



République Tunisienne
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Carthage - Ecole Supérieure de la Statistique et de l'Analyse de l'Information



Rapport de Projet de Fin d'Etudes présenté pour l'obtention du
Diplôme National d'Ingénieur en Statistique et Analyse de l'Information



Par

Nader GHRAB

Modélisation des éléments du Business Plan

Soutenu le 21/07/2020 devant le jury composé de :

M. Mokhtar KOUKI	Président
Mme Ines ABDELJAOUED	Rapporteur
M. Hichem RAMMEH	Encadrant

Projet de Fin d'Études encadré par

Mme Rym Ben Salah

Fait à Assurances Zitouna Takaful



Dédicaces

*A mes chers parents,
qui ont tout sacrifié pour faire de moi ce que je suis aujourd'hui. Aucun mot ne pourrait exprimer à
leur juste valeur l'amour et la gratitude que je vous porte. Que Dieu vous procure une longue vie
pleine de santé et de bonheur.*

*A ma soeur et mon frère,
Je vous dédie ce travail en témoignage de gratitude, d'estime et d'attachement. Puisse dieu vous
accorder santé, longue vie , prospérité et que le sourire éclaire toujours votre vie.*

*A toute ma famille et mes meilleurs amis,
qui n'ont jamais cessé de m'encourager*

Remerciements

La réalisation de ce stage a été possible grâce à la contribution de plusieurs personnes à qui je voudrais témoigner toute ma reconnaissance et ma gratitude pour leur générosité et la patience qu'ils ont eue envers mon travail.

Tout d'abord, j'adresse mes remerciements à mon encadrant universitaire M. Hichem Rammeh, pour la confiance qu'il m'a accordée en acceptant d'encadrer ce travail, pour ses multiples conseils et pur son soutien tout au long du stage.

Mes remerciements sont adressés aussi aux membres du jury, pour avoir l'extrême gentillesse de bien vouloir évaluer mon travail.

Finalement, je tiens à remercier ma famille et mes amis de m'avoir soutenu tout au long de mon travail et de mon parcours universitaire.

Résumé

Le présent projet a été réalisé à chez la compagnie Assurance Zitouna Takaful au sein de la direction Enterprise Risk Management (ERM).

Il s'inscrit dans le cadre d'un projet de fin d'études en vue de l'obtention d'un diplôme d'ingénieur en statistique et analyse de l'information.

Le rôle du business-plan évolue en passant d'un outil convaincant pour l'obtention d'un financement dans la phase de création d'un projet à un outil d'aide à la décision durant la vie de l'entreprise. Le but du projet est de créer un business-plan incluant les différents indicateurs détaillés d'une compagnie d'assurance Takaful ainsi que la création d'un modèle statistique pour l'estimation de son chiffre d'affaires.

Ce modèle vise à fournir aux gestionnaires un outil d'aide à la décision qui permet, dans une optique de prédiction, de concevoir ou de modifier les éléments du business plan. Dans ce contexte, nous construisons différents modèles prédictifs des séries temporelles avec des méthodes basées sur l'apprentissage automatique et l'auto regressivité.

Mots-clés : Business plan, Plan d'affaires, Chiffre d'affaires, IBNR, Séries Temporelles, LSTM, SARIMA, Chain Ladder.

Table des matières

Table des matières	v
Table des figures	vi
Liste des tableaux	viii
Liste des algorithmes	ix
Introduction	1
1 Présentation du Projet de Fin d'Études	3
1.1 La structure du marché des assurances	3
1.2 Présentation de l'organisme d'accueil (Zitouna Takaful)	5
1.2.1 Présentation générale	5
1.2.2 Branches	5
1.2.3 Missions	6
1.2.4 Valeurs	6
1.2.5 Les modèles de gestion des entreprises d'assurance Takaful (en particulier Zitouna Takaful)	6
1.2.6 Département d'accueil	6
1.3 Problématique et objectif	7
1.4 Conclusion	7
2 Définition et notions de bases d'une compagnie d'assurance	8
2.1 Les éléments du business plan	8
2.1.1 Le business-plan	8
2.1.2 Chiffre d'affaires	9
2.1.3 Sinistre	10
2.1.4 Réassurance	10
2.1.5 Budget	10
2.1.6 Trésorerie	10
2.1.7 Ratios	11
2.2 Quelques concepts de base	11
2.3 Conclusion	13
3 Formulation des modèles pour la prévision du chiffre d'affaires	14
3.1 Serie temporelle	14
3.2 Méthodes de prévision des series temporelles	15

