, le flux informationnel et les données dont dispose l'entreprise.

L'entreprise Tchîn-lait évolue dans deux types de marchés qui sont le marché du lait et le marché du jus ; il est donc judicieux de faire un état des lieux concernant ces deux industries.

2.1. Diagnostic externe

2.1.1. Marché du lait

a. Caractéristiques

Avec une dépense de près de 2 Milliards de dollars sur le lait en poudre en 2014, enregistrant une croissance de plus de 60% par rapport à l'année 2013 (LaitDZ, 2018), l'Algérie se positionne parmi les plus grands importateurs de cette matière et occupe une 4^{ème} place derrière la Chine et la Malaisie. L'Algérie a consommé près de 17% des échanges mondiaux du lait en poudre durant la période 2009-2014, confirmant ainsi son appétit pour ce produit (Yahiaoui, 2017).

Toujours en prenant la Chine et la Malaisie comme repères, les grandes quantités de lait en poudre importées par ces deux pays sont principalement destinées à la production des produits laitiers qui sont par la suite exportés (environ 40% de la production est exportée). L'Algérie, par contre, importe cette matière première qui est consommée localement. L'exportation ne touche que deux pays voisins qui sont la Libye et la Mauritanie avec un faible flux qui ne dépasse pas les 600.000\$ par an (Yahiaoui, 2017).

Selon une étude, plus de 80% des entreprises existantes sur le marché fonctionnent à un régime inférieur à 75% de leurs capacités. L'étude a révélé que 64% des entreprises produisent du lait pasteurisé conditionné (lait en sachet subventionné par l'état), 25% sont spécialisées dans la production de lait cru local et pas plus de 11%, dont Tchîn-lait, produisent le lait UHT (Yahiaoui, 2017).

Des statistiques ont montré qu'en Algérie, la consommation du lait, par personne, est la plus élevée dans les pays du Maghreb, affichant ainsi une moyenne annuelle de 130 L/habitant contre 85 au Maroc et 65 en Tunisie. S'il est prévu que le taux garde sa stabilité, la consommation totale est en croissance suivant ainsi la croissance démographique du pays (LaitDZ, 2018).

b. Subventions de l'état

Dans le but de réduire la facture d'importation et afin de dynamiser ce secteur, plusieurs actions ont été entreprises par l'état, nous citerons :

- Agir en amont en appuyant la production du lait de vache.
- Subventionner les acteurs de la logistique, en particulier le transport à partir des centres de collecte.
- Subventionner la production.

Aujourd'hui, et grâce à cette politique, l'Algérie produit annuellement près de 2.2 milliard de litres de lait de vache grâce aux dix entreprises existantes. Ce chiffre demeure très faible par rapport à la demande globale (LaitDZ, 2018).

c. Elasticité de la demande

Une étude réalisée par l'économiste français D. Bergman révèle que la consommation de lait présente sur le long terme une stabilité, malgré des fluctuations dues à la saisonnalité.

Dans cette même étude, il a été vérifié que la demande du lait est faiblement impactée par la variation des prix ce qui affirme une faible élasticité.

Pour des raisons techniques et économiques, les producteurs de laits opèrent dans un marché à client unique (monopsonne) et un nombre limité de firmes (oligopole) (Bergman, 2016).

d. Pénuries

Depuis de longs mois, les pénuries de lait en sachet (coûtant 25DA) sont devenues très fréquentes et font partie du quotidien du citoyen algérien. Les enquêtes qui avaient pour but de révéler les raisons de ces pénuries répétitives ont divergé sur la cause principale. En conséquence de la politique d'austérité appliquée par le gouvernement depuis 2015 ou à cause d'une mauvaise gestion par les distributeurs, les conséquences sont les mêmes : la pénurie est une réalité et le consommateur en pâtit (AlgérieEco, 2018).

Face à cette situation, une bonne partie des consommateurs trouvent l'alternative dans les boîtes de lait en carton (entre autres Candia), engendrant ainsi une forte demande sur ces produits. « A chaque fois qu'il y a pénurie de lait, la demande sur le lait conditionné dans les boîtes en carton s'accroît fortement. La crise profite donc à certains » déclare un responsable de l'Union Générale des Commerçants et Artisans Algériens (El-Watan, 2018)

2.1.2. Marché du jus

Pour ce qui est du jus, le marché algérien a été qualifié par des experts de l'union européenne de croissant et à potentiel élevé avec des perspectives de développement certaines. Un marché où le volume d'écoulement annuel est de 2.4 milliards de litres avec un budget de 104.8 milliards de dinars et qui affiche une croissance avoisinant les 10% pendant les 5 dernières années (BSA, 2018).

L'Algérie évolue dans le secteur des boissons rafraichissantes sans alcool (BRSA) vers un degré de concentration du marché proche de celui observé en Europe, même si les leaders (NCA Rouïba, Ramy, N'Gaous) s'emparent de 80% des parts du marché.

La production nationale couvre majoritairement la demande et l'importation est marginale. Menacées par la fluctuation des prix des matières premières (eau, PET, ...), les entreprises se retrouvent dans l'obligation d'innover constamment afin de réduire leurs coûts et de faire face à la concurrence.

Les consommateurs affirment avoir l'habitude de consommer régulièrement du jus avec un pourcentage de 75% chez les enfants, 66% chez les adolescents et 43% chez les adultes. Ceci nous mène à conclure que la consommation des jus dépend des paramètres (BSA, 2018):

- Démographiques : taux de naissances, ...
- Economiques : évolution des revenus des ménages, ...
- Réglementaires : lois sur la composition, étiquetage, ...

Tchin-lait commercialise beaucoup de produits sur des marchés différents, ceci est dû à sa stratégie de diversification et à sa volonté d'exister sur le marché du jus.

Des entretiens effectués avec le personnel du département marketing ont relevé que le mois de ramadhan et la saison estivale sont deux périodes dynamiques caractérisés par une hausse globale des ventes. Sans préciser les produits impactés ou le degré de cette hausse pour chacun des produits.

2.2. Diagnostic interne

2.2.1. Distribution

a. Réseau de distribution

A la sortie de l'usine, les produits de Tchin-lait traversent deux étapes intermédiaires qui sont, les dépositaires puis les points de ventes, avant d'arriver chez le consommateur final.

La flotte de Candia est responsable de l'approvisionnement des dépositaires, qui eux à leur tour, se chargent d'alimenter les points de vente.

Le schéma présenté dans la figure 4 illustre le réseau de distribution chez Candia.

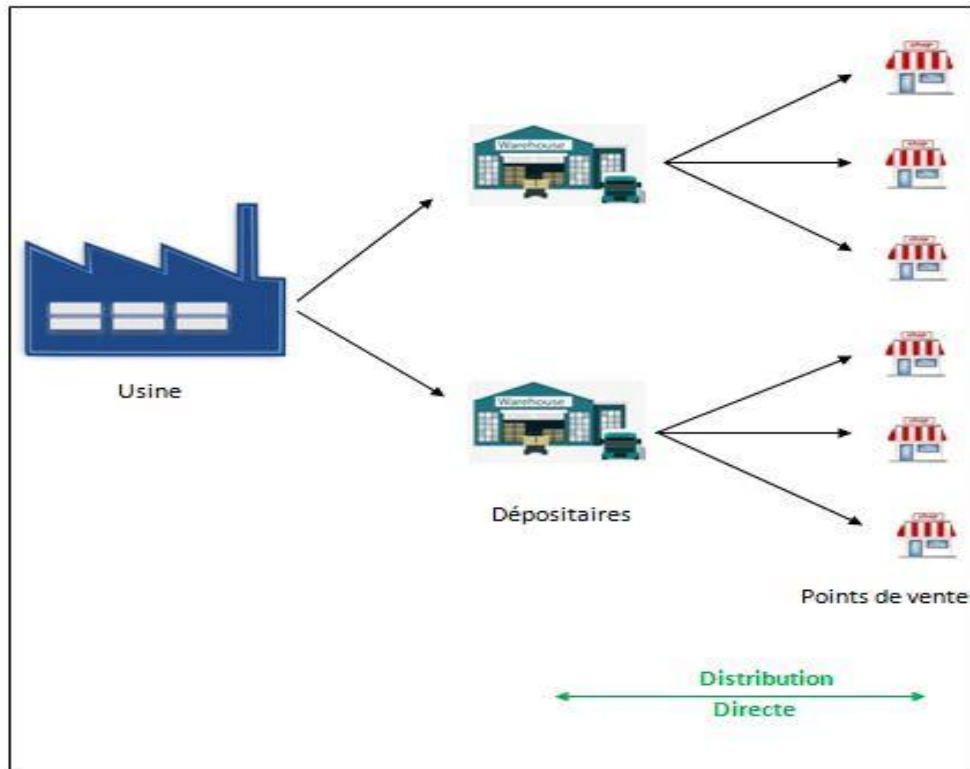


Figure 3 : Réseau de distribution chez Candia

b. Les acteurs de la distribution

- **L'usine**

- Il existe deux usines opérationnelles, une à Bejaia et une à Alger
- Elles produisent 27 références de produits
- Un projet d'une usine à Sétif est en cours réalisation.

- **Les dépositaires**

- Il existe 52 dépositaires sur le territoire national qui sont répartis par région : 14 au Centre, 16 à l'Est, 10 à l'Ouest et 12 au Sud. La liste des dépositaires par wilaya est jointe en annexe C.
- Ce sont des entrepôts de différentes capacités de stockage.
- Les dépositaires travaillent en étroite collaboration avec Candia ; ils sont reliés par des contrats avec l'entreprise après une sélection minutieuse.
- Chaque dépositaire possède une flotte de camions qui assure la livraison vers les points de ventes.

- **Les vendeurs**

- Ce sont les camionneurs. Ils assurent la livraison des produits depuis les entrepôts jusqu'aux points de vente.
- Les vendeurs sont liés aux dépositaires par des contrats.
- Il existe deux types de vendeurs :
 - * Type A : pour les laits blancs (4 références).
 - * Type B : pour le complément de gamme (23 références).
- Chaque vendeur est responsable de 5 circuits différents. Un circuit est un ensemble de points de vente situés dans une même région et visités une fois par semaine (en temps normal).
- Chaque vendeur est équipé d'un appareil PDA(Personal Digital Assistant) qui lui permet d'enregistrer toutes les ventes réalisées
- Chaque fin de journée, le vendeur regagne le dépôt pour faire les comptes de la journée et charger les produits du lendemain.

- **Les points de vente**

- Ce sont les superettes, les hypermarchés, les kiosques, ... Ils sont la dernière étape avant que les produits n'arrivent chez les consommateurs finaux.
- Ils sont aussi appelés « clients »

- **Les superviseurs**

Ce sont les hommes du terrain pour Candia. Ils ont pour mission de :

- Contrôler les dépositaires et les vendeurs.
- Visiter les points de vente et s'assurer de leur bon approvisionnement.
- Relever les anomalies dans la distribution

c. « ASSABIL », outil de gestion de la distribution

L'entreprise Tchou-lait contrôle le flux informationnel grâce à un dispositif très performant constitué d'une solution logicielle « ASSABIL » qui est connecté en temps réel avec les appareils PDA. Grâce à cet outil, le département force de vente possède une bonne visibilité sur tout le maillon de la distribution.

Parmi les informations que l'entreprise peut récolter grâce à ASSABIL on peut citer:

- L'identifiant du vendeur, le point de vente et la quantité vendue pour chaque transaction en temps réel.
- La quantité de stock chez les dépositaires.
- La quantité de stock restante chez le vendeur.
- Les points de vente visités et non visités, et le vendeur responsable de la visite.

2.2.2. Prévision de vente et méthodes de chargement

Durant notre visite chez l'un des dépositaires situé à Bejaia, et après des entretiens avec les vendeurs et le chef de zone, nous avons pu comprendre comment un vendeur détermine les quantités prévues pour alimenter les points de vente de son circuit. Dans les faits, les vendeurs décident des quantités à transporter en se basant uniquement sur leur expérience. Les prévisions des ventes à ce niveau sont donc faites de manière qualitative. Une pratique qui donne souvent lieu à des taux de retour élevés vu que les vendeurs essaient généralement de charger le maximum afin d'éviter les retours aux dépôts (obligatoire en cas de rupture pour n'importe quel produit).

Il convient de préciser que pour chaque point de vente, une quantité $Q(n)$ pour la semaine (n) est prévue sur le PDA du vendeur, mais que ce dernier ne peut être au courant de cette information que lorsqu'il est chez le client.

La quantité $Q(n)$ est calculée selon la formule suivante :

$$Q(n) = 1.2 * [Q(n - 1) + Q(n - 2) + Q(n - 3)] / 3$$

Où : $Q(n-1)$; $Q(n-2)$; $Q(n-3)$ sont les quantités achetées par le point de ventes lors des trois semaines précédentes.

- Si le client effectue un achat inférieur à la quantité $Q(n)$, le camionneur demande au client à effectuer un prélèvement de stock
- Si la quantité de stock existant + quantité achetée < $Q(n)$. Alors le client est programmé pour une deuxième visite dans la même semaine.

2.2.3. Processus de vente

Les étapes du processus de vente sont présentées dans la figure 5.

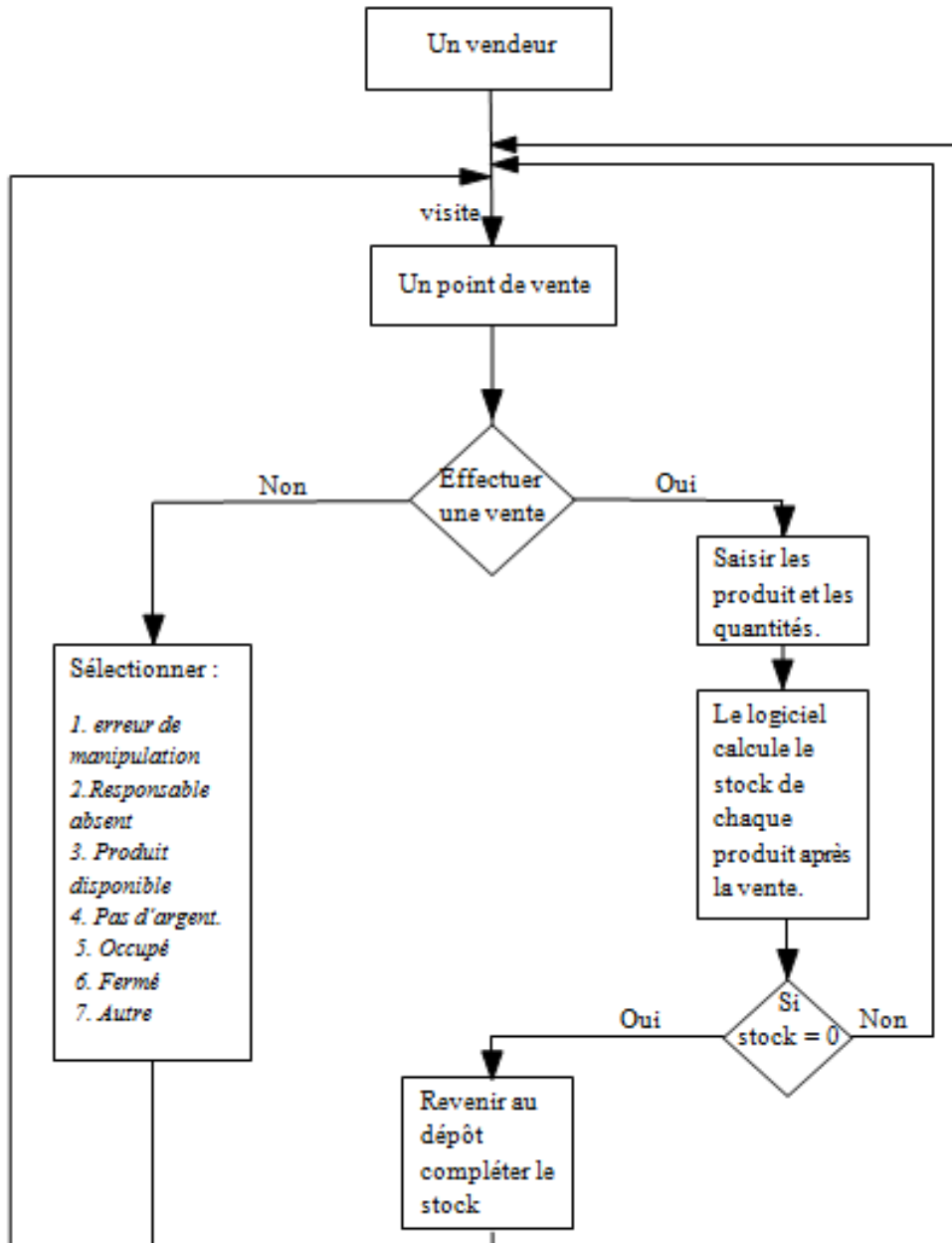


Figure 4 : Schéma du processus de vente

2.2.4. Analyse des données

Les données sont un élément indispensable pour comprendre l'environnement d'une entreprise, son évolution et le comportement de ses clients. L'importance des données est encore plus grande lorsqu'il s'agit d'un système intelligent et dynamique comme c'est notre cas. En réalité, la qualité de tout le système repose principalement sur des données fiables, complètes et en quantité suffisante.

Il est donc nécessaire de consulter la base de données de Tchik-lait, et principalement les enregistrements des ventes car c'est la partie qui nous intéresse dans ce projet.

L'entreprise a fourni un document qui représente les ventes depuis le 01/01/2017 jusqu'au 31/04/2018. Le document regroupe toutes les ventes qui ont été effectuées durant cette période, dans la ville de Tizi-Ouzou et ce pour l'ensemble des produits.

Dans la partie qui suit, plusieurs traitements seront réalisés dans l'objectif d'avoir une idée sur le comportement des points de ventes, les tendances, les similitudes, les problèmes et autres informations qu'il faudra prendre en compte dans cette étude.

a. Analyse du comportement

Dans cette partie, il s'agit de vérifier si les points de vente adoptent un comportement similaire dans leurs politiques d'approvisionnement ; ceci permettrait de mettre en place un seul système qui modélise leur comportement et qui est applicable pour tous. Dans le cas échéant, chaque point de vente nécessitera une étude spécifique qui aboutirait à un modèle spécifique. Une deuxième analyse sera effectuée, cette fois sur un même point de vente, dans l'objectif de vérifier s'il adopte une même politique d'approvisionnement pour tous les produits.

- **Par produit**

Les tests réalisés sur les tendances d'approvisionnement d'un même point de vente par rapport aux différents produits ont montré que le client (point de vente) n'a pas un comportement unifié pour tous les produits. En effet, chaque produit a sa propre tendance avec des périodes de pics, de hausse et de baisse spécifiques. La figure 6 ci-dessous est un exemple qui montre l'évolution des achats de 4 produits différents par un même point de vente.

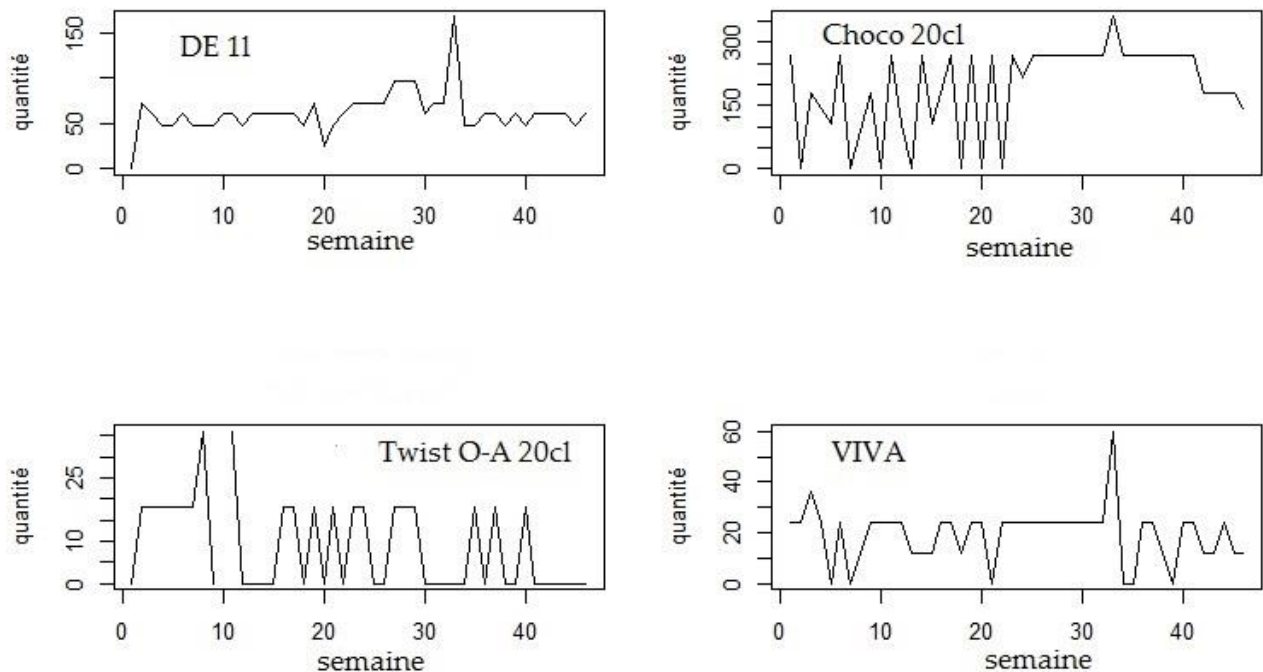


Figure 5. Evolution des approvisionnements de 4 produits différents chez un même point de vente

• **Par point de vente**

Comme pour le test précédent, celui-ci a pour but de vérifier des similitudes dans l'approvisionnement des différents points de vente avec un même produit. Les résultats sur la figure 7 ont montré qu'il existe une différence dans la politique d'achat, les hausses et les baisses correspondent à des périodes différentes chez chaque point de vente.

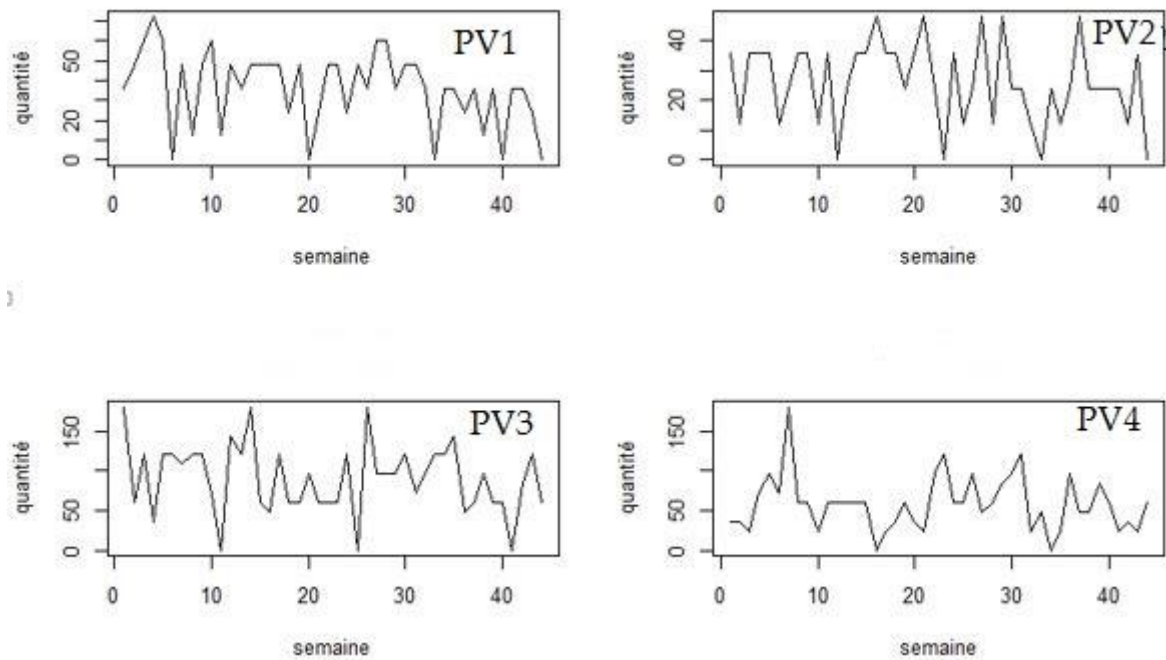


Figure 6. Evolution des approvisionnements du produit DE 11 chez 4 points de vente différents

Synthèse : Les deux résultats exposés ci-dessus permettent de conclure que :

- Il ne peut exister un modèle unique qui régit le comportement d'achat d'un même produit par tous les points de vente
- Il n'existe pas un modèle unique qui modélise le comportement d'un point de vente pour tous les produits

Ainsi, faire des prédictions sur l'ensemble des points de vente reviendrait à élaborer plus de 50.000 modèles (2100 points de vente * 27 produits), ce qui ne peut être envisagé ni dans le cadre de ce PFE ni par l'entreprise vu sa difficulté technique et ses coûts économiques.

b. Problèmes contraignant la prévision

• **Fréquence des achats**

Nous constatons d'après la figure 8 que plus de 25% des points de ventes ont effectué moins de 20 achats (l'achat d'au moins un produit lors d'une visite est considéré comme un achat, se référé au paragraphe 2.2.3.) en une année alors qu'ils sont visités avec une fréquence hebdomadaire. Une analyse approfondie sur ces points de vente a montré aussi une irrégularité dans la fréquence des achats. Cette information souligne la difficulté de prédire le

comportement des points de vente qui effectuent des achats isolés et de manière non régulière. En effet, ces points de vente peuvent causer des ruptures chez le vendeur en cas d'achat intense, comme ils peuvent être la cause d'un retour important de produits.

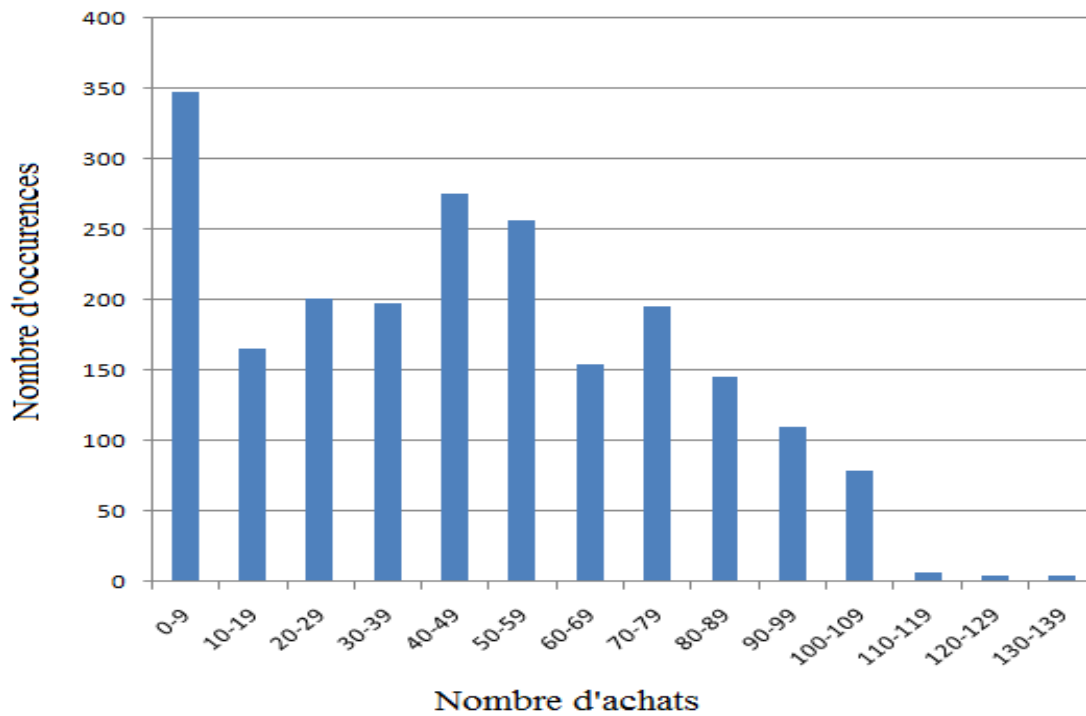


Figure 7. Fréquence des achats

- **Ruptures**

Quel que soit l'outil qui sera retenu pour répondre à cette problématique, ils partagent tous un point commun qui est « l'apprentissage par les données ».

Durant l'analyse effectuée sur les ventes des différents produits, nous avons constaté que plusieurs produits se distinguent par de longues périodes de rupture. Cela représente un handicap de taille lorsqu'il s'agit d'effectuer des prédictions quantitatives et fiables sur les produits qui ont un manque important dans les données. Par conséquent, plusieurs produits ne peuvent pas être intégrés dans cette étude.

La figure 9 montre que les périodes de rupture enregistrées pour 6 produits varient de 10 à plus de 40 semaines.

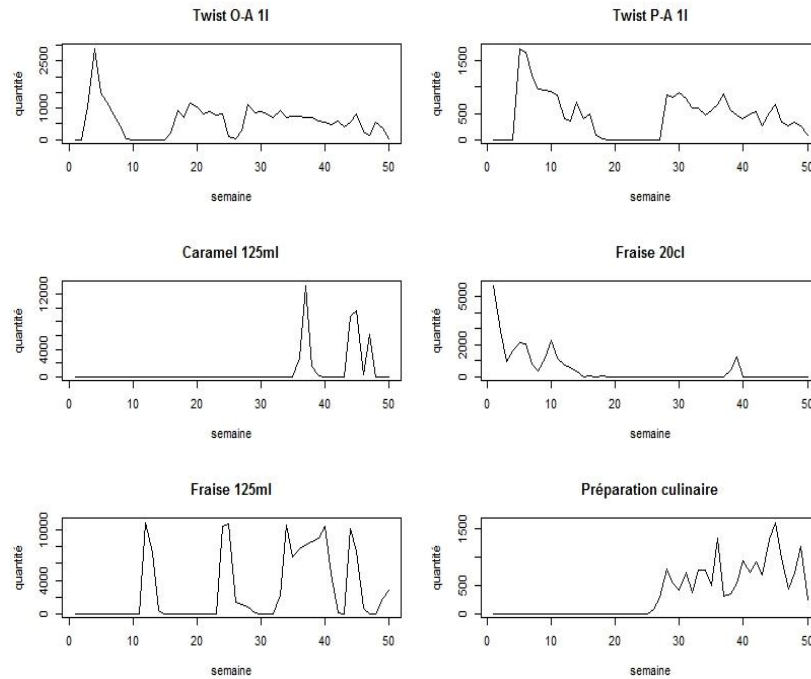


Figure 8. Les ruptures dans certains produits

Conclusion

L'analyse externe et interne de l'entreprise a permis de comprendre l'environnement de l'entreprise, le fonctionnement de la distribution ainsi que la relation de l'entreprise avec ses points de vente.

L'analyse du marché du lait a montré que la vente de ce produit est caractérisée par une certaine stabilité par rapport à la saisonnalité mais qu'elle est impactée par d'autres événements comme les ruptures chez les concurrents, un événement qui peut impacter les ventes d'une manière très importante.

Les entretiens avec le personnel de l'entreprise ont révélé quelques indices sur d'autres paramètres pouvant influencer la vente, et surtout la vente des jus, en l'occurrence le mois de Ramadhan et la saison estivale.

L'analyse des données a montré qu'il existe une répulsivité dans le comportement des points de vente et par conséquent il ne serait pas possible de réaliser un seul modèle qui soit applicable pour tous les points de ventes. Ainsi, une prédiction à un niveau de détail aussi poussé devient une mission très compliquée techniquement puisqu'il faut élaborer plus de 50.000 modèles pour une seule ville.

L'analyse des données a révélé aussi que, à l'heure actuelle et en fonction des données existantes, il ne serait pas possible de faire des prédictions par point de vente en raison d'un manque dans les données et l'aléa qui est inhérent aux comportements des clients.

Dans le chapitre suivant, il s'agira d'explorer les outils de l'apprentissage automatique qui peuvent être appliqués pour répondre à la problématique.

Chapitre II : ETAT DE L'ART :
Machine Learning et réseaux de
neurones

Introduction

Depuis quelques années, l'ascension de l'intelligence artificielle et principalement l'apprentissage automatique (Machine Learning) a donné une nouvelle dimension à la relation de l'homme avec la machine. Cet outil a intégré un nouveau paradigme de programmation qui consiste en l'apprentissage des machines par les données, les machines ne sont donc pas explicitement programmées, contrairement à la programmation classique qui consiste à exécuter une séquence de commandes.

L'apprentissage automatique est appliqué sur un large éventail d'applications comme la reconnaissance des formes, la classification, la prédiction, ... mais avant de développer encore plus cet outil, nous allons faire un brève aperçu sur les techniques de prévision (équivalent à la prédiction en machine learning) qui ont été les plus utilisées dans le passé.

1. Historiques de la prévision mathématique

Dans les prévisions économiques, les séries chronologiques figurent parmi les techniques incontournables pour donner les appréciations futures d'une variable quantitative. Ces techniques développées en Annexe D reposent sur l'analyse des observations historiques de la même variable afin d'élaborer un modèle décrivant les relations sous-jacentes. Le modèle est ensuite utilisé pour extrapoler la série temporelle dans le futur (Zhang, 2004).

Il existe différentes approches de modélisation basée sur les séries chronologiques. Les modèles statistiques traditionnels tels que la moyenne mobile, le lissage exponentiel et ARIMA (Auto Régressive Integrated Moving Average) ont été les principaux axes de recherche et les outils les plus appliqués durant les dernières décennies et ce, en raison de leur relative simplicité théorique et de mise en œuvre.

Bien qu'une certaine amélioration ait été notée avec ces modèles, le gain de leur utilisation pour les problèmes généraux de prévision est limité car ils ont été développés sur et pour des modèles non-linéaires spécifiques et sont par conséquent incapables de s'adapter à de nouveaux problèmes plus complexes (Zhang, 2004).

Depuis quelques années, l'ascension de l'intelligence artificielle et principalement l'apprentissage automatique (Machine Learning) a donné une nouvelle dimension aux problèmes de prédiction. En effet, cet outil a fondamentalement changé la donne en adoptant une approche « *data-driven* » (piloté par les données) qui vient remplacer les techniques classiques qui sont des « *rule-based model* » (modèles explicitement programmés).

En effet, l'avantage de l'apprentissage automatique réside dans le fait qu'ils ont la capacité de s'adapter avec les données et de se corriger avec « l'expérience », fournissant ainsi des résultats plus performants avec l'augmentation dans la quantité de données qui les alimente. Contrairement aux techniques précédemment citées qui sont figées et qui nécessitent une identification, et une description, plus ou moins formelle, des composantes de la série chronologique observée.

2. L'apprentissage automatique

2.1. Introduction

« Machine Learning » ou apprentissage automatique est une technique d'analyse de données qui consiste à enseigner à des ordinateurs ce dont les humains sont naturellement capables de faire: tirer des leçons de leurs expériences. Les algorithmes de Machine Learning font appel à des méthodes de calcul pour « apprendre » des informations directement à partir des données sans qu'il ne soit nécessaire de s'appuyer sur une équation prédéterminée en tant que modèle. Les algorithmes s'adaptent et deviennent plus performants au fur et à mesure que le nombre d'échantillons disponibles pour l'apprentissage augmente » (Mathworks, 2018).

Yoshua Bengio, Professeur au département informatique et recherche opérationnelle à l'université de Montréal donne la définition suivante « l'apprentissage automatique fait partie de la recherche sur l'intelligence artificielle, cherchant à transmettre des connaissances aux ordinateurs à l'aide de données, d'observations et d'interactions avec le monde. Cette connaissance acquise permet aux ordinateurs de généraliser correctement sur les nouvelles situations » (Bengio, 2017).

L'essor de cet outil est dû aux conséquents développements qu'ont connus les technologies supports. En effet, l'évolution des machines de calcul, la disponibilité des algorithmes appliqués sur des modèles mathématiques et les quantités gigantesques de données générées quotidiennement par l'activité humaine ont grandement contribué à booster cette discipline qui existait pourtant depuis les années 80s.

Il existe plusieurs algorithmes d'apprentissage faisant partie du *Machine Learning* ; ils sont tous regroupés dans deux catégories :

- L'apprentissage supervisé.
- L'apprentissage non supervisé.

2.2. Types d'apprentissage automatique

2.2.1. Apprentissage supervisé

Les techniques de l'apprentissage supervisé sont caractérisées par la présence de données d'entrée et de données de sortie pour la construction du modèle et chaque individu se présente sous la forme du couple (input-output). L'algorithme fait associer la variable de sortie de chaque individu avec sa ou ses valeurs d'entrée et fait entraîner le modèle à réaliser des prédictions fidèles sur les données futures (Murphy, 2012).

L'échantillon \mathcal{D} de taille N impliqué dans l'apprentissage du modèle se présente comme suit :

$$\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^N$$

Tel que :

- $X_i = (x_{i1}, x_{i2}, \dots, x_{im})$: vecteur d'entrée de l'individu i et m le nombre de variables
- Y_i = variable à prédire.

Après une étape d'entraînement, le modèle doit être testé avec un second échantillon de données où, dans cette étape, il reçoit seulement les variables d'entrée et fait des prédictions sur les sorties. Ces prédictions sont ensuite comparées avec les résultats réels pour mesurer la performance du modèle. Plus les prédictions du modèle sont proches des valeurs réelles plus le système est performant.

Ces techniques sont appelées ainsi car les données de sortie permettent à l'algorithme de comparer et de se corriger au fur et à mesure. En d'autres termes, elles vont « guider » l'algorithme vers les résultats qui doivent en ressortir.

Lorsque la valeur de sortie est une variable continue on parle d'une *prédiction*. Par exemple, anticiper la température dans une période de l'année est un problème de prédiction. Par contre, si l'output est une variable discrète alors il s'agit d'une *classification* (exemple : classifier les clients dans deux catégories Fiables/Non-Fiables ou classifier les cellules cancéreuses en Bénigne/Maline) (Youcef, 2014).

Les algorithmes qui sont utilisés pour résoudre les problèmes de classification incluent notamment les machines à vecteurs de support (SVM), la classification naïve bayésienne, la méthode des k plus proches voisins, ... (Mathworks, 2018).

Concernant les algorithmes de régression, on citera la régression linéaire, neuro flou adaptatif, les réseaux de neurones, ...

2.2.2. Apprentissage non supervisé

Pour l'apprentissage non-supervisé, seules les données d'entrée sont disponibles pour la construction du modèle, le but étant de trouver des informations cachées et de faire émerger des similitudes complexes à l'intérieur de ces données. *Le clustering*, qui consiste en la séparation des individus dans des groupes homogènes, est la principale technique de l'apprentissage non-supervisé (Youcef, 2014).

L'avantage de cette méthode qui est connue comme un outil d'exploration, est qu'elle peut apporter de nouvelles informations dont l'utilisateur ignorait l'existence. Un exemple pratique dans la segmentation des clients, l'algorithme peut renvoyer des groupes en fonction de certains paramètres qu'un humain n'aurait jamais pris en compte à cause des préjugés préétablis et à cause d'une inertie psychologique.

Contrairement à la première technique où le modèle peut comparer les outputs prédits avec les outputs réels, il est moins évident de mesurer la performance d'un modèle de clustering car les valeurs de sortie correctes ne sont pas connues au préalable. La performance dépend donc de si des informations intéressantes dans les données ont été révélées par les clusters ou non (Youcef, 2014).

Parmi les algorithmes utilisés en apprentissage non-supervisé nous citerons l'algorithme de K-moyennes, le clustering hiérarchique, les modèles de Markov cachés(Mathworks, 2018).

La figure 10 résume la classification des algorithmes d'apprentissage automatique.

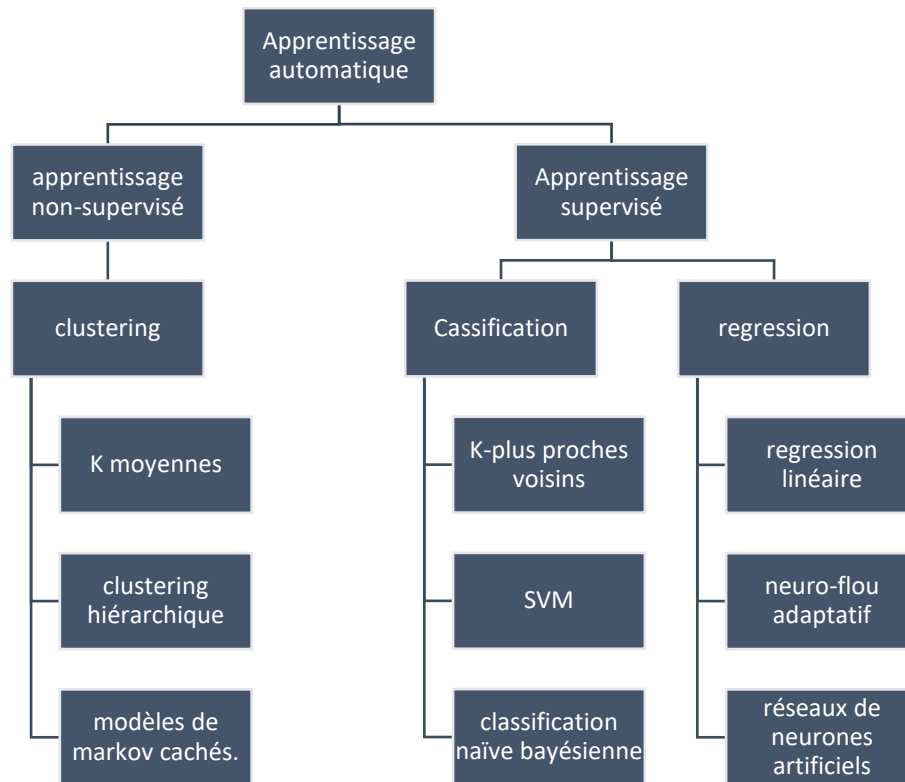


Figure 9. Classification des algorithmes du machine learning

2.3. Intérêt de l'apprentissage automatique

On pourrait se demander quelle est la raison derrière tout cet engouement mondial pour le *Machine Learning* ? Pourquoi toutes les disciplines cherchent à intégrer cet outil dans la résolution des problèmes alors que depuis de longues années les ingénieurs ont conçu des programmes ingénieux qui ont grandement impacté la vie des individus. En réalité, il existe plusieurs raisons techniques qui expliquent cela dont voici une liste non exhaustive (Nilson, 2012):

- Nous sommes incapables de définir quelques phénomènes d'une manière correcte et précise sinon par des exemples. En présentant des échantillons sous une forme entrée/sortie, la machine est capable d'ajuster sa structure afin de donner un modèle qui généralise et qui peut résoudre les futurs problèmes.
- Il est possible que des relations complexes soient cachées à l'intérieur des grandes quantités de données; l'apprentissage automatique (plus spécialement l'apprentissage profond) peut être utilisé pour révéler ces relations non aperçues par l'analyste.
- Certaines tâches peuvent être trop grandes pour un codage explicite.