3.2.1	Méthodologie traditionnelle (modèle SARIMA)	16
3.2.2	Approche deep learning	19
3.2.3	Métrique d'évaluation	21
3.3	Algorithmes	22
3.4	Conclusion	24
4	Application des modèles : modélisation du chiffre d'affaires	25
4.1	Chiffre d'affaires global	25
4.2	SARIMA	25
4.2.1	Erreur de prédiction (année 2019)	34
4.2.2	Analyse des résidus	37
4.3	LSTM	38
4.3.1	Préparation de données :	38
4.3.2	Normalisation des données :	39
4.3.3	Préparation du vecteur d'entrée pour le LSTM :	39
4.3.4	Architecture du modèle :	40
4.3.5	Prédiction du chiffre d'affaires de l'année 2019 :	40
4.3.6	Prédiction du chiffre d'affaires de l'année 2020 :	42
4.4	Comparaison (SARIMA vs LSTM) :	44
4.5	Chiffre d'affaires par branche	47
4.5.1	Branche Automobile	47
4.5.2	Branche Risques divers et spéciaux (RDS)	49
4.5.3	Branche Incendie	51
4.5.4	Branche Transport	53
4.5.5	Branche Family (assurance vie)	54
4.6	Conclusion	56
5	Sinistralité : modélisation de l'IBNR	57
5.1	Présentation des données	57
5.2	Méthode de chain ladder	58
5.2.1	Triangle Dommages Automobile	59
5.2.2	Traçage de triangles	60
5.2.3	Application de la méthode "chain Ladder"(triangle automobile)	61
5.3	Conclusion	63
	Conclusion	64
	Annexe	65
	A Structure du business-plan	65
	B Structure du Chiffre d'affaires	67
	C LSTM	68
	D La classe TimeseriesGenerator	69
	Bibliographie	72

Table des figures

1.1	Rrincipaux indicateurs d'activité 2017-2018	4
1.2	Evolution des principaux indicateurs d'activité 2017-2018	4
4.1	Chiffre d'affaires global	26
4.2	"auto_arima"	27
4.3	1ère differencation	28
4.4	Augmented Dickey-Fuller	28
4.5	KPSS test	29
4.6	ACF	30
4.7	PACF	31
4.8	les modèles	32
4.9	les modèles selectionnés	32
4.10	les modèles significatifs	32
4.11	Modèle 1	33
4.12	Modèle 2	33
4.13	Forecast 2019	34
4.14	RMSE	34
4.15	Prédiction du chiffre d'affaires global	35
4.16	Chiffre d'affaires 2020 avec le modèle SARIMA	36
4.17	Les résidus du modèle SARIMA	37
4.18	Préparation de données	38
4.19	Normalisation des données	39
4.20	La classe TimeseriesGenerator	39
4.21	Architecture du modèle LSTM	40
4.22	Apprentissage du modèle LSTM	40
4.23	prediction	41
4.24	Prediction LSTM vs valeur réelles : année 2019	41
4.25	Prediction LSTM 2019	42
4.26	RMSE du modèle LSTM	42
4.27	Prédiction du chiffre d'affaires 2020 avec LSTM	43
4.28	LSTM 2020	43
4.29	Valeurs réelles&Sarima&LSTM : année 2019	44
4.30	Comparaison entre SARIMA et LSTM avec RMSE	44
4.31	Valeurs réelles vs SARIMA vs LSTM : année 2019	45
5.1	Triangle de liquidation	58
5.2	Triangle	59