- l'instabilité et l'évolution rapide de l'environnement poussent les utilisateurs à compter plus sur des machines dynamiques qui s'adaptent avec les changements plutôt que sur des systèmes figés, qui ont continuellement besoin d'être ajustés et révisés.

2.4. Comparaison entre programmation classique et apprentissage automatique

Traditionnellement, un programme est modélisé comme une série de commandes se déroulant d'une manière séquentielle et procédurale, indiquant que le programme doit choisir un chemin spécifique. Entre temps les données sont au repos.

En revanche, la programmation de type *data driven* (ou *dataflow*) met l'accent sur le mouvement des données et façonne le programme (Johnson, 2004).

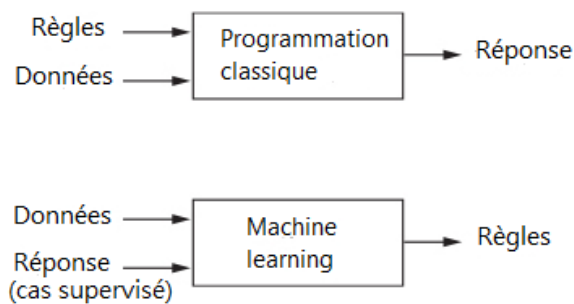


Figure 10. Différence entre la programmation classique et l'apprentissage automatique (Shukla, 2018).

3. Les réseaux de neurones artificiels

3.1. Introduction

Les réseaux de neurones artificiels (RNA) sont des modèles informatiques qui ont été inspirés des systèmes nerveux biologiques (Annexe F) pour réaliser des tâches de reconnaissance de formes et autres traitements informatiques.

Un RNA est un réseau composé de nombreuses unités de calcul simples appelées neurones ou cellules, qui sont fortement interconnectées et organisées en couches. Chaque neurone effectue une simple tâche de traitement de l'information en convertissant des entrées reçues en des sorties. Grâce aux arcs de liaison entre ces neurones, la connaissance peut être générée et stockée.

Les recherches scientifiques réalisées sur les réseaux de neurones artificiels (RNA) ont affirmé que cet outil possède une grande capacité de prédiction et s'inscrit parmi les méthodes de modélisation mathématique les plus robustes.

Leur application sur des problématiques réelles et dans différents domaines a donné une grande satisfaction et a confirmé ainsi les conclusions scientifiques. Aujourd'hui, les RNA

sont utilisés pour anticiper l'évolution des indices boursiers, la consommation énergétique, la vérification d'authenticité, ... (Zhang, 2004).

Les RNA présentent plusieurs avantages qui permettent de les propulser parmi les alternatives les plus prometteuses dans la prédiction. En effet, la structure inhérente aux réseaux leur donne la capacité d'approcher n'importe quelle fonction dans un intervalle fermé. Les RNA peuvent non seulement modéliser les formes linéaires mais sont aussi très performants face à des modèles non linéaires quelconques comme cela a été prouvé dans *the Universal approximation theorem* (Hornik, 1990).

Un autre aspect très important est que les RNA sont des méthodes de type *data driven* qui n'exigent pas beaucoup d'hypothèses ou de contraintes sur le processus en question. Cette caractéristique de «l'apprentissage par les données ou par l'expérience» est hautement souhaitable dans diverses situations de prédiction où les données sont généralement plus faciles à collecter, mais sont régies par un mécanisme pas clair et pas spécifié (Danilo, 2001).

3.2. Propriétés fondamentales des réseaux de neurones

Deux caractères fondamentaux intrinsèques aux réseaux de neurones sont abordés dans ce qui suit. Ces propriétés justifient le choix de cet outil pour notre problème.

3.2.1. Approximation universelle

« Toute fonction bornée suffisamment régulière peut être approchée avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire » (Hornik, 1990).

Cet avantage qui, certes représente un atout considérable, n'est pas une spécificité exclusive aux réseaux de neurones artificiels et plusieurs autres familles de fonctions paramétrées partagent cette propriété. L'avantage réel des RNA réside dans le caractère « parcimonieux » de l'approximation (Dreyfus, 2004).

3.2.2. Approximation parcimonieuse

« À un même degré de précision, certains réseaux de neurones nécessitent moins de paramètres à ajuster que les autres approximateurs » (Barron, 1993).

En effet, le nombre de paramètres nécessaires pour atteindre une approximation fiable demeure un critère très important dans le choix de l'approximateur. Le concepteur doit toujours opter pour l'approximateur avec le plus faible nombre de paramètres possible, dans ce cas on dit que nous cherchons l'approximation la plus parcimonieuse (Dreyfus, 2004).

Intérêt du caractère parcimonieux

Il est utile de préciser que l'approximateur parcimonieux en paramètres est aussi parcimonieux en données (Dreyfus, 2004).

Il sera développé ultérieurement que l'apprentissage du réseau de neurones consiste en l'ajustement des paramètres à partir des exemples. Dans le but de minimiser la fonction du « coût », le réseau doit traiter plusieurs cas avant d'arriver à une combinaison optimale des paramètres. Or, dans beaucoup de problèmes, disposer d'une grande quantité de données peut être une mission très compliquée (travailler sur des événements à fréquence hebdomadaire ou mensuelle par exemple), ce qui conforte le besoin d'un outil qui réalise une fonction avec le minimum de ressources (Dreyfus, 2004).

3.3. Fonctionnement d'un réseau de neurones

Un réseau de neurones artificiel est un ensemble de structures élémentaires appelées les « neurones ou nœuds » qui transforment un signal d'entrée x en un signal de sortie y en réalisant une opération appelée fonction de transfert (ou fonction d'activation).

Les neurones qui composent un réseau sont arrangés dans des couches. Généralement, un réseau est composé de trois couches appelées : couche d'entrée, couche interne et couche de sortie. Il existe des modèles avec plusieurs couches internes, on parle alors d'un apprentissage profond ou *deep Learning*.

Chaque neurone du réseau est connecté avec tous les neurones qui composent la couche suivante, cette connexion est pondérée par une valeur ω_{ij} qui représente le poids de la liaison entre le neurone i et le neurone j . L'information circule de la couche d'entrée vers la couche de sortie en passant par la couche interne (Laurne, 1993).

Chaque réseau de neurones artificiel est un modèle unique qui est caractérisé par :

- Le nombre de couches internes du réseau
- Le nombre de neurones dans chaque couche
- Les paramètres ω_{ij} de chaque connexion
- L'entrée inhibitrice θ_j
- La fonction de transfert

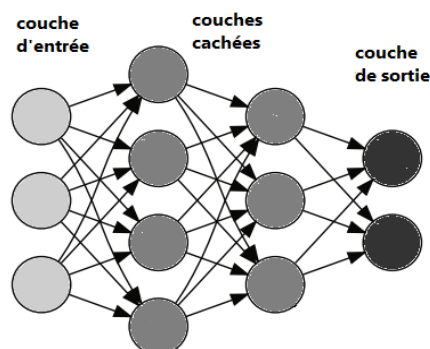


Figure 11. Architecture d'un RNA (Zhang, 2004).

3.4. Le neurone formel

Comme mentionné précédemment, le neurone est l'élément chargé de transformer un signal d'entrée en un signal de sortie. Les valeurs sur lesquelles agit le neurone sont appelées les entrées du neurone, le résultat de tout le processus est appelé sortie du neurone.

Cette opération se fait en deux étapes résumées dans la figure 13 (Laurne, 1993).

1. Réaliser une somme pondérée des signaux d'entrée :

$$Z = \omega_1x_1 + \omega_2x_2 + \dots + \omega_nx_n$$

2. Exécuter la fonction de transfert sur Z :

$$Y = f(Z + \theta)$$

f : fonction d'activation ou fonction de transfert

θ : biais ou seuil.

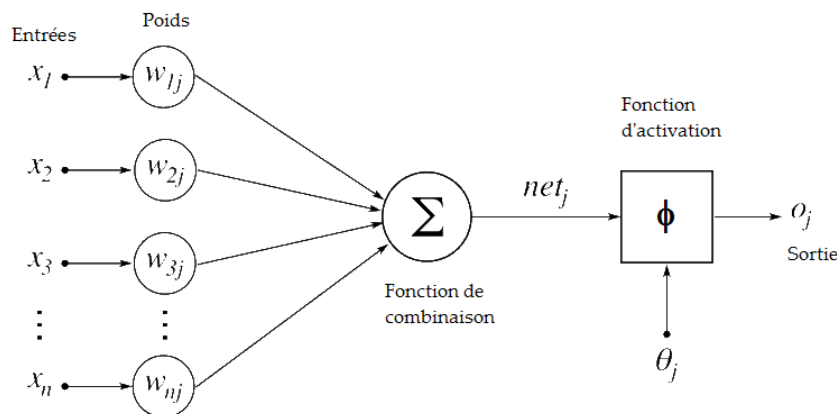


Figure 12. Schéma d'un neurone artificiel (Aouedad, 2018).

3.5. Fonctions de transfert

En biologie, la fonction d'activation peut être imaginée comme une fonction binaire OUI/NON ou 0/1 ; elle porte le nom de *Thereshold activation function* et n'est activée que si la somme des signaux d'entrée est supérieure à un seuil. Cette fonction qui peut être très indiquée pour expliquer le phénomène biologique est rarement utilisée dans les réseaux de neurones artificiels.

Les fonctions non-linéaires sont les plus fréquemment utilisées grâce à leurs caractéristiques de modélisation des signaux. Parmi ces fonctions, il y a les sigmoïdes (la fonction logistique et la fonction tangente hyperbolique Tanh) et plus récemment, la fonction ReLu. Ces fonctions ont la capacité de généraliser ou de s'adapter à une variété de données tout en différenciant la sortie (Sharma, 2017).

Afin qu'une fonction de transfert puisse être appliquée de manière efficace, elle doit vérifier les deux critères suivants :

- Différentiable : un changement dans les x implique un changement dans y
- Fonction monotone : une fonction qui est toujours croissante ou toujours décroissante.

3.5.1. Fonction logistique

Elle a une forme en S et donne des valeurs dans l'intervalle $[0, 1]$. Son gradient très important pour les valeurs proches de 0 fait que la moindre variation de x causera une variation significative de y , ce qui rend cette fonction très performante pour les entrées proches de 0 (Sharma, 2017). La fonction logistique, dont le graphe est donné dans la figure 14, s'écrit comme suit:

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

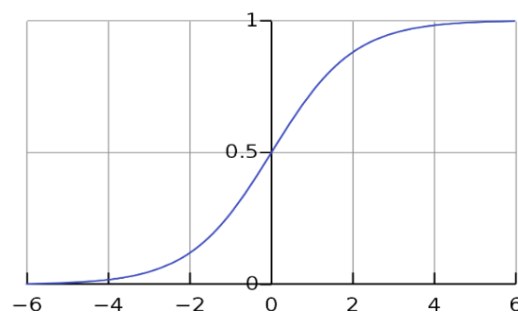


Figure 13. Tracé de la fonction logistique

3.5.2. Fonction tangente hyperbolique

Elle présente beaucoup de similitudes avec la fonction logistique. Les valeurs de sortie de cette fonction varient entre -1 et +1, ce qui se traduit par un gradient plus important (Sharma, 2017). L'expression de la fonction \tanh s'écrit comme suit :

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

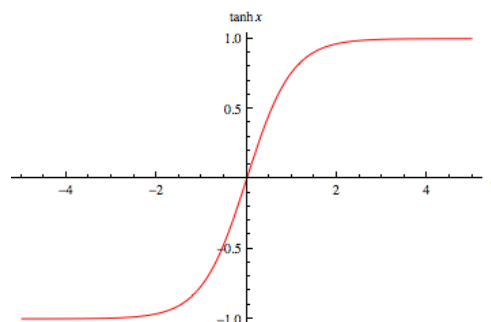


Figure 14. Tracé de la fonction Tanh

3.5.3. Fonction ReLU

C'est la fonction d'activation la plus récente et qui est très répandue actuellement. Elle est aussi bien utilisée dans les réseaux simples que dans les réseaux d'apprentissage profond. Il existe d'autres variantes à cette fonction qui sont LeakyReLU, PeLU, ... Elle est exprimée par (Sharma, 2017) :

$$F(x) = \max(0, x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{sinon} \end{cases}$$

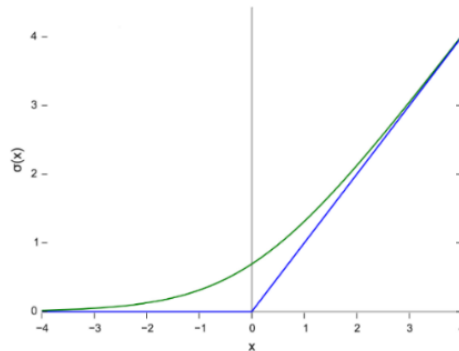


Figure 15. Tracé de la fonction ReLU

3.6. Types de réseaux de neurones

3.6.1. Réseaux non récurrents

Dans ce type de réseaux, l'information circule uniquement de la couche d'entrée vers la couche de sortie et le chemin inverse n'est donc pas possible comme illustré dans figure 17. D'un point de vue graphique, le réseau non récurrent est l'équivalent d'un réseau acyclique (Dreyfus, 2004).

Plusieurs types de réseaux s'inscrivent parmi les réseaux non bouclés, nous citerons principalement le FFNN (FeedForward Neural Network) ou réseau à propagation directe, qui est le plus communément utilisé au point où si le type du réseau n'est pas précisé, c'est qu'il s'agit d'un FFNN.

Dans cette architecture, Les neurones d'entrée correspondent aux variables indépendantes (x) qui sont nécessaires pour prévoir la variable (y). Le FFNN est donc utilisé lorsque la variable de sortie est expliquée par plusieurs variables d'entrée appelées variables explicatives. Les neurones dans la couche cachée sont la clé pour détecter la relation entre les variables d'entrée et la variable de sortie. Avec des fonctions de transfert non linéaires, les neurones cachés peuvent traiter des informations complexes reçues des neurones d'entrée, puis envoyer ces informations traitées à la couche de sortie pour un traitement supplémentaire avant de générer des prédictions.

Il convient d'ajouter que les réseaux non bouclés sont aussi appelés réseaux statiques du fait que le temps ne joue aucun rôle et que les fonctions sont exécutées instantanément.

Cas particulier : Le Perceptron Multicouches

Le perceptron multicouches est un type spécial parmi les réseaux de neurones non récurrents dont la caractéristique est que la fonction d'activation est un sigmoïde pour tous les neurones excepté les neurones d'entrée qui exécutent une fonction d'activation linéaire. Cette structure est très largement utilisée et on la trouve parfois avec la notation MLP (Multi-Layer Perceptron).

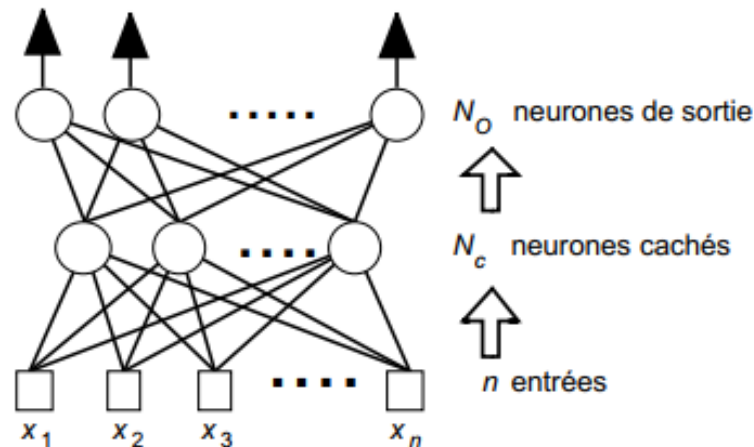


Figure 16. Réseau de neurones non bouclé(Dreyfus, 2004).

3.6.2. Réseaux récurrents

Dans la partie précédente, il a été expliqué que les réseaux de neurones non récurrents sont soumis à la contrainte d'absence de boucle et par conséquent l'information ne peut circuler que dans un seul sens.

Si nous détenons cette condition, et que nous laissons apparaître au moins un cycle à l'intérieur du réseau, nous obtenons alors un réseau récurrent où l'information peut revenir en arrière pour alimenter les neurones.

Contrairement aux réseaux *feedforward* qui appréhendent les vecteurs d'entrée et de sortie d'une manière indépendante, les RNN (Recurrent neural network) peuvent faire intégrer l'historique de l'ensemble des entrées précédentes pour chaque sortie, ce sont donc des systèmes dynamiques.

Les réseaux récurrents permettent de traiter des données séquentielles. En effet, à l'instant t ils calculent leur sortie en fonction de l'état aux temps précédents. Ainsi, ils font évoluer un état interne qui fait office de mémoire à court terme et qui permet de prendre en compte les dépendances temporelles que manifestent les sorties (Tristan, 2016).

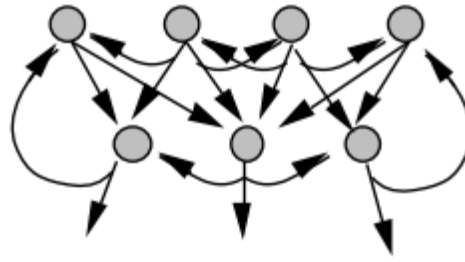


Figure 17. Réseau de neurones récurrent (Touzet, 1992).

La récurrence à l'intérieur du réseau peut être locale ou globale. Une récurrence locale est, soit une connexion entre deux neurones appartenant à la couche interne (ou entre deux couches internes), soit une boucle à l'intérieur d'un même neurone.

Une récurrence globale est, quant à elle, réalisée par une liaison partant de la couche de sortie vers la couche d'entrée (Danilo, 2001).

3.6.3. Réseau NARX

a. Définition

NARX (Non linéaire AutoRegressive avec variables eXogènes) est un réseau de neurones dynamique, constitué d'un Perceptron Multicouches (MLP) où la future valeur de la variable à prédire $y(t)$ est en fonction des n_y précédentes valeurs de sortie et des n_u précédentes valeurs d'une autre variable appelée variable exogène (Dreyfus, 2004). L'équation qui modélise le système s'écrit comme suit :

$$Y(t) = f(y(t-1), y(t-2), \dots, y(t-n_y), u(t-1), u(t-2), \dots, u(t-n_u))$$

Où

- u_t et y_t sont respectivement l'entrée et la sortie du modèle à l'instant t .
- n_u et n_y sont respectivement l'ordre du retard de l'entrée et ordre du retard de la sortie.

b. Architecture

Le modèle dynamique estimé est construit en ajustant le poids des connexions ω_{ji} et les biais b , via la fonction d'apprentissage. Le réseau NARX peut être représenté en adoptant l'une des deux architectures suivantes (Topus, 2010) :

- **Mode série parallèle**

Le réseau est alimenté par les vraies valeurs de sorties et le système est défini ainsi :

$$\hat{y}(k+1) = f(y(k), y(k-1), \dots, y(k-n_y), u(k+1), u(k), \dots, u(k-n_u))$$

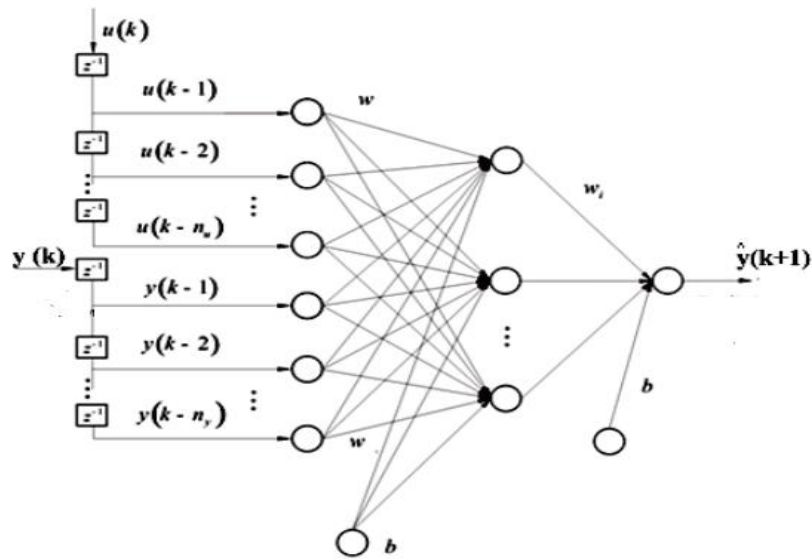


Figure 18. Architecture NARX, Séries-Parallèles (Topus, 2010).

- **Mode parallèle**

Dans ce cas, les sorties estimées sont ré-envoyées et incluses dans l'entrée de la structure qui est définie comme suit :

$$\hat{y}(k+1) = f(\hat{y}(k), \hat{y}(k-1), \dots, \hat{y}(k-n_y), u(k), u(k-1), \dots, u(k-n_u))$$

Alors que le réseau de type NARX-SP est utilisé en phase d'apprentissage, le réseau de type NARX -P est utilisé dans la phase de test.