5.3	Tableau d'évolution des sinistres du triangle Automobile, avec une ligne par période d'origine	60
5.4	Tableau d'évolution des sinistres du triangle automobile, avec des panneaux individuels pour chaque période d'origine	61
5.5	Extrapolation log-linéaire des facteurs d'âge	62
5.6	schéma de developpement des sinistres attendus	62
5.7	triangle intégral	63
5.8	Perte totale	63
A.1	Structure du Business Plan	66
B.1	Structure du Chiffre d'affaires	67
C.1	Structure de LSTM	68
D.1	LSTM	69
D.2	La classe TimeseriesGenerator	70
D.3	Paramètres de la classe TimeseriesGenerator	71

Liste des tableaux

4.1	Comparaison entre le réel et la prédiction 2019 SARIMA LSTM	45
4.2	Comparaison entre LSTM et SARIMA 2020	45
4.3	Points forts et faibles (LSTM & SARIMA)	46
4.13	Evolution chiffre d'affaires par branche (2019-2020)	56

Liste des Algorithmes

1	SARIMA	22
2	LSTM	23

Introduction générale

De nos jours, le besoin d'avoir des analyses fines sur l'activité d'une entreprise devient une clé de réussite. D'où l'apparition des différentes méthodes d'analyse selon les activités de la société. Le rôle du business plan pour une entreprise est de fixer des objectifs sur un horizon d'un an et plus et c'est, essentiellement, pour deux buts. Le premier est de trouver des financements et le deuxième est de clarifier la vision de l'avenir de l'entreprise pour l'entrepreneur. Le business plan sert à réaliser des autres objectifs. En fait, il est utilisé comme un repère à partir duquel les entrepreneurs peuvent faire une comparaison de la différence entre la réalité et l'estimation de sorte qu'ils peuvent trouver des explications sur l'origine du différentiel. De ce fait, tant que l'entreprise élabore un business plan avec un niveau de détails fins, elle pourrait analyser sa situation d'une façon plus précise.

Il s'agit d'une étape absolument indispensable lors de la création ou de la reprise d'une compagnie, mais pas uniquement, il est également pertinent de mettre à jour son business plan de façon régulière au cours de la vie de l'entreprise afin d'en planifier le développement de façon structurée.

Dans l'univers des assurances, une convention réunit l'assureur et les assurés. En vertu de cet accord, les assurés versent une prime à l'assureur qui s'engage à les garantir d'une indemnisation en cas de dommage, de maladie, de perte, etc. Le "business model" assurance a été imaginé depuis de nombreux siècles et il repose sur la mutualisation des risques et leur prise en charge. Grâce à l'accumulation des primes versées par les clients, les compagnies ont la possibilité de rembourser financièrement les demandes d'indemnisation. Pour accroître les gains, les assureurs réalisent par ailleurs divers placements et investissements avec les primes collectées.

D'où le besoin d'une bonne prédiction de certains indicateurs d'activité.

Pour ce besoin, la compagnie d'assurance Zitouna Takaful nous a proposé de modéliser les éléments de son business-plan. Ce dernier doit répondre aux détails qu'elle veut voir sur ses différents indicateurs d'activité et qu'il permet d'estimer, à partir d'une méthode statistique, son chiffre d'affaires de la période de 2020 jusqu'à 2024.

Ce présent rapport contient cinq chapitres. Le premier est dédié à la présentation du cadre du projet où nous allons présenter le secteur des assurances, l'entreprise d'accueil ainsi que la présentation du projet.

Ensuite, dans le deuxième chapitre, nous parlons des définitions de quelques notions utiles à la modélisation des éléments du plan d'affaires.

Puis, nous illustrons dans le troisième chapitre pour la formulation des modèles que nous allons utiliser dans la prévision du chiffre d'affaires. Le quatrième chapitre est consacré pour l'application et les résultats de notre modélisation du chiffre d'affaires.

Enfin, le cinquième chapitre est consacré à la modélisation d’une partie de la sinistralité : les sinistres tardifs.

Chapitre 1

Présentation du Projet de Fin d'Études

Resumé :

Dans ce chapitre introductif, nous allons présenter le cadre du projet de fin d'études.