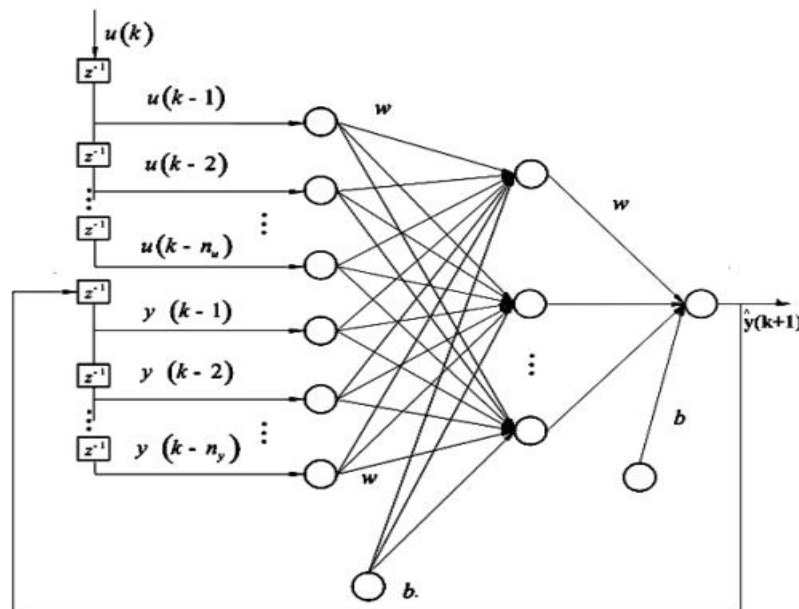


Figure 19. Architecture NARX en mode parallèle (Topus,2010).

3.7. L'apprentissage

3.7.1. Introduction

Un modèle de réseau de neurones tient sa puissance principalement, et avec une large mesure, à la manière dont les poids (connexions entre neurones) sont ajustés au fil des itérations. Ils sont ajustés par un algorithme d'apprentissage et qui approche la fonction cible par une précision suffisante.

L'apprentissage est défini comme l'ajustement des poids synaptique et des biais d'un réseau. Cette méthode est également connue sous le nom d'algorithme d'apprentissage. Le raisonnement derrière l'entraînement est que les poids sont mis à jour d'une manière itérative afin de minimiser l'écart entre les valeurs réels et les valeurs prédite cette fonction objectif est aussi appelé coût de l'apprentissage.

Lors de l'apprentissage, les données sont divisées en deux ensembles, un ensemble d'entraînement et un ensemble de tests. L'ensemble d'entraînement permet d'estimer les poids dans le modèle. Habituellement, l'ensemble de test est constitué de 10% à 30% de l'ensemble de données total. Il existe plusieurs algorithmes d'apprentissage utilisés pour estimer les valeurs des paramètres du réseau, leur utilisation dépend du type du réseau étudié(Cancellà, 2008).

a. Algorithme de rétro-propagation

L'algorithme de *Back propagation* (rétro propagation) est de loin l'algorithme d'apprentissage des réseaux de neurones le plus populaire. Il est souvent utilisé pour effectuer l'entraînement des perceptrons multicouche *feedforward* ainsi que les réseaux récurrents.

Le principe de cette méthode est de transmettre la responsabilité des écarts à chacun des éléments de traitement dans le réseau en propageant le gradient de la fonction d'activation à travers le réseau vers chaque couche cachée antérieure jusqu'à arriver à la première couche. La modification des poids est, par la suite, faite de manière à minimiser l'erreur quadratique moyenne entre la cible et la prédiction du réseau.

En réalité, le réseau traite une sortie. Cette sortie dépend des poids auxquels des valeurs aléatoires ont été attribuées initialement ; des valeurs d'entrée, des couches cachées et des fonctions de transferts.

La différence entre la prédiction du réseau et la sortie réelle (cible), appelée erreur, est utilisée pour mettre à jour les poids du réseau. Ce processus est répété jusqu'à ce que l'erreur totale soit minimale.

La fonction de l'erreur peut être définie comme suit :

$$E(\vec{\omega}) = \frac{1}{2} \sum_{i=1}^N (T_i - Y_i)^2$$

Où :

- T_i est la valeur cible et Y_i la sortie calculée pour la $i^{\text{ème}}$ itération.
- $E(\vec{\omega})$ est ainsi l'erreur totale des prédictions pour tout le processus d'entraînement.
- $\vec{\omega}$ est le vecteur poids du réseau.

Les erreurs de prédiction des données d'apprentissage individuelles sont à leurs tours égales à la somme des différences entre les valeurs de sortie produites par le réseau et les valeurs cibles.

Le but d'un algorithme d'apprentissage est de minimiser $E(\vec{\omega})$ pour un ensemble particulier de données destinées à l'apprentissage. Il y a plusieurs façons d'y parvenir, une méthode courante consiste à utiliser la méthode du gradient descendant.

L'algorithme d'apprentissage par la retro propagation selon le Gradient Descendant

Cette méthode peut être résumée en quatre étapes :

1. Initialisation aléatoire des poids du modèle.
2. Calcul du gradient de la fonction d'erreur $\nabla E(\vec{\omega})$:

$$\nabla E(\vec{\omega}) = \left(\frac{\partial E(\vec{\omega})}{\partial \omega_1}, \frac{\partial E(\vec{\omega})}{\partial \omega_2}, \dots, \frac{\partial E(\vec{\omega})}{\partial \omega_n} \right)$$

3. Ajustement des poids en se déplaçant dans la direction inverse du gradient de la fonction d'erreur. C'est à dire se déplacer sur une distance suffisamment petite pour diminuer l'erreur de façon maximale.
4. Répéter les étapes 2 et 3 jusqu'à ce que le gradient de la fonction d'erreur se rapproche du zéro. Si ce gradient est négatif, il faut augmenter les poids pour aller directement vers le minimum, et diminuer les poids s'il est positif.

3.7.2. Déroulement (Mazur, 2018)

Tout d'abord, un réseau de neurones est créé et initialisé ; c'est-à-dire que les poids sont fixés aléatoirement. Ensuite, et jusqu'à ce que la condition d'arrêt (par exemple l'erreur quadratique moyenne du réseau soit inférieure à un certain seuil d'erreur) soit satisfaite, tous les exemples d'apprentissage sont "appris" par le réseau.

Les entrées de chaque donnée d'apprentissage sont envoyées au réseau et traitées à partir de la couche d'entrée en passant par la couche cachée jusqu'à arriver à la couche de sortie.

De cette manière, le vecteur Y_i des valeurs de sortie produites par le réseau est obtenu. À l'étape suivante, les poids du réseau doivent être ajustés.

Fondamentalement, cela se produit lorsque la valeur de mise à jour du poids, $\Delta\omega$, doit "déplacer" le poids dans la direction de la descente la plus raide de la fonction d'erreur E par rapport au poids, ou de la dérivée partielle de E par rapport au poids :

$$\Delta\omega = -\eta \frac{\partial E}{\partial \omega_{ji}}$$

Où η est le taux d'apprentissage qui détermine la taille de l'étape utilisée pour "se déplacer" vers le minimum de E .

S'il est trop petit, alors la convergence vers un point optimal peut être faible, conduisant à une convergence lente du réseau de neurones artificiels. Inversement, s'il est trop grand, l'algorithme peut ne pas converger du tout. Habituellement : $\eta \in \mathfrak{R} ; 0 \leq \eta \leq 1$.

Notons que ω ne peut influencer le RNA que par la valeur nette, c'est-à-dire la somme pondérée des entrées pour l'unité j . Nous pouvons donc utiliser la règle de la chaîne pour écrire :

$$\frac{\partial E}{\partial \omega_{ji}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial \omega_{ji}}$$

Tel que net_j représente la sortie du neurone j .

Notons que :

$$\frac{\partial net_j}{\partial \omega_{ji}} = \frac{\partial (\sum_z \omega_{jz} x_{jz})}{\partial \omega_{ji}} = \frac{\partial (\omega_{j0} x_{j0} + \dots + \omega_{jz} x_{jz})}{\partial \omega_{ji}}$$

$$\frac{\partial net_j}{\partial \omega_{ji}} = (0 \cdot x_{j0} + \dots + 1 \cdot x_{ji} + \dots + 0 \cdot x_{jz})$$

$$\frac{\partial net_j}{\partial \omega_{ji}} = x_{ji}$$

Nous aurons donc :

$$\frac{\partial E}{\partial \omega_{ji}} = \frac{\partial E}{\partial net_j} x_{ji}$$

Les poids des nœuds cachés sont également mis à jour. Toutes les expressions sont les mêmes pour les neurones de sortie et ceux de la couche cachée, à l'exception de la dérivation de δ_j :

$$\delta_j = \frac{\partial E}{\partial net_j}$$

Dans le cas des neurones de sortie net_j , ces sorties ne peuvent influencer le réseau de neurones artificiels qu'à travers Y_j . Comme il a été fait ci-dessus avec ω_{ji} , nous pouvons enchaîner la règle pour écrire:

$$\frac{\partial E}{\partial net_j} = \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial net_j}$$

Notons que :

$$\begin{aligned} \frac{\partial E}{\partial Y_j} &= \left(\frac{1}{2} \sum_z (T_j - Y_j)^2 \right) dY_j = \frac{1}{2} ((T_0 - Y_0)^2 + \dots + (T_z - Y_z)^2) dY_j \\ &= \frac{1}{2} (0 + \dots + (-2T_j + 2Y_j) + \dots + 0) \\ &= -(T_j - Y_j) \end{aligned}$$

Aussi :

$$\frac{\partial Y_j}{\partial net_j} = \frac{\partial transfer(net_j)}{\partial net_j} = transfer'(net_j)$$

L'équation devient donc :

$$\frac{\partial E}{\partial net_j} = -(T_j - Y_j). transfer'(net_j)$$

En considérant les expressions précédentes, la mise à jour des poids $\Delta\omega$ pour les neurones de sortie peut être exprimée par :

$$\begin{aligned} \Delta\omega &= -\eta \frac{\partial E}{\partial \omega_{ji}} = -\eta \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial \omega_{ji}} = -\eta \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial net_j} x_{ji} \\ &= -\eta \left(-(T_j - Y_j). transfer'(net_j) \right) x_{ji} \\ &= \eta \left((T_j - Y_j). transfer'(net_j) \right) x_{ji} \end{aligned}$$

Pour les couches cachées cette dérivation doit prendre en considération la manière indirecte par laquelle ω_{ji} peut influencer les neurones de sortie et par conséquent l'erreur.

Il existe plusieurs autres algorithmes améliorés telles que le terme d'impulsion, la perte de poids,... Cependant, la rétroaction multicouche en combinaison avec l'algorithme d'apprentissage de descente de gradient stochastique est la technique des réseaux de neurones artificiels la plus utilisée dans la pratique. Une autre caractéristique importante de cet algorithme d'apprentissage est qu'il suppose une fonction d'erreur quadratique, donc il

suppose qu'il n'y a qu'un minimum. En pratique, la fonction d'erreur peut avoir - mis à part le minimum global - de multiples minima locaux. Il y a un risque que l'algorithme atteigne l'un des minima locaux et ne soit donc pas capable de réduire l'erreur au maximum possible en atteignant un minimum global (Cancelli,2008).

Pour conclure, dans la figure 19 une illustration qui résume le processus du rétro-propagation est présentée

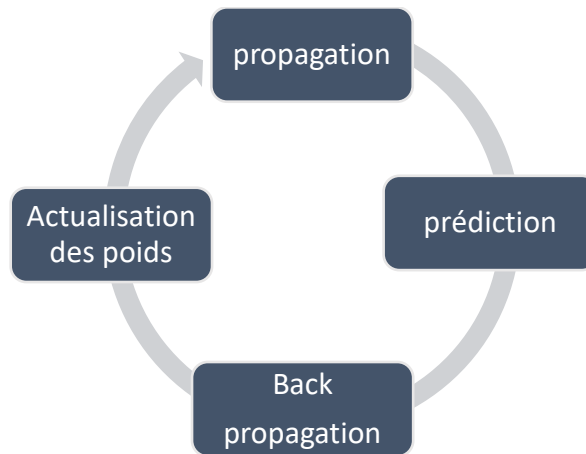


Figure 20. Cycle de l'apprentissage

Conclusion

Dans ce chapitre, une revue sur l'apprentissage automatique a permis de se familiariser avec les différents algorithmes existants ainsi que leurs cas d'utilisation. Il a été vérifié que les problèmes de prédiction s'inscrivent dans le domaine de l'apprentissage supervisé et que les réseaux de neurones artificiels figurent parmi les meilleurs outils de prédiction grâce à leurs qualités d'approximation et de parcimonie.

Dans une seconde partie, une explication de « l'apprentissage » a été détaillée, permettant ainsi une meilleure compréhension de ce mécanisme qui semble abstrait et mystérieux à première vue.

Maintenant que les réseaux de neurones récurrents ont été retenus, le choix du type de réseau le plus adéquat sera traité dans le prochain chapitre. Ce choix sera effectué en se basant sur les caractéristiques de cette problématique, la nature des données disponibles et après une revue sur le domaine d'application de chaque type. Un *benchmarking* avec des problématiques similaires sera aussi réalisé pour conforter ce choix.

Chapitre III : CHOIX DE LA METHODE ET SIMULATION

Introduction

Dans le chapitre précédent, et lorsque les types de réseaux de neurones ont été abordés, une différence cruciale entre les FFNN et les RNN a été révélée et qui va nous orienter pour le choix du réseau. Cette différence concerne la manière dont chaque type de réseau appréhende les données d'entrée. En effet, les réseaux récurrents ont la capacité de garder l'aspect séquentiel dans les données d'entrée, l'apprentissage est inscrit dans le temps et les prédictions futures dépendent des précédentes entrées. Contrairement aux réseaux à propagation directe qui traitent les entrées indépendamment du temps d'où le nom de réseaux statiques. Cela nous oriente clairement à choisir un réseau récurrent et éliminer les réseaux de propagation directe. Le tableau 2 qui montre le type de réseau de neurones et leur application est une référence pour valider notre réflexion.

Tableau 1 : Type de réseau et leur application (Heaton, 2015)

	Clust	Regr	Class	Predic	Robot	Vision	optim
FeedForward		***	***	**	**	**	
Hopfield			*			*	*
Boltzman Machine			*				**
DeepFeedForward		***	***	**	***	**	
CPPN					***	**	
ConvolutionalNetwork		*	***		***	***	
Elman Network		**	**	***			
Jordan Network		**	**	**	**		
Recurrent Network		**	**	***	**	*	

- Clust : problème de clustering
- Regis : problème de régression, la valeur de la sortie dépend des variables explicatives.
- Classif : problème de classification.
- Predic : problème de prédiction. Le réseau prédit des événements dans le temps.
- Robot : problème de robotique
- Vision : problème de reconnaissance des images
- Optim : problème d'optimisation.

Comme seconde référence, le tableau 3 montre quelques problèmes où les réseaux de neurones récurrents ont été utilisés. Toutes les prédictions présentes dans ce tableau sont basées sur l'évolution historique d'une même variable.

Tableau 2 : Travaux réalisés par des réseaux récurrents

Topic	Authors	Year
Predictive head tracking for virtual reality system	Saad, Caudell, and Wunsch	1999
Financial prediction using recurrent neural network	Giles, Lawrence, Tsoi	1997
Electric load forecasting	Costa, Pasero, Piglione, and Radasanu	1999
Natural water inflows forecasting	Coulibaly, Anctill, and Rousselle	1999
The use of NARX neural network to predict chaotic time series	Eugen Diaconescu	2008

Finalement, les travaux réalisés par VedatTopus indiquent qu'il est possible d'intégrer des facteurs explicatifs dans un réseau récurrent qui ne prenait –initialement- en compte que des valeurs historique de la variable prédite. Ce nouveau type d'architecture appelé NARX (*Non-linear Auto Regressivewithexogenous variable*) a été appliqué sur la prédiction de flux de la circulation. En effet, ce réseau semble parfaitement adéquat pour réaliser des prédictions basées sur des ventes historiques et autres paramètres ayant une influence sur la vente, nous allons donc alimenter le système par les ventes historiques et par des variables exogènes après les avoir sélectionnées (Topus, 2010).

1. Sélection des variables exogènes

1.1. Introduction

Dans cette partie, il s'agira d'identifier les paramètres qui ont une influence sur la vente des différents produits.

Si le consommateur algérien à un comportement assez stable pour la consommation du lait, qui demeure un produit incontournable et de première nécessité, la demande sur les jus et les laits-jus est susceptible d'être influencée par le changement des saisons ou par le mois de ramadhan qui représente une période particulière.

Après des entretiens avec les responsables de la force de vente de Tchén-lait, nous avons formulé des hypothèses sur les facteurs qui influent sur la vente, nous citons :

- Pour les laits : saison estivale, pénurie chez les concurrents (surtout le lait subventionné), mois de Ramadhan, promotion, périodes des examens, ...
- Les jus : saison estivale, Ramadan.
- Les laits-jus : vacances (saison estivale), rentrée scolaire, ...

Afin de vérifier les hypothèses précédemment annoncées, nous utilisons l'analyse de la variance. Il convient de mentionner que ce test est un outil quantitatif basé sur les données, il donc sera impossible de vérifier l'impact des paramètres dont nous ignorons les intervalles de temps où ils ont existé (i.e. vu que nous ne connaissons pas les périodes de pénuries chez les concurrents, il ne sera pas possible de mesurer l'influence de ce paramètre).

1.2. ANOVA

1.2.1. Principe de la méthode

ANOVA ou analyse de la variance est une technique statistique qui permet d'étudier l'influence d'un ou de plusieurs paramètres sur une variable quantitative. Cette méthode permet d'expliquer si la variabilité observée sur les données (réparties en classes ou groupes) est due au hasard de l'échantillonnage ou s'il existe réellement des facteurs externes influant sur les résultats. Ces facteurs sont des variables discrètes appelées « variables explicatives ».

La décomposition de la variance totale de l'échantillon en deux variances inter et intra-classes est à l'origine de cette appellation.

a. Conditions de l'analyse de la variance

- Indépendance des échantillons
- Normalité de la distribution

b. Données

- Il existe p groupes d'observations
- Pour chaque groupe k ($x_{k,1}, x_{k,2}, \dots, x_{k,n}$) l'espérance mathématique est noté μ_k
- $N = n_1 + n_2 + \dots + n_p$ est le nombre total des valeurs observées.

c. Test d'hypothèse

- $H_0 : \mu_1 = \mu_2 = \dots = \mu_p$ (le facteur f n'a aucune influence sur la population)
- H_1 : au moins une moyenne est différente des autres (le facteur f influe sur la population)

1.2.2. Déroulement

1. Calculer la moyenne empirique m_k de chaque classe :

$$m_k = \frac{x_{k,1} + x_{k,2} + \dots + x_{k,n_k}}{n_k}$$

2. Calculer la moyenne empirique totale M de l'échantillon :

$$M = \frac{n_1 m_1 + n_2 m_2 + \dots + n_p m_p}{N}$$

3. Calculer la variance empirique V_k de chaque classe :

$$V_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_{k,i} - m_k)^2$$

4. Calculer la variance intra-classes V_{intra} :

$$V_{intra} = \sum_{k=1}^p \frac{n_k}{N} V_k$$

5. Calculer la variance des moyennes ou variance inter-classes V_{inter} :

$$V_{inter} = \sum_{k=1}^p \frac{n_k}{N} (m_k - M)^2$$

6. Calculer F du test :

$$F = \frac{V_{inter}/(p-1)}{V_{intra}/(N-k)}$$

Extraire la statistique $F_{critique}$ du test, distribuée selon une loi de Fisher à $(p-1)$, $(N-k)$ degrés de liberté et à un seuil de signification α .

Règle de décision

Si $F < F_{critique} \rightarrow$ accepter H_0 (au risque α)
sinon rejeter H_0

1.2.3. Analyse de la variance appliquée sur VIVA

Dans cette partie, l'analyse de la variance est appliquée à l'ensemble des produits de Tchén-lait. Cependant, seule la démarche pour le produit VIVA sous l'influence du facteur « Ramadhan » est présentée en détail. La même démarche est appliquée pour les autres facteurs sur le reste des produits.

a. Préparation des données

Pour vérifier l'impact d'un paramètre sur un produit quelconque, nous isolons les données enregistrées sous l'influence de ce paramètre avec deux autres échantillons « neutres » ou « témoins » (enregistrés en l'absence du paramètre en question). Les trois groupes seront soumis à une analyse de la variance.

Dans cet objectif, les paramètres suivants ont été retenus vu que nous disposons des données nécessaires pour l'analyse.

Saison estivale :

L'été peut avoir un impact sur la consommation des produits de Candia car le consommateur risque de changer ses habitudes durant cette période de vacances. Aussi, la population de la ville risque de changer (départ de certains et arrivée d'autres).

Dans le cadre de ce test, la saison estivale est la période allant du 25 juin 2017 au 31 Août 2017 ; cette période comprend 51 observations. La logique derrière le choix de cet intervalle de temps est qu'il commence après le mois de Ramadhan jusqu'à la dernière

semaine du mois d'Août qui marque, en général, la fin des vacances pour les écoliers, les professionnels et les estivants.