Dans un premier temps, nous présenterons le marché des assurances. Ensuite, nous allons parler de l'organisme d'accueil, Assurance Zitouna Takaful. Enfin, nous développerons la problématique et les objectifs de ce travail.

1.1 La structure du marché des assurances

Le marché des assurances en Tunisie compte 23 entreprises résidentes et 6 compagnies off-shore ; 19 entreprises résidentes opèrent sous le statut de société anonyme et 4 sont constituées en société à forme mutuelle [11]. La plupart des compagnies opèrent en multi-branches et certaines d'entre elles sont spécialisées dans une activité particulière d'assurance [11] :

- **Cinq en assurance vie** : Hayett, Gat Vie, Magherbia Vie, Carte Vie et Attijari Assurance
- **Une en assurance des exportations et crédits à l'exportation** : Cotunace qui a absorbé à la fin de l'année 2012 la compagnie «Assurcrédit» spécialisée en assurance des crédits commerciaux locaux lors d'une opération de fusion par absorption
- **Une en réassurance** : Tunis Ré
- **Trois en assurance Takaful** : Zitouna Takaful, El Amana Takaful et Assurances Takafulia.

Le secteur des assurances contient aussi un réseau étendu sur toute la Tunisie moyennant les agents d'assurance qui comptent 918, les bureaux de courtage qui sont 82 et les 64 producteurs d'assurance vie [11].

Quant aux experts, le marché compte près de 920 experts et 100 commissaires d'avaries. Par contre, il ne contient que 19 actuaires [11].

Les principaux indicateurs d'activité sont le chiffre d'affaires, indemnisations réglées, provisions techniques, charges techniques et résultat technique. Le tableau et le graphe ci-dessous illustre leur évolution au cours de deux années 2017-2018 [12].

PRINCIPAUX INDICATEURS D'ACTIVITE 2017-2018

Indicateurs (M.D)	2017	2018	Tx d'évolution (%) 2018/2017
Chiffres d'Affaires	2 087,9	2 252,4	7,9
Indemnisations Régliées	1 053,8	1 262,8	19,8
Charges Techniques	598,4	640,5	7
Provisions Techniques (1)	4506,6	5 026,8	11,5
Placements cumulés (2)	4 913,8	5 443,3	10,8
Taux de couverture (%) (3)	104,4	103,3	-1,1
Résultat Technique	66,7	115,9	73,8

(1) : Total des provisions techniques selon les données des bilans des sociétés d'assurance directe + Provisions pour prévisions de recours à encaisser à la clôture (B 1-1 et B 1-2)

(2) : Total des placements (Source: Bilans des sociétés d'assurance: AC3 + AC4)

(3) : Taux de couverture des provisions techniques par les actifs placés en représentation (sans tenir compte de l'insuffisance de provisions dégagée après les travaux de contrôle)