Le mois de Ramadhan :

La consommation de certains produits alimentaires est généralement impactée durant ce mois particulier. 23 observations sont disponibles pour le mois de ramadhan de l'année 2017 et qui a coïncidé la période allant du 26 Mai 2017 au 24 Juin 2017.

b. Vérification de la normalité

Pour l'argument suivant, une approche quantitative ainsi qu'une validation graphique sera retenue afin de vérifier la normalité des échantillons : ANOVA est un test robuste (Schmider, 2010). Les tests robustes sont des tests avec de bonnes performances pour les données tirées d'une distribution qui n'est pas parfaitement normale.

- **Test de Shapiro et Wilk**

Ce test de normalité est spécifique pour des échantillons plus ou moins réduit comme est le cas de notre échantillon.

La démarche à suivre est :

1) Ordonner l'échantillon :

Ordonner l'échantillon de manière décroissante tel que

$$X_1 > X_2 > \dots > X_n$$

2) Calculer la moyenne de l'échantillon :

$$\bar{X} = \frac{1}{n} \left(\sum_{i=1}^n X_i \right)$$

3) Calculer S^2 :

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

4) Calculer b^2 :

$$b^2 = \sum_{j=1}^{\lfloor n/2 \rfloor + 1} a_j * d_j$$

Où :

$$a_j = X_j - X_{n-j+1}$$

Et d_j des valeurs propre au test à lire à partir du tableau de Shapiro et Wilk

5) Calculer W :

$$W = \frac{b^2}{S^2}$$

6) Comparer à W_{critique} .

Lire sur la table de Shapiro et Wilk la valeur de W_{critique} qui correspond à un échantillon de taille n .

7) Valider ou rejeter la normalité.

Si $W > W_{\text{critique}}$ de la table, la normalité est acceptée.

Si $W < W_{\text{critique}}$ de la table, la normalité est rejetée.

Dans notre cas :

$n=24$; $W_{\text{critique}}= 0.884$ avec un risque $\alpha =0.05$.

Pour les trois échantillons nous avons :

$W_1=0,96119275$; $W_2=0.89389198$; $W_3= 0,97387346$.

Les trois échantillons suivent une distribution normale selon le test de Shapiro et Wilk avec un risque de 5%.

- **Histogramme de la distribution**

La figure 22 représente la distribution des ventes du produit Viva, cette distribution prend la forme d'une distribution normale

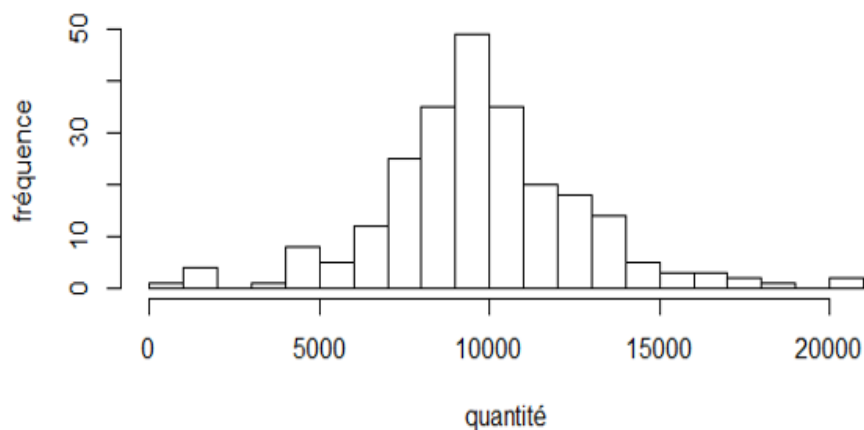


Figure 21. Distribution de l'échantillon VIVA

- **Diagramme Quantile-Quantile**

Le Diagramme Q-Q est un outil statistique graphique qui permet de vérifier l'ajustement d'une distribution par rapport à un modèle théorique (dans notre cas, nous vérifions la distribution de l'échantillon VIVA par rapport à la loi normale). Si la distribution des données suit le modèle théorique, les points doivent se positionner suivant la première diagonale.

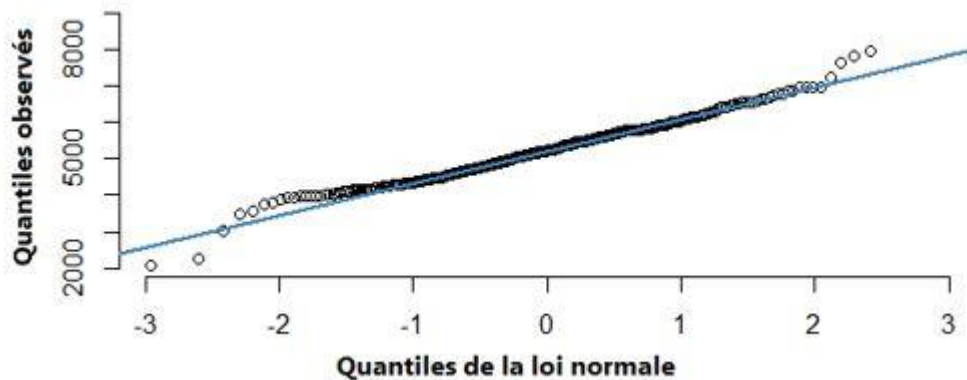


Figure 22. Diagramme Q-Q de l'échantillon VIVA

D'après l'histogramme de la distribution et le diagramme Q-Q où les points sont quasiment superposés sur la diagonale, la normalité de l'échantillon est donc admise pour la suite de cette application.

c. Application

Ce test est réalisé à l'aide du logiciel Microsoft Excel. Les données de la première colonne ont été collectées pendant le mois de Ramadhan alors que les données des deux autres colonnes ont été récoltées dans deux périodes différentes et hors mois de Ramadhan (mois de mars et mois d'octobre respectivement). Le risque α sera maintenu à 0,05.

- H_0 : les moyennes sont égales et le facteur « Ramadhan » n'influe pas sur la vente.
- H_1 : le facteur « Ramadhan » influe sur les ventes.

Tableau 3 : Total des ventes Viva au mois de Ramadhan et hors Ramadhan

Ramadhan	H.Ramadhan	H.Ramadhan
8556	8652	13248
9047	9132	7968
11616	9648	11472
11400	7440	7428
10744	8040	10272
8844	10140	9504
7212	8028	12408
9564	7272	13008
8688	4572	9606
6432	13476	12252
10224	7764	10688
10116	10476	11136
10188	9636	12204
9276	7464	5046
12696	10044	9506
15084	552	9384
4836	8148	10716
6948	9660	8381
8760	9288	8340
8292	8664	9936
12228	10692	11728
5868	9504	9876
9108	9156	8784
1080	7956	7404

d. Résultat de l'analyse

- **Rapport détaillé**

Tableau 4: Rapport détaillé

Groupes	Nombre	Somme	Moyenne	Variance
colonne 1 (R)	24	216768	9032	8082427,83
colonne 2 (HR)	24	205344	8556	5596678,96
colonne 3 (HR)	24	240324	10013,5	4033327,04

Lecture : Dans les tableaux 4 et 5, nous remarquons que les moyennes des trois groupes (9032, 8586 et 10013,5) sont très proches, avec le mois de Ramadhan qui affiche un résultat au milieu des deux autres.

- **Analyse de la variance**

Tableau 5:Résultat du test ANOVA sur le produit Viva sous l'influence du facteur «Ramadhan ».

Source de variation	Somme des carrées	Degré de liberté	F ₀	probabilité	f-critique
Entre groupes	26513976	2	2,24	0.11	3,13
A l'intérieur du groupe	407385978	69	5904144		
Total	433899774	71			

Lecture : Dans le tableau 6 nous avons la valeur $F_0=2,24$ et qui est inférieure à $F_{critique}(3,13)$

- **Conclusion**

$F < F_{critique}$, l'hypothèse H_0 est accepté.

Nous pouvons dire que VIVA 11 n'est pas influencé par le mois de Ramadhan.

e. Résultat de l'analyse sur tous les produits

La même démarche qui a été appliquée sur les ventes du produit VIVA pendant le mois de Ramadhan est reproduite avec le reste des produits et sous l'influence d'un deuxième facteur qui est la saison estivale.

Tableau 6: Résultats de l'analyse de la variance sur les produits Candia à $\alpha = 0,05$

Produit	« Ramadhan »		« été »	
	F	F ₀	F	F ₀
Viva	2,45	3,13	21,3	3,05
DE 1L	0,67	3,13	35,5	3,05
Silhouette	1,22	3,13	0,2	3,05
Entier	5	3,13	16,9	3,05
Twist P-A20cl	5,25	3,13	1,22	3,15
Twist O-F-B 20cl	2,13	3,13	1,96	3,15
Twist O-M 20cl	1,89	3,13	2,42	3,15
Twist O-A 20cl	1,68	3,13	2,33	3,15
Choco 20cl	1,92	3,13	2,08	3,05
Cocktail 20cl	1,85	3,13	2,43	3,05
Orange 20cl	1,45	3,13	1,67	3,05
Choco 1L	2,71	3,13	12,33	3,05
Twist O-F-B 1L	2,64	3,13	2,02	3,05
Orange 1l	18,47	3,13	27,2	3,05
Nectar de Grenade	26,9	3,13	2,85	3,05
Citronnade	51	3,13	38,45	3,05
Cocktail 1l	2,2	3,13	28,6	3,05
DE 50cl	DM	/	1,69	3,05
Twist O-A 1L	DM	/	DM	/
Twist P-A 1L	DM	/	DM	/
Twist O-M 1L	DM	/	DM	/
Choco 125ml	DM	/	1,78	3,05
Candy Caramel 125ml	DM	/	DM	/
Fraise 20cl	DM	/	DM	/
Fraise 125 ml	DM	/	DM	/
Crème de cuisson	DM	/	DM	/

Les tableaux 8 et 9 présentent respectivement les produits influencés par les facteurs « Ramadhan » et « été » ainsi que le taux de variation par rapport à la moyenne annuelle.

Tableau 7: Liste des produits impactés par le facteur « Ramadhan »

Produit	Hausse (H) / Baisse (B)	Taux de variation par rapport à la moyenne annuelle
Entier	B	-10%
Twist P-A 20cl	B	-20%
Orange 1l	H	+70%
Nectar de grenade	H	+120%
Citronnade	H	+160%

Tableau 8 : Liste des produits impactés par le facteur «été»

Produit	Hausse (H) / Baisse (B)	Taux de variation par rapport à la moyenne annuelle
Viva	H	+ 25%
DE 1L	H	+ 28%
Entier	H	+ 27%
Choco 1l	H	+ 33%
Orange 1l	H	+ 30%
Citronnade	H	+ 60%
Cocktail 1l	H	+ 60%

1.3. Synthèse

L'analyse de la variance a permis de mettre en évidence l'influence de deux facteurs très importants sur l'activité de l'entreprise Tchou-lait à savoir le mois de Ramadhan et la saison estivale.

Les résultats de cette analyse permettront au concepteur du system basé sur les RNA d'intégrer les facteurs exogènes identifiés de les inclure comme variables d'entrées.

Dans le cas de cette étude, le facteur F_i qui aura un impact sur le produit P_j sera intégré comme variable discrète exogène dans la construction du réseau de neurones.

2. Simulation

Le logiciel de calcul numérique et de visualisation graphique Matlab a été retenu pour la simulation. Ce choix est motivé par le fait qu'il dispose d'une bibliothèque de réseaux de neurones permettant de tester différents problèmes de clustering, de classification ou de régression.

Matlab qui est un langage de programmation très ergonomique dispose aussi d'une plateforme conviviale et d'un module d'aide très efficace qui permet d'orienter l'utilisateur au cours de son travail. En réalité, il suffit d'introduire le nom de la commande précédée par aide (aide <commande>) pour que le logiciel explique en détail comment utiliser cette commande.

Les étapes de la simulation sont représentées dans la figure 24 :

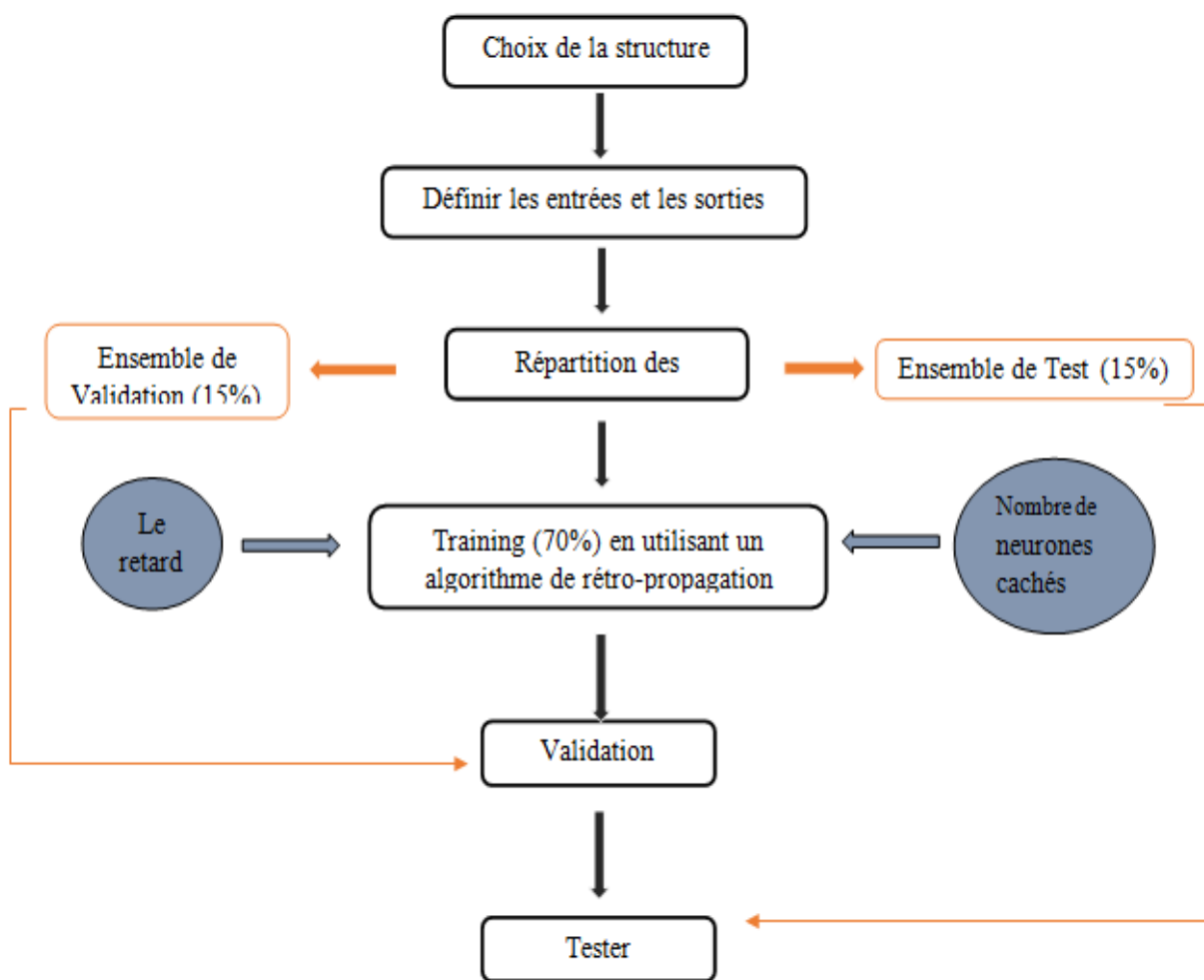


Figure 23 : Etapes de la simulation

2.1. Les étapes de construction du modèle

La construction du réseau de neurones qui prédit le comportement d'achat des points de vente comporte les étapes suivantes :

2.1.1. Préparation des données

Les traitements effectués à l'aide du tableau croisé dynamique sur Excel nous ont permis d'extraire le total des ventes journalières par produit. VIVA est le produit retenu pour la simulation et la série qui représente la vente totale journalière est définie comme la variable de sortie $y(t)$. 320 observations ont été récoltées et qui représentent la vente durant la période allant du 02/01/2017 jusqu'au 23/03/2018.

Puisque dans le cas présent la vente dépend principalement de la période de l'année, des variables exogènes seront introduites afin de cadrer l'évolution dans le temps. Pour ce faire, deux variables d'entrée seront utilisées : les 5 jours de la semaine (samedi=1, dimanche=2, lundi=3, mardi=4 et mercredi=5) et aussi la semaine de l'année (1:52). Ces deux variables seront très utiles pour l'apprentissage du système. En effet, lorsque le système dispose de l'information sur le jour, la période, la saison, ... cela facilitera la détection des informations cachées et les périodes qui causent une augmentation ou une baisse dans la sortie.

D'autres variables pourront être introduites, il s'agit des facteurs qui influent sur la vente. Les résultats de l'étude statistique ANOVA nous ont permis de confirmer l'influence de quelques paramètres comme la saison d'été et le mois de Ramadhan sur la vente des produits. Par conséquent un vecteur logique –présenté dans la figure 25- sera introduit pour marquer la présence du facteur.

N°jour	N°sem	F	vente
2	1	0	3012
3	1	0	5052
4	1	0	4848
5	1	0	5760
1	1	0	3936
2	2	0	4572
3	2	0	4080

Figure 24. Les 7 premières lignes de la matrice qui alimente le réseau pour la simulation

2.1.2. Sélection du réseau

Après l'ouverture de Matlab, la commande `nnstart` permet d'accéder à la bibliothèque Réseaux de neurones. Plusieurs assistants sont disponibles pour différents problèmes comme présenté dans la figure 26. Dans notre cas « time series app » est sélectionné.

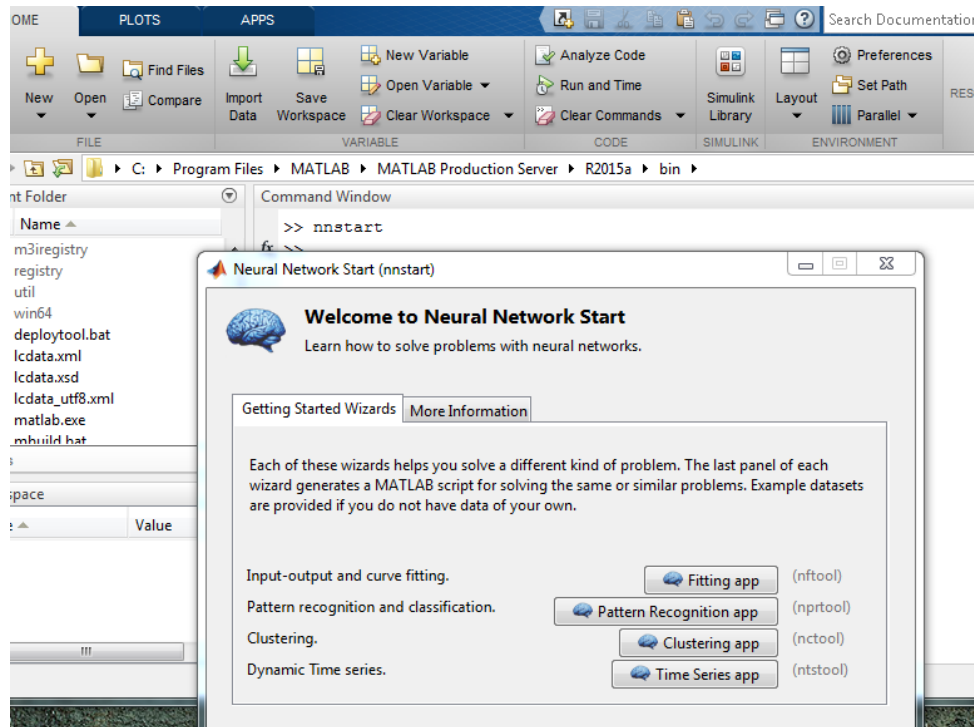


Figure 25. Illustration de l'interface pour les RNA sur Matlab

2.1.3. Sélection des entrées et sorties

- Données d'entrée : comme mentionné précédemment, trois vecteurs seront introduits en *input* à savoir : vecteur « jour », vecteur « semaine » et vecteur « facteur ».
- Donnée de sortie : le vecteur qui représente les ventes de VIVA, qui est la variable à prédire, sera introduit comme *output*. La récurrence présente dans l'architecture NARX permet à cette variable d'alimenter le réseau (donc l'*output* est considéré, lui aussi, comme une variable d'entrée).

2.1.4. Séparation des données

Maintenant que les données sont importées et prêtes à l'utilisation, il s'agit de procéder à une séparation de ces données.

La séparation peut aboutir à deux ou trois ensembles selon la nature du problème ou selon l'algorithme responsable de l'apprentissage. Les données sont réparties en un ensemble d'entraînement et un ensemble de test, un ensemble de validation s'ajoute lors de l'application de certains outils comme dans les réseaux de neurones.

Les données dont nous disposons seront réparties en 70% pour l'entraînement, 15% pour la validation et 15% pour le test comme le prescrit l'usage, explicité dans la figure 27.