FIGURE 1.1 – Principaux indicateurs d'activité 2017-2018

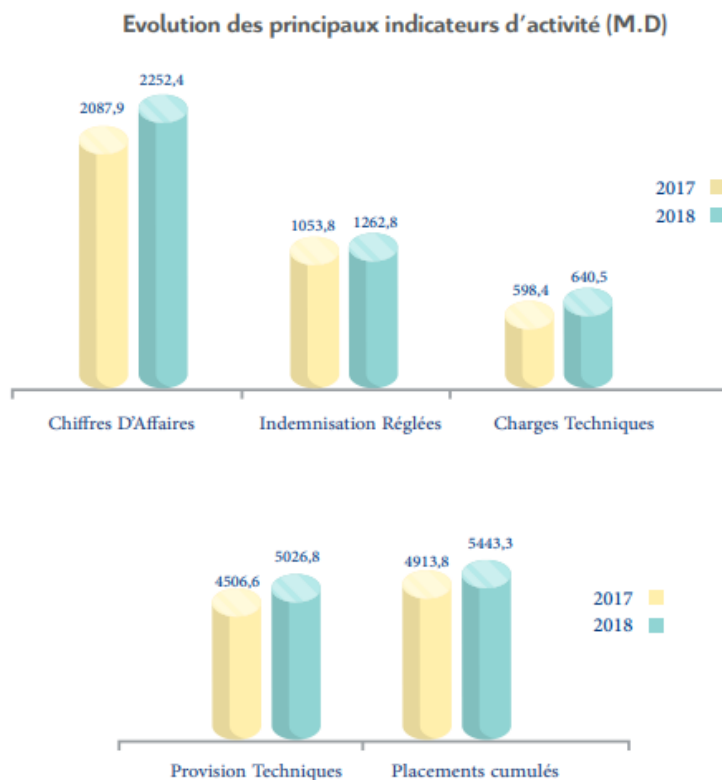


FIGURE 1.2 – Evolution des principaux indicateurs d'activité 2017-2018

1.2 Présentation de l'organisme d'accueil (Zitouna Takaful)

Nous avons effectué le stage de projet de fin d'études au sein de Zitouna Takaful plus précisément au niveau de la Direction Enterprise Risk Management (ERM).

1.2.1 Présentation générale

Zitouna Takaful est la première compagnie d'assurances et de réassurances spécialisée en assurance Takaful en Tunisie [23]. Elle a été créée en 2011 avec un capital de 15 000 000 DT [23] . Elle opère selon le concept Takaful. Ce dernier est caractérisé par :

- Séparation des fonds
- Placement dans des actifs conformes à la Charia'a
- Distribution des excédents techniques
- Supervision par le Comité Charia'a

La compagnie d'assurance Zitouna Takaful a connu une évolution importante de l'année 2012 à l'année 2019 puisqu'elle a démarré son activité en 2012. En prime, elle a atteint en 2019 60 992 203 DT contre 23 933 279,23 DT coût de sinistre.

Le principal rôle de ZITOUNA TAKAFUL est de veiller à la bonne gestion du fonds des adhérents selon les principes de finance islamique qui est basée sur l'interdiction des quatre éléments suivants [23] :

- **El Gharar** : L'incertitude e manque de connaissance ou l'absence d'informations précises, en effet il existe une disproportion considérable entre la prime payée par l'assuré et le capital ou l'indemnité qui il va peut-être recevoir, cette disproportion associée à ce caractère aléatoire ce qui est interdit par la chariaa.
- **El Maysir** : La spéculation et le jeu de hasard, l'assuré risque de perdre totalement la prime d'assurance en cas de non-réalisation du risque.
- **El Riba** : Investir le fonds de l'assurance dans les moyens de financement (obligations, actions... etc.) générant des intérêts.
- **Haram** : Placer le fonds dans des activités illicites et interdites par la chariaa comme l'alcool, le jeu...

1.2.2 Branches

Les produits de Zitouna Takaful sont classifiés en six branches, cinq en assurance vie et une en assurance non vie :

- Automobile
- Maladie
- Risque Divers et Spéciaux (RDS)
- Incendie
- Family (assurance vie)

1.2.3 Missions

Les principales missions de la compagnie Zitouna Takaful sont [23] :

- Offrir une panoplie de produits et services Takaful Général et Takaful Family de haute valeur ajoutée conformes aux principes de la finance islamique.
- Participer à l'innovation et à la diversification des produits d'assurances Tunisiens.
- Contribuer au développement économique du pays, à la pérennité des entreprises et au bien-être social des particuliers.

1.2.4 Valeurs

- Transparence : C'est la pierre angulaire du modèle de fonctionnement.
- Conformité : Le mode de gestion ainsi que les produits sont approuvés par un comité charia' a constitué par d'éminentes personnalités de la finance islamique en Tunisie.
- Responsabilité sociale : Se placer au diapason des changements que connaît l'environnement et contribuer à l'édification du bien-être social de participants.
- Innovation : les solutions proposées sont le reflet des besoins évolutifs des clients.
- Professionnalisme : Chaque client mérite un accompagnement personnalisé et une réponse pertinente et rapide à ses attentes [23].

1.2.5 Les modèles de gestion des entreprises d'assurance Takaful (en particulier Zitouna Takaful)

Pour ce qui est de la disposition opératoire