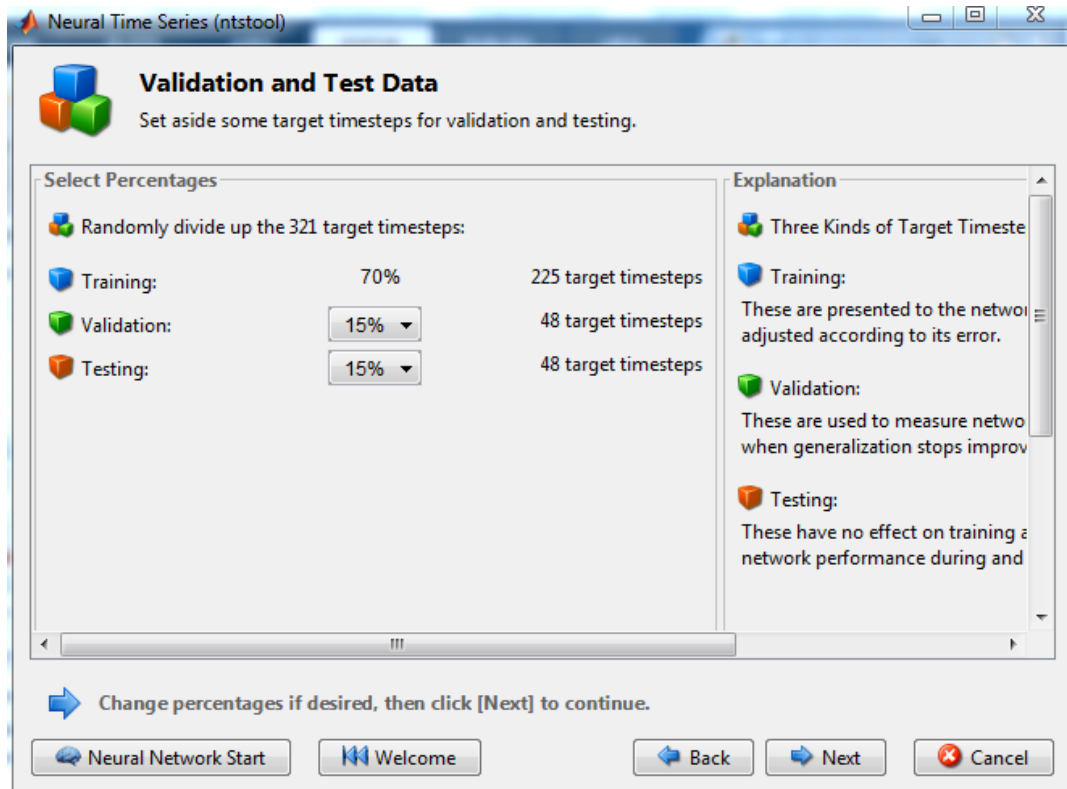


Figure 26 :Répartition des données

2.1.5. Ensemble d’entraînement

L’ensemble d’entraînement est utilisé pour ajuster les poids sur le réseau de neurones (phase d’apprentissage). Dans cette partie, les données d’entrée sont présentées au modèle au même titre que les données de sortie.

2.1.6. Ensemble de validation

Cet ensemble de données est utilisé pour empêcher un problème très répandu qui est le sur-ajustement ou (*overfitting*). Il ne s’agit pas ici d’ajuster les poids du réseau mais de vérifier simplement que toute augmentation de précision par rapport à l’ensemble de données d’entraînement augmente réellement la précision sur un ensemble de données qui n’a pas été montré au réseau auparavant (ensemble de test) et augmente donc la performance du modèle. Si la précision sur l’ensemble de données d’entraînement augmente, mais que la précision sur l’ensemble de données de validation reste la même ou diminue, alors il s’agit d’un sur-ajustement.

Définition du sur-apprentissage (*overfitting*) et le sous-apprentissage (*underfitting*)

Le sur-apprentissage (ou *overfitting*) est la situation dans laquelle le réseau prend le bruit et les fluctuations aberrantes en tant que concept. Le problème est que ces aberrations ne s’appliquent pas aux futures données et auront un impact négatif sur la capacité du modèle à généraliser. L’architecture du réseau est généralement la cause de ce problème qui se manifeste par des performances élevées en apprentissage mais fournit de faibles prédictions lors du test (on dit que le modèle est trop flexible) (Yuu, 2007).

En revanche, un réseau de neurones avec trop peu de paramètres (sous-apprentissage ou *underfitting*) ne sera pas assez complexe pour correspondre à la complexité de la fonction de régression (inconnue), donc il ne sera pas en mesure d'apprendre les données d'entraînement. Les structures en sous-apprentissage sont caractérisées par une faible performance lors de l'apprentissage comme lors du test (Yuu, 2007).

2.1.7. Ensemble de test

Cet ensemble de données est utilisé uniquement pour tester la solution finale afin de confirmer la capacité prédictive réelle du réseau. Evidemment, dans cette étape le système prend seulement les données d'entrée, réalise des prédictions et compare avec les valeurs réelles.

2.2. Exécution

« Il n'existe, à l'heure actuelle, aucun résultat théorique permettant de prévoir le nombre de neurones cachés nécessaires pour obtenir une performance spécifiée, compte tenu des données disponibles. Il faut donc nécessairement mettre en œuvre une procédure numérique de conception de modèle » (Dreyfus, 2004).

A partir de ce résultat, la conception du réseau optimal passe par une série de simulations et d'ajustement avant d'optimiser les paramètres qui donnent les meilleurs résultats.

Pour obtenir un résultat optimal, plusieurs simulations sont effectuées en modifiant après chaque simulation les valeurs de D (retard) ou NNC (nombre de neurones cachés) ou les deux en même temps.

Tableau 9: Résultat simulation NARX

NNC	D	Train	Test	Overfitting/underfitting
2	1	0.5	0.54	underfitting
3	2	0.83	0.68	overfitting
4	2	0.82	0.83	
4	1	0.51	0.48	underfitting
4	5	0.88	0.67	overfitting
4	10	0.90	0.61	overfitting
5	2	0.83	0.84	
5	3	0.87	0.76	
5	8	0.90	0.65	overfitting
6	2	0.84	0.8	
6	4	0.89	0.51	overfitting
8	2	0.86	0.73	
8	6	0.95	0.61	overfitting
10	2	0.89	0.60	overfitting
10	5	0.92	0.73	overfitting
16	2	0.87	0.66	overfitting
22	2	0.85	0.67	overfitting
22	8	0.97	0.6	overfitting

2.3. Discussion

- Lorsque le nombre de neurones cachés est faible ($NNC < 4$), les modèles fournissent des résultats relativement faibles (test = 0.54 et 0.68 pour $NNC = 2$ et 3 respectivement). Nous pouvons conclure qu'un nombre réduit de neurones cachés cause un sous-ajustement du modèle.

- Les grandes valeurs de retard D provoquent un sur-ajustement du modèle. En effet, lors de l'entraînement avec $D = (10, 8, 6)$ le modèle a affiché un résultat élevé dans l'entraînement mais cela n'a pas été reproduit lors du test. Donc D ne doit pas prendre de grandes valeurs afin d'éviter la construction d'un système figé incapable de généraliser sur les données futures.

- Tout comme pour les valeurs élevées de D , lorsque le NNC est grand, le système souffre d'un sur-ajustement et affiche une performance moyenne lors du test; cette performance est encore plus faible quand le retard est grand.

- Les meilleurs résultats ont été obtenus pour les petites valeurs de D et le $NNC < 9$. La structure qui a donné la meilleure performance lors de ce test correspond à $NNC = 5$ et $D = 2$, avec une corrélation de 83% lors de l'entraînement et 84% lors du test. Donc le modèle ne souffre pas d'un sur-ajustement.

La figure 28 illustre la structure du modèle avec les paramètres qui offrent la meilleure performance.

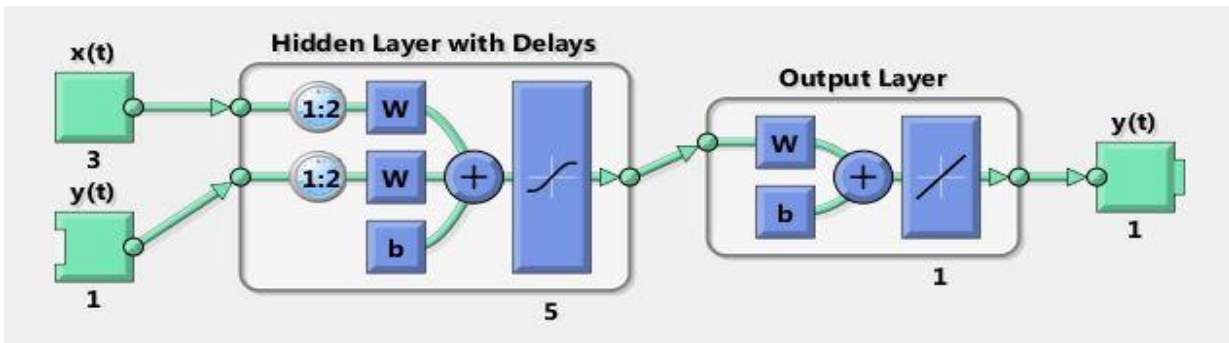


Figure 27. Structure du modèle retenu

2.4. Analyse de la performance et validation

2.4.1. Diagramme de régression

Ce diagramme reflète la performance du modèle, il représente à quel point les valeurs prédites sont proches des valeurs réelles, plus R est proche de 1 plus le modèle réussit à approcher les valeurs réelles et plus il est fiable. Pour notre modèle ses valeurs son représentés sur la figure 29.

R : corrélation entre les valeurs prédites et les valeurs réelles.

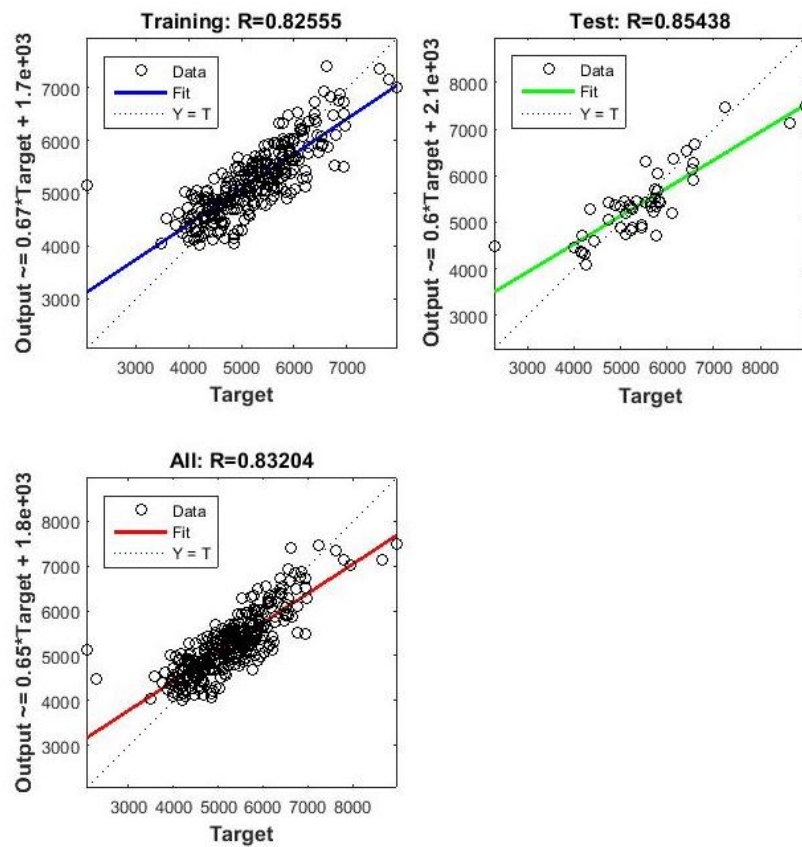


Figure 28. Corrélation entre les valeurs de prédiction et les valeurs réelles (NARX)

2.4.2. Analyse de l'erreur

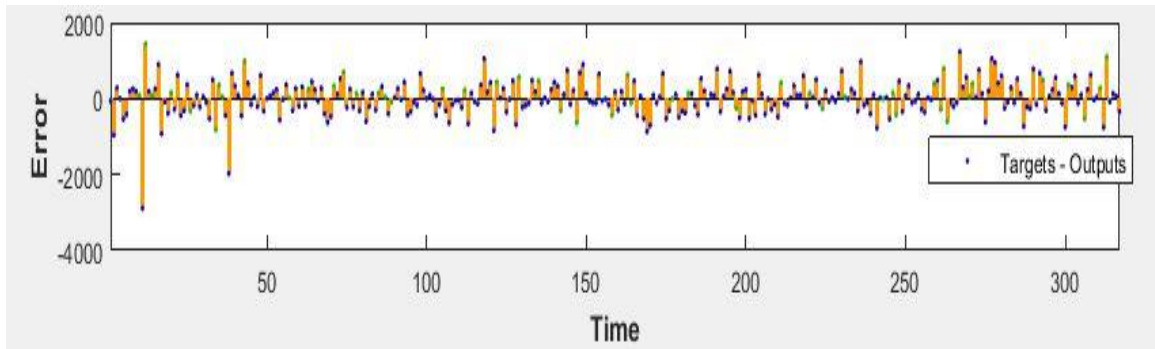


Figure 29. Diagramme de l'erreur

Ce diagramme montre l'écart entre la valeur prédite par le modèle et la valeur réelle de la sortie. Nous constatons que chaque fois les valeurs réelles atteignent un pic d'une manière subite, l'erreur devient importante ce qui veut dire que notre modèle n'est pas encore suffisamment performant pour anticiper ce genre d'événement.

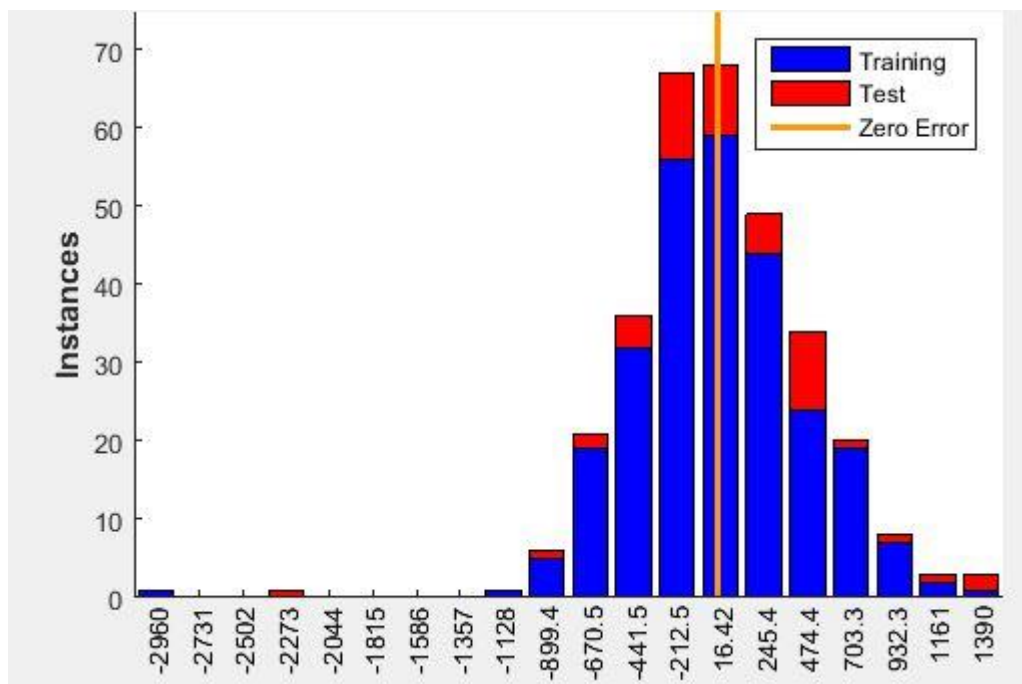


Figure 30. Histogramme illustrant le cumul de l'erreur

Même si les valeurs prédites ne correspondent pas exactement aux valeurs réelles, et que le modèle peut afficher une erreur supérieure à 500, l'histogramme dans la figure 31 montre que l'erreur est généralement proche de 0 ce qui est un point positif pour le modèle. L'information qu'il faut retenir c'est que globalement le modèle affiche de bonnes performances malgré quelques erreurs importantes. Ces erreurs arrivent lors d'un changement brusque dans la demande que le système n'est pas encore prêt à les anticiper.

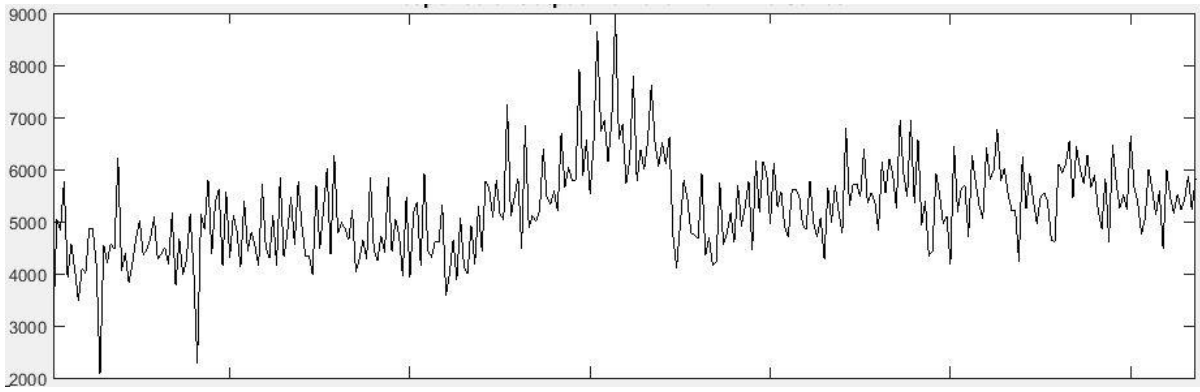


Figure 31. Diagramme des valeurs réelles

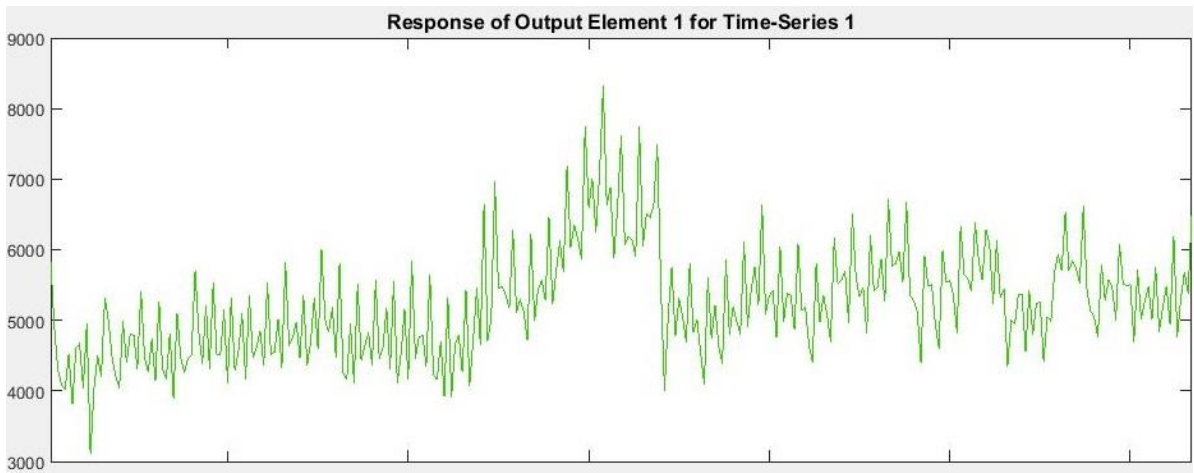


Figure 32. Diagramme des valeurs prédites.

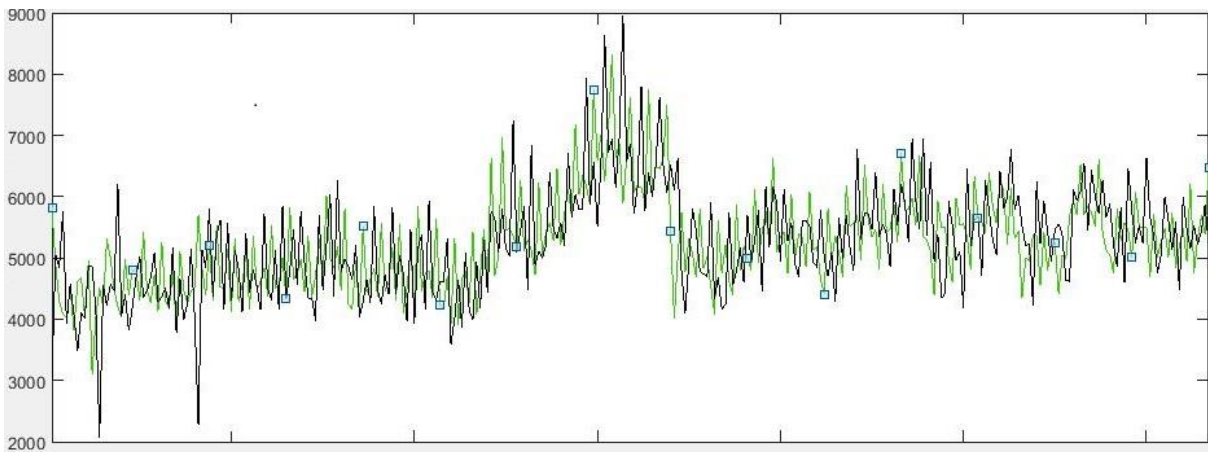


Figure 33. Diagramme des valeurs réelles et prédites

2.5. Amélioration de la performance

Dans cette seconde étape, un ensemble plus large de données a été considéré pour la simulation. Cet ensemble représente le total des ventes depuis le 01/01/2016 jusqu'au 30/03/2018. Nous rappelons que pour la première simulation, les données allaient du 01/01/2017 au 30/03/2018. Donc cet ensemble n'est qu'un prolongement dans le passé des données précédemment utilisées.

Le tableau 11 présente une comparaison entre les résultats obtenus avant et après l'intégration des nouvelles données.

Tableau 10: Comparaison de la performance du réseau

Avec les nouvelles données	Sans les nouvelles données
<p style="text-align: center;">NNC = 5 D = 2 Avec une corrélation :</p> <p>R entraînement = 0.82 R test = 0.82</p>	<p style="text-align: center;">NNC = 5 D = 2 Avec une corrélation :</p> <p>R entraînement = 0.83 R test = 0.84</p>

Discussion :

Le prolongement des données injectées dans le système n'a pas aidé ce dernier à fournir une meilleure performance, au contraire la corrélation a légèrement baissé. Connaissant le comportement des outils de l'apprentissage automatique qui s'améliorent au fur et à mesure avec la croissance des données disponibles, cela apparaît non logique à première vue.

En effet, le premier ensemble de données qui représente les ventes sur 15 mois a permis au système de réaliser une corrélation de 84%. Toutefois, l'extension des observations sur 27 mois à aboutit à une corrélation R test=82%.

Ce comportement peut être expliqué par le fait que la série chronologique qui représente l'évolution des ventes (de VIVA comme des autres produits d'ailleurs) peut être définie comme une série « Chaotique » ou série « quelconque ». Ce type de série ne présente pas une forme qui se reproduit après une période déterminée (absence de saisonnalité et absence de périodicité). Un caractère qui rend la prédiction de leur évolution une mission plus compliquée. Les deux figures 35 et 36 illustrent cette explication.

Conclusion : les nouvelles données n'ont pas aidé le modèle à fournir de meilleurs résultats. Donc une corrélation de 84% peut être la meilleure performance de ce modèle.

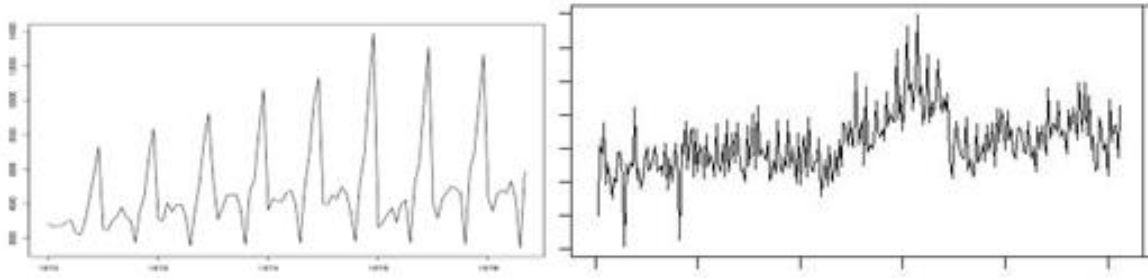


Figure 34: Comparaison de série périodique (gauche) et série quelconque (droite)

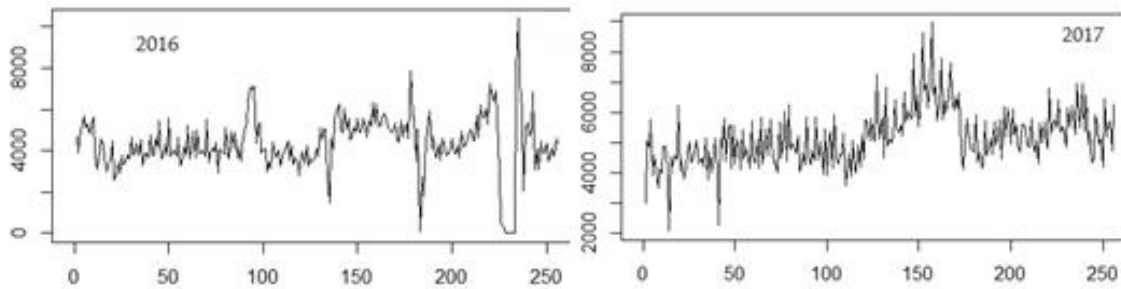


Figure 35 : Comparaison des ventes Viva en 2016 et 2017

2.6. Importance des variables exogènes

Afin de montrer l'importance des variables exogènes et leur influence sur la performance du réseau, une simulation sur un réseau de type NAR (Non linear Auto-Regressive) est appliquée. Ce réseau possède les mêmes caractéristiques que NARX excepté l'intégration des variables exogènes. Ainsi, le système n'aura que les ventes historiques pour prédire les futures ventes.

Tout comme pour la précédente simulation, un échantillon de 320 observations a été récolté et qui représente les ventes durant la période du 02/01/2017 au 23/03/2018. Après plusieurs itérations pour la sélection des paramètres N et D, le couple (N=8, D=3) a présenté le meilleur taux de corrélation R qui est en effet le principal critère de sélection.

Synthèse : les résultats de ce test qui sont nettement inférieurs à ceux produits par NARX montrent en effet l'importance des variables exogènes dans la réalisation des prédictions fiable. Ces variables présentent des informations supplémentaires considérablement utiles pour que le réseau puisse mieux interpréter les variations et maîtriser les causes des fluctuations.

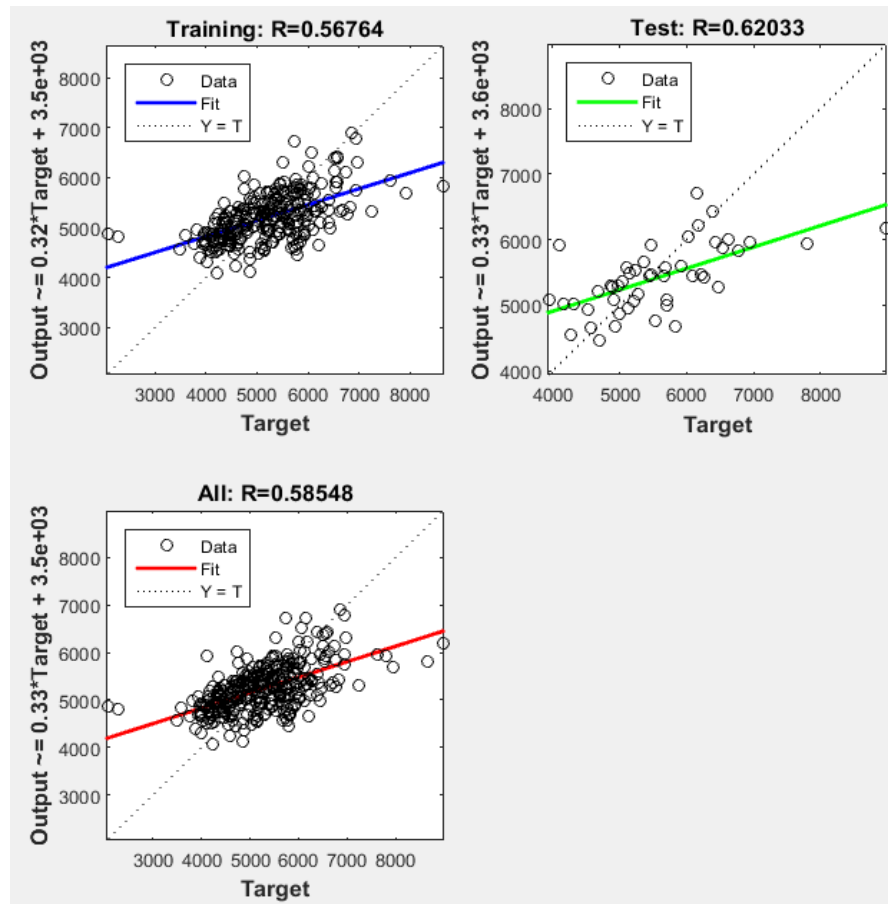


Figure 36 : Corrélation entre les valeurs de prédiction et les valeurs réelles (NAR)

3. Recommandations

Avant d'aborder les recommandations, il est important de présenter une vision globale d'un système de prédiction. Pour cela nous faisons appel aux travaux de Mentzer et Moon qui présentent un schéma en trois couches incluant cinq étapes successives et qui sont (Mentzer, 2004)

- a) Analyse de l'environnement
- b) Définition de l'organisation
- c) Choix de la technique
- d) Implémentation du système informatique
- e) Mesure de la performance

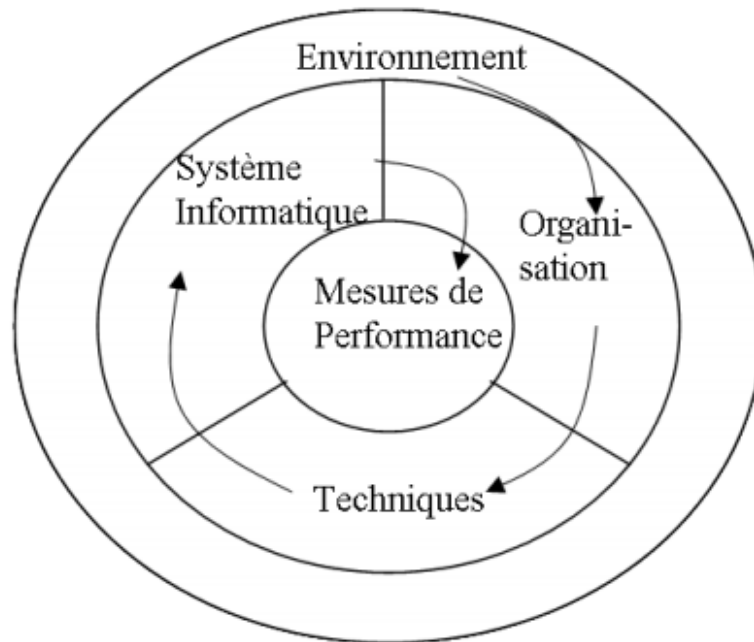


Figure 37. Processus de prévision de vente (Mentzer, 2004)

Dans la partie qui suit, nous développons brièvement les étapes de conception du système et les recommandations à prendre en considération lors de chaque étape

3.1. Analyse de l'environnement

La couronne externe représente l'environnement dans lequel évolue l'entreprise. Il est divisé en facteurs sur lesquels l'entreprise peut agir et d'autres sur lesquels il n'est pas possible d'agir, souvent ces deux éléments peuvent être assimilés à l'environnement externe et interne de l'entreprise.

Dans le présent travail, Cette étape a été réalisée lors du diagnostic.

3.2. Organisation

En vue d'adapter l'organisation de l'entreprise à accueillir le nouveau système de prédiction les points suivants sont proposés:

3.2.1. Prédiction par circuit

D'après l'étude du système de distribution de Tchik-lait et l'étude comparative des deux modes de prédictions résumées dans le tableau 12, nous recommandons d'effectuer la prévision par circuit. Les avantages de cette proposition sont explicités ci après.

Tableau 11: Comparaison de la prédiction à l'échelle macro et micro

Prédiction par point de vente (échelle micro)	Prédiction par ville (échelle macro)
<p style="text-align: center;"><u>Les avantages :</u></p> <ul style="list-style-type: none"> - Permet de Connaitre les besoins de chaque client - Eviter les ruptures de stocks au niveau des points de vente - Optimiser le transport entre le dépositaire et le point de vente - Personnaliser le service - Fidéliser les clients <p style="text-align: center;"><u>Les obstacles et les inconvénients :</u></p> <ul style="list-style-type: none"> - Manque de données sur le comportement de chaque point de vente - Nécessite de très grandes ressources techniques et économiques - Les aléas sont très fréquents donc le risque est élevé et quasiment non maitrisable - Données récoltées avec une faible fréquence (1 observation/par semaine) 	<p style="text-align: center;"><u>Les avantages :</u></p> <ul style="list-style-type: none"> - Réalisable techniquement et économiquement - Données récoltées avec une grande fréquence (1 observation /par jour) - Permet de maitriser la planification et la production <p style="text-align: center;"><u>Les obstacles et les inconvénients :</u></p> <ul style="list-style-type: none"> - Ne permet pas de connaitre le besoin de chaque point de vente. - Pertes considérables en cas d'erreur - Autres

En effet, les résultats présentés dans le tableau ci-dessus nous ont poussé à explorer une solution qui permet en même temps de bénéficier des avantages qu'offrent les prédictions par point de vente et les avantages des prédictions par ville, tout en réduisant les obstacles et les inconvénients.

La chaine de distribution de Candia permet d'exploiter une prédiction avec un degré de précision intermédiaire qui est la « prédiction par circuit ».

Le Circuit : un ensemble de points de vente caractérisés par une proximité géographique, approvisionnés le même jour, une fois par semaine par un même vendeur.

Les prédictions par circuits permettent de garder les avantages d'une prédiction par point de vente tout en apportant les améliorations suivantes :

- Eliminer les aléas inhérents au comportement d'un point de vente.
- Permet de réduire les erreurs de prédiction par point de vente grâce à une compensation mutuelle (plus on monte dans le niveau de granularité plus les erreurs se compensent).
- Nécessite moins de ressources.

La prédiction par circuit n'a pas été réalisée dans le présent travail pour les raisons suivantes :

- Le manque de données : c'est la raison la plus contraignante. Sur l'ensemble de l'année 2017, pas plus de 50 observations ont été disponibles dans les cas les plus favorables.
- La structure : jusqu'à la fin de l'année 2017, la ville de Tizi-Ouzou n'a toujours pas été entièrement divisée en circuits.

3.2.2. La veille concurrentielle

La veille concurrentielle qui est appelée aussi l'intelligence économique est un processus coordonné qui englobe les actions de collecte, de traitement, de diffusion et de protection de l'information au sein de l'entité économique.

Afin d'assurer une bonne vision sur l'environnement et d'anticiper les événements qui peuvent impacter les ventes de l'entreprise, le dirigeant doit mettre en place un système qui fait remonter l'information nécessaire pour la prise de décision. En effet, le diagnostic a révélé que plusieurs variables comme les pénuries de lait en sachet peuvent avoir un impact positif sur les ventes chez Candia.

Pour que cette information soit intégrée dans le système intelligent, un processus en trois étapes doit être réalisé :

- a) Collecte de l'information : le département force fore de vente doit appliquer les actions nécessaires qui permettent d'enregistrer les périodes de pénurie chez les concurrents.
- b) Traitement : les informations récoltées lors de la première étape vont être soumises à une analyse pour mesurer si réellement les pénuries affectent les ventes de Candia (principalement le demi écrémé et Viva).
- c) Exploitation : les résultats du traitement vont être intégrés dans le système (une variable exogène dans le réseau NARX) afin que ce dernier puisse améliorer les prédictions.
- d) Protection.

Le processus de veille est résumé dans la figure 38.

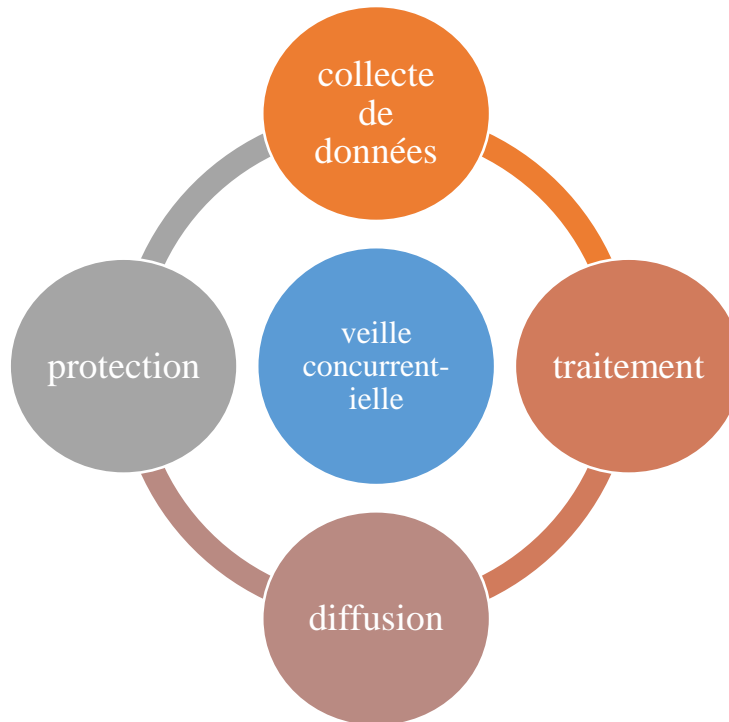


Figure 38 : Processus de veille

3.3. Technique

Une recherche a été réalisée sur les différents outils qui permettent d'appliquer l'apprentissage automatique sur la problématique annoncée par l'entreprise.

Après avoir éliminé les méthodes qui ne sont pas adéquates au problème, le choix s'est porté sur les réseaux de neurones récurrents à cause de leur développement technique, leur adéquation avec notre problème et nos données et aussi du fait qu'ils ne sont pas seulement des concepts théoriques mais qui ont été réellement appliqués sur les cas pratiques, les ressources technologiques qui permettent d'implémenter un réseau de neurones récurrent sont donc disponibles.

La simulation réalisée sur le logiciel Matlab a montré que la technique utilisée est capable de modéliser et de répondre à notre problème, reste à améliorer le système avec le temps.

3.4. Système informatique

La concrétisation du système intelligent basé sur les réseaux de neurones et qui prédit le comportement des points de vente peut être confiée à une entité spécialisée dans les systèmes d'information. Les recherches que nous avons menées nous ont permis de proposer l'utilisation de la bibliothèque open source Tensorflow de Google qui est une bibliothèque dédiée aux calculs numériques utilisant des graphes de flux de données.

Cette bibliothèque donne la possibilité aux développeurs d'échanger rapidement les données et de construire des applications plus facilement grâce aux multiples algorithmes déjà développés.

La multiplication des contributions des experts dans les toutes dernières années a conduit à un développement remarquable de la bibliothèque qui devient ainsi très riche et particulièrement adaptée pour les problèmes basés sur les réseaux de neurones artificiels.

Pour la nouvelle version 1.0 sortie en 2017, de nombreuses nouvelles API (*Application Programming Interface*) ont été ajoutées (les API pour les langages de programmation tel que Matlab, python, C++ existent bien évidemment), avec également l'introduction du nouveau module `tf.keras` qui offre la compatibilité avec Keras, la bibliothèque populaire de réseaux de neurones de haut niveau.

Grâce à cette solution, le concepteur aura la possibilité de choisir les indicateurs de performance les plus importants afin de les intégrer dans l'interface. La solution peut être hébergée soit dans un Cloud soit dans l'un des serveurs de l'entreprise.

Conclusion

Ce dernier chapitre nous a permis de choisir le réseau le plus adéquat pour implémenter le système de prédiction de l'entreprise. Le réseau NARX a réalisé des résultats très prometteurs grâce à une mémoire qui lui permet d'intégrer les précédentes ventes dans les futures prédictions. Un autre caractère crucial dans les réseaux NARX consiste en les variables exogènes qui permettent d'orienter le modèle à réaliser des résultats plus performants, c'est donc pour cette raison qu'il a été recommandé à l'entreprise d'entreprendre quelques actions qui permettent de récolter plus de facteurs exogènes (veille concurrentielle).

Finalement, ce système qui est en effet basé totalement sur les données a un réel potentiel à se développer grâce aux quantités de données que peut générer l'entreprise dans le futur.

Conclusion générale

L'objet de ce travail était de préparer les bases d'un système intelligent qui prédit le comportement des points de vente de Tchîn-lait et principalement la demande journalière. Cet outil, qui constituerait un assistant du manager, lui permettra d'avoir une meilleure visibilité sur ses clients représentés par les points de vente et d'anticiper leurs demandes.

La problématique a été proposée par l'entreprise Tchîn-lait qui est l'un des principaux acteurs du marché du lait UHT en Algérie, en plus d'être aussi un acteur important dans d'autres marchés tel que les jus et les laits-jus grâce à la large gamme de produits qu'elle commercialise.

Pour répondre à cette problématique, un diagnostic externe a été réalisé et qui nous a permis de comprendre les caractéristiques de l'environnement dans lequel évolue l'entreprise et de relever les éléments qui ont un impact direct sur les ventes. Le diagnostic interne a quant à lui été bénéfique pour connaître la relation de l'entreprise avec ses points de vente et les mécanismes qui régissent le flux physique et informationnel entre les deux parties.

L'analyse des données a été une étape très importante dans ce projet. Ainsi, les différents traitements effectués ont permis non seulement de mieux connaître l'activité de l'entreprise mais aussi d'avoir une idée précise sur le potentiel informationnel de l'entreprise en vérifiant la qualité et la quantité des données.

Partant des besoins de l'entreprise qui souhaiterait disposer d'un système dynamique pouvant s'adapter aux changements qui surviennent dans l'environnement, et après avoir vérifié la faisabilité de cet ambitieux projet, une large revue de littérature a été menée. Cette étape nous a permis de découvrir le monde de l'intelligence artificielle et principalement sa branche la plus prospère, en l'occurrence le *Machine Learning*.

Le *Machine Learning* (ou apprentissage supervisé) présente un nouveau paradigme dans l'interaction avec la machine. Cet outil a la capacité d'apprendre à partir de l'expérience –tout comme l'homme en effet- et d'améliorer ses performances en fonction des données qui lui sont injectées. Cela a grandement motivé notre orientation.

Après avoir classifié les algorithmes de l'apprentissage automatique selon les problèmes pour lesquels ils sont destinés, notre choix s'est porté sur les réseaux de neurones artificiels qui sont des outils en plein essor, très utilisés dans les problèmes de prédiction grâce à leurs notables qualités de prédiction.

Parmi les différents types de réseaux de neurones, le réseau NARX était le plus adéquat pour le cas d'étude. En effet, ce réseau permet de prédire les futures ventes en fonction des précédentes ventes tout en intégrant les facteurs exogènes qui influent sur la vente. Afin de vérifier la performance du modèle retenu, une simulation des ventes globales du produit Viva de Candia dans la ville de Tizi-Ouzou a été réalisée.

Cette simulation à permis de valider le choix du modèle et de montrer la capacité des réseaux NARX à effectuer des prédictions fiables avec peu de données. Cependant plusieurs améliorations peuvent être remarquées grâce à la disponibilité de plus de données.

A la fin de ce mémoire, un certain nombre de recommandations ont été formulées et qui permettent à l'entreprise de préparer l'environnement favorable pour accueillir ce futur système de prédiction. Celles-ci touchent à l'implémentation du système ainsi qu'aux supports qui assurent son bon fonctionnement.

Une fois le système mis en place, il constituera un outil conséquent d'aide à la décision qui assistera les managers à tous les niveaux de l'entreprise, stratégique comme opérationnel, à adapter leur plan d'action. En effet, les prédictions pourront être exploitées par le département force de vente (qui recherche une prédiction à l'échelle micro) afin d'éviter les ruptures chez les points de ventes et optimiser la flotte qui assure l'approvisionnement. Ce même système pourra être utilisé par le département production (qui est intéressé par une prédiction à l'échelle macro) afin d'ajuster les plans de production.

Enfin, ce travail était une immersion de l'ingénieur dans une réelle problématique d'entreprise en étant responsable sur une partie d'un projet ambitieux. Lors de ce projet nous avons été confrontés à de multiples contraintes techniques qui nous ont forcés à sortir de notre zone de confort et à apporter tout notre savoir-faire acquis durant la formation. Ce travail était aussi une très bonne opportunité pour nous afin de découvrir et de manipuler les outils de l'apprentissage automatique, une compétence qui devient de plus en plus importante dans les différents domaines de l'industrie et de l'ingénierie.

Bibliographie

- 12 Fevrier 2018. [https://in.mathworks.com/discovery/machine-learning.html?requestedDomain=.](https://in.mathworks.com/discovery/machine-learning.html?requestedDomain=)
- AlgerieEco, pourquoi les crises du lait sont elles récurrentes. *AlgérieEco*. 17 Decembre 2017. <https://www.algerie-eco.com/2017/12/17/algerie-crise-lait-recurrente/>.
- Aoudad. 2 Mai 2018. http://aoudad.perso.centrale-marseille.fr/TPE/ii_maths.php.
- Barron, Andrew. «Universal approximation bounds for superpositions of a sigmoidal function.» IEEE, 1993.
- Bengio, Toshua. 2 Septembre 2017. <https://www.techemergence.com/what-is-machine-learning/>.
- Bergman, M.D. «les bases théoriques de l'organisation du marché du lait.» 2016.
- BSA. 2018. http://www.boisson-sans-alcool.com/marques_jus-fruit-algerie.html.
- Cancella, Angela Marisa. «Comparative study of artificial Networks and Box & Jinkins ARIMA.» 2008.
- Danilo P. Mandic, Jonathan A. Chambers. «Recurrent Neural Network for Prediction.» 2001.
- Despagne, W. *Construction, Analyse et implémentation d'un modèle de prévision*. HAL, 2010.
- Dreyfus, Gerard. *Neural Networks, Methodology and Application*. Paris: Eyrolles, 2004.
- E. Shmider, Mathias Ziegler, Markus Buner. «Reinvestigating the Robustness of ANOVA.» Berlin, 2010.
- Eugen, Dianescu. «the use od Narx Neural Network to predict Chaotic Time series.» Romania, 2010.
- Graves, Alex. *Supervised Sequence Labeling with Recurrent Neural Network*. Vol. 385. Springer, 2012.
- Hava T. Siegelman, Bill G. Hornes and C. Lee Giles. «Computational Capabilities of Recurrent Neural Networks.» IEEE, 1997.
- Heaton, Jeff. «Artificial Intelligence for Humans Volume 3: Neural Networks and Deep Learning.» 2015.
- Hornik, Kurt. «Approximation Capabilities of FFNN.» Vienna, 1990.
- Johnson, J.P Hanna, R.J Millar. «Advances in Dataflow programming language.» 2004.
- laitDZ*. 2018 Mai 2018. http://www.boisson-sans-alcool.com/marques_lait-algerie.html.
- Lantz, Brett. *Machine Learning with R*. Mumbai: Pachet Publishing, 2013.
- Laurne, F. *Fundamental of Neural Network*. 1993.

- M.W Jonhson, J. P Hanna and R. Milner. *Advances in Dataflow Programming languages*. 2004.
- Mazur, Matt. 2018. https://www.google.com/search?source=hp&ei=Mo8fW-f3laKv6ATTnoPIBw&q=backpropagation+algorithm+mate+mazur&oq=backpropagation+algorithm+mate+mazur&gs_l=psy-ab.3..33i21k1j33i160k1l2.786.7852.0.7999.17.16.0.0.0.0.239.2134.0j7j4.11.0....0...1c.1.64.psy-ab..6.1.
- Mentzer, M. Moon and J. *Sales forecasting Managment: a Demand Management Approach*. SAGE Publications, 2004.
- Murphy, K. P. *Machine Learning a Probabilistic perspective*. 2012.
- Nilson, Nils J. «Introduction to Machine Learning.» Stanford University, 2012.
- Rapacci, Bernard. «dure, dure les series chronologiques.» aout 1993.
- Sharma, Sagar. 6 septembre 2017. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
- . *Activation function: Neural Network*. 2017. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
- Shukla, Nishant. *Machine Learning with Tansor Flow*. Manning, 2018.
- Topus, Vadat. «Hourly trafic flow predication.» Marmara University, 2010.
- Touzet, Claude. *Introduction au connexionisme*. 1992.
- Tristan, Sterin. *réseaux de neurones récurrnts*. Iyoun, Université Lyon 1, Département Informatique, 2016.
- W. Johnson, J.P Hanna, R.J Millar. «Advances in Dataflow programming language.» 2004.
- Watan, El. 18 Janvier 2018. http://www.elwatan.com/regions/centre/alger/la-penurie-de-lait-en-sachet-dure-depuis-des-semaines-08-01-2018-360121_148.php.
- Yahyaoui Salah, Kaci Mohamed. «Etude lait conditionné et boisson lactées en Algérie.» 2017.
- Youcef, Malik. *Micro RNA Biology and Computational Analytics*. Hatfield: Humana Press, 2014.
- Yuu, L. *Foreign rate exchange with Artificial Neural Networks*. 2007.
- Zhang, G.P. *Neural Networks for Businnes Forecasting*. Georgia: Idea Publishing, 2004.

Annexes

Annexe A : Liste des produits.

L'entreprise Tchik-lait commercialise une gamme de 27 produits différents :

- Les laits blancs : (5 Références)
 - * Lait Demi écrémé 1 L.
 - * Lait Demi écrémé 0.5 L.
 - * Lait Entier 1L.
 - * Lait Viva 1L.
 - * Lait Silhouette 1L.
-
- Les Candys : (7 Références)
 - * Candy Choco 1L.
 - * Candy Choco 20 cl.
 - * Candy Fraise 20 cl.
 - * Candy Choco 125 ml.
 - * Candy Caramel 125 ml.
 - * Candy Fraise 125 ml.
 - * Candy Banane 125 ml.
-
- Boissons aux fruits : (6 Références)
 - * Orange 1L.
 - * Orange 20cl.
 - * Cocktail 1L.
 - * Cocktail 20cl.
 - * Citronnade 1L.
 - * Nectar de grenade 1L.
-
- Twist (Lait et jus) : (8 Références)
 - * Orange/Ananas 1L.
 - * Orange/Ananas 20cl.
 - * Pêche/Abricot 1L.
 - * Pêche/Abricot 20cl.
 - * Orange/Mangue 1L.
 - * Orange/Mangue 20cl.
 - * Orange/Fraise/Banane 1L.
 - * Orange/Fraise/Banane 20cl.
-
- Autres : (1 Référence)
 - * Préparation culinaire 20cl.

Annexe B : Classification ABC.

Tableau 1 Annexe B : Classification ABC des produits

Produit :	Abréviation	Impact sur CA	Classe :
Lait Demi écrémé 1 L	DE 1l	32.34%	A
Candy Choco 20cl	Choco 20cl	27.02%	A
Lait Viva 1L	Viva	18.71%	A
Candy Choco 1L	Choco 1l	5.63%	B
Candy Choco 125 ml	Choco 125ml	3.55%	B
Twist Orange/Mangue 1L	Twist O-M 1l	3.25%	B
Lait Entier 1L	Entier	2.53%	B
Lait Silhouette 1L	Silhouette	1.65%	B
Lait Demi écrémé 50cl	DE 50cl	0.29%	C
Orange 1L	Orange 1l	0.29%	C
Twist Orange/Fraise/Banane 20cl	Twist O-F-B 20cl	0.87%	C
Twist Orange/ananas 20cl	Twist O-A 20cl	0.85%	C
Twist Pêche/Abricot 20cl	Twist P-A 20cl	0.76%	C
Citronnade 1L	Citronnade	0.37%	C
Cocktail 1L	Cocktail 1l	0.30%	C
Cocktail 20cl	Cocktail 20cl	0.30%	C
Orange 20cl	Orange 20cl	0.29%	C
Twist Orange/Mangue 20cl	Twist O-M 20cl	0.28%	C
Nectar de Grenade 1L	Nectar de Grenade 1l	0.19%	C
Twist Orange/Fraise/Banane 1L	Twist O-F-B 1l	0.14%	C
Candy Banane 125ml	Banane 125ml	0.13%	C
Twist Orange/ananas 1L	Twist O-A 1l	0.13%	C
Twist Pêche/Abricot 1L	Twist P-A 1l	0.12%	C
Préparation culinaire 20cl	Préparation culinaire	0.10%	C
Candy Fraise 125ml	Fraise 125ml	0.08%	C
Candy Fraise 20cl	Fraise 20cl	0.04%	C
Candy Caramel 125 ml	Caramel 125ml	0.01%	C

Annexe C : Liste des dépositaires.

Tableau 1 Annexe C : Liste des dépositaires.

Batna	1	Tlemcen	1
Biskra	1	Mascara	1
Ouargla	1	Sid Bel Abbes	1
Ghardaia	1	Ain Temouchent	1
El Oued	1	Mostaganem	2
Oum El Bouaghi	2	Alger	4
Laghouat	1	Boumerdes	1
Chlef	1	Blida	1
Ain Defla	1	Tipaza	1
Djelma	1	Tizi-Ouzou	1
Médéa	1	Bouira	1
Tiaret	1	Bejaia	3
Rélizane	1	JJEL	1
Annaba	1	Mila	1
Skikda	1	Sétif	2
Tebessa	1	BBA	1
Khanchela	1	M'sila	1
Guelma	1	Tindouf	1
Constantine	1	Tamenrasset	1
El Taref	1	Adrar	2
Oran	2	Tlemcen	1

Annexe D : Méthodes de prévisions.

La prévision est une branche des mathématiques appliquées très développée. Il existe deux types de prévisions :

- Les méthodes qualitatives : se basent sur un jugement subjectif.
- Les méthodes quantitatives : se basent sur des modèles statistiques bien définis.

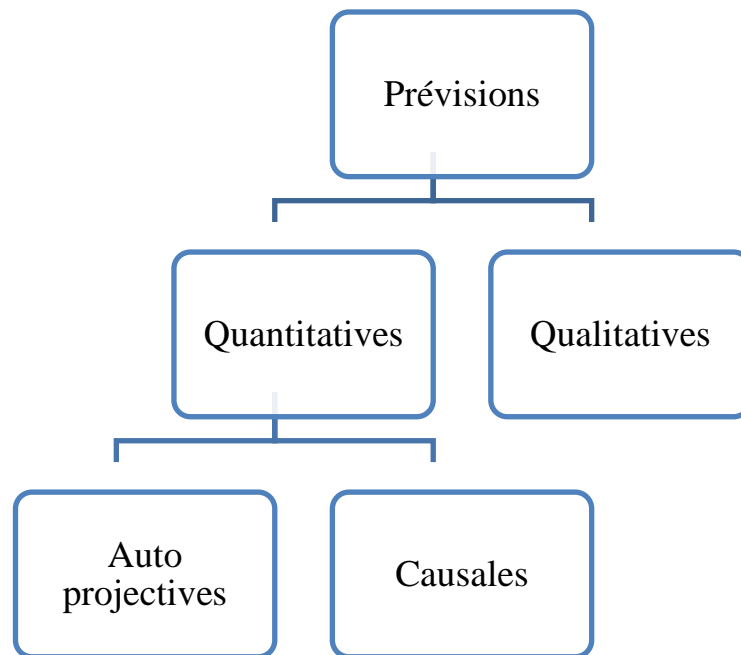


Figure 1 Annexe D : Types de prévisions et leurs méthodes.

Les méthodes qualitatives :

Ces méthodes sont utilisées par l'entreprise dans le cas où elle ne dispose pas de beaucoup de données, ou que le phénomène étudié est sans précédent. Cependant ses méthodes sont très coûteuses et ne sont fiables que sur le long terme. Parmi ces méthodes, il existe :

- La méthode Delphi.
- L'étude de marché.
- Les méthodes de comparaison (les analogies historiques).
- La méthode de sondage des opinions.

Les méthodes auto projectives :

Ces méthodes se basent sur les données historiques d'une variable X à des instants antérieurs t , $t-1$, ... $t-n$. la valeur de la variable X à l'instant $t+1$ est fonction de valeurs de X aux instants précédents. L'ensemble des valeurs de X est appelé série chronologique. Parmi les principales méthodes auto projectives, il existe :

- La méthode des moyennes mobiles (MA).

- Le modèle Autorégressif (AR).
- La méthode de lissage exponentiel simple.
- La méthode de lissage exponentiel double (Holt).
- La méthode de lissage exponentiel triple (Holt-Winters).
- La méthodologie de Box et Jenkins.

Les méthodes causales :

Ces méthodes expliquent une variable endogène par plusieurs variables explicatives ou exogènes. Celles-ci nécessitent la connaissance de toutes les variables exogènes ainsi que leurs valeurs historiques, chose qui est très difficile.

Annexe E : Historique des Réseaux de neurones.

Les réseaux de neurones artificiels sont des structures électroniques qui sont exécutés sur des machines de calcul ou même sur des circuits dédiés. Ils ont été inspiré principalement par le système nerveux humain afin de réaliser des tâches de classification, de regroupement de données, reconnaissance des formes, d'identification, ...

L'histoire des réseaux de neurones artificiels commence en 1943 dans un article publié par MC CULLOCH et PITTTS qui ont proposé l'idée simplificatrice du neurone formel : une opération (neurone) binaire interconnectée à d'autres opération par des « synapses » excitatrices ou inhibitrices.

En 1949, D.O.HEBB introduisit la notion de « plasticité synaptique », c'est-à-dire le mécanisme d'actualisation progressive des couplages inter neurones responsable d'une modification dans leurs propriétés collectives, phénomène connu aujourd'hui comme « l'apprentissage automatique ». Une règle par laquelle les couplages se modifient proportionnellement aux corrélations entre neurones, que ces corrélations soient négatives «inhibition » ou positives «activation ».

C'est en 1958 que ROSENBLATT décrivit le premier modèle opérationnel de réseaux de neurones, mettant en œuvre les idées de HEBB, MC CULLOCH et PITTTS, et annonçant ainsi un réel début d'une discipline qui connaîtra un grand essor.

En 1969, MINSKY et PEPART, ont démontré des limites théoriques relatives au perceptron. Un certain manque d'approfondissement de la théorie et le mouvement concurrent de l'intelligence artificielle étaient comme un frein au développement des réseaux de neurones, ainsi très peu de travaux scientifiques ont été publiés entre 1970 et 1980 et peu de chercheurs étaient toujours intéressés.

Les débuts des années 1980 ont assisté à une renaissance frappante de cette discipline grâce à HOPFIELD 1982, qui a détaillé dans un article le bon usage du réseau de neurones formels en s'appuyant sur l'organisation en couches et la simulation numérique par ordinateur. Aussi, WIDROW et HOFF ont beaucoup contribué avec leurs travaux sur les algorithmes adaptatifs.

En 1986, RUMBELHART, HINTON et WILLIAMS publient l'algorithme de rétro-propagation de l'erreur qui permet d'optimiser les paramètres d'un réseau de neurones à plusieurs couches. À partir de ce moment, la recherche sur les réseaux de neurones connaît un ressort fulgurant et les applications commerciales de ce succès académique suivent au cours des années 90[BIS].

Annexe F : Comparaison entre RN Biologique et RN Artificiel.

Les structures artificielles et les structures biologiques se ressemblent principalement dans les deux caractères suivants [ROK]:

- L'acquisition de la connaissance se réalise grâce à « l'apprentissage »
- Tout comme pour les réseaux de neurones biologiques, la connaissance d'un RNA est stockée dans les connexions entre les neurones avec une valeur appelé « poids synaptique ».

Le neurone biologique possède la structure simplifiée suivante :

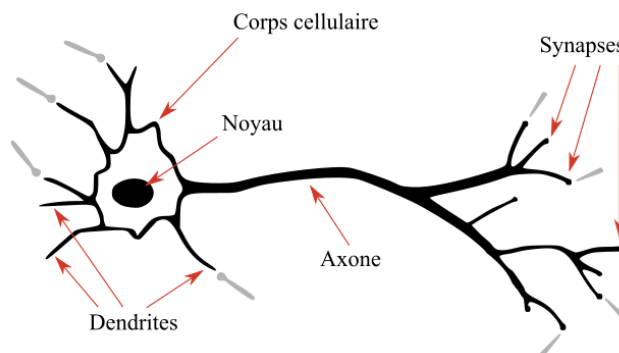


Figure 1 ANNEXE F : Structure du neurone biologique

Avec :

- Dendrite : reçoit les signaux des autres neurones en entrée.
- Corps cellulaire : additionne tous les signaux d'entrées.
- Axone : déclenche un signal de sortie en cas où la somme des entrées dépasse un seuil.
- Synapses : le poids d'interconnexion d'un neurone avec un autre.

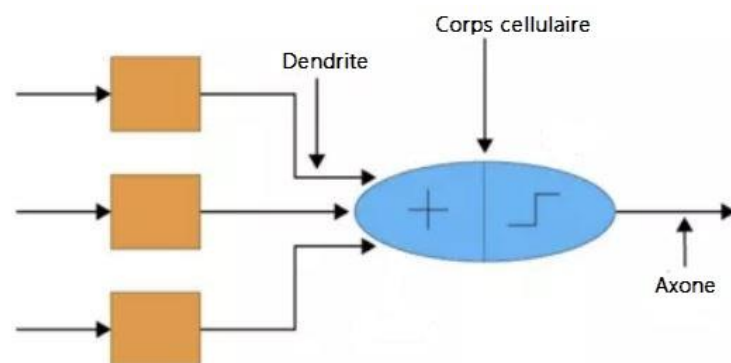


Figure 2 ANNEXE F : Analogie réseau de neurones artificielle et biologique

Annexe G : Algorithme du modèle.

Script MATLAB :

```
% Résolution d'un problème autoregressif avec NARX Neural Network.
% dans ce script les variables suivantes doivent être définies:
%
%   in - input time series.
%   out - feedback time series.

X = tonndata(in,false,false);
T = tonndata(out,false,false);

% choisir la fonction de transfert
% pour voir la liste de toutes les fonctions de transfert: help
  ntrain
% 'trainlm' est généralement plus rapide.
% 'trainbr' est lente mais efficace pour les problèmes complexes.
trainFcn = 'trainbr'; % Bayesian Regularization backpropagation.

% création du réseau
inputDelays = 1:2;
feedbackDelays = 1:2;
hiddenLayerSize = 5;
net =
  narxnet(inputDelays,feedbackDelays,hiddenLayerSize,'open',trainF
  cn);

% préparation des données pour l'entraînement et la simulation
% la fonction PREPARETS prépare les inputs et les outputs de la
  série chronologique
% l'utilisation de preparets permet de conserver l'ordre dans les
  données d'origine
[x,xi,ai,t] = preparets(net,X,{},T);

% séparation des données en 3 ensembles de : Training, Validation,
  Testing
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

% entraîner le réseau
[net,tr] = train(net,x,t,xi,ai);

% tester le réseau
y = net(x,xi,ai);
e = gsubtract(t,y);
performance = perform(net,t,y)

% voir le réseau
view(net)

% Plots
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, ploterrhist(e)
%figure, plotregression(t,y)
%figure, plotresponse(t,y)
```

```

%figure, ploterrcorr(e)
%figure, plotinerrcorr(x,e)

% Closed Loop Network
% la fonction CLOSELOOP remplacele feedback input avec une
  connexion direct à partir de l'output
netc = closeloop(net);
netc.name = [net.name ' - Closed Loop'];
view(netc)
[xc,xic,aic,tc] = preparets(netc,X,{},T);
yc = netc(xc,xic,aic);
closedLoopPerformance = perform(net,tc,yc)

% prediction pas en avant
nets = removedelay(net);
nets.name = [net.name ' - Predict One Step Ahead'];
view(nets)
[xs,xis,ais,ts] = preparets(nets,X,{},T);
ys = nets(xs,xis,ais);
stepAheadPerformance = perform(nets,ts,ys)

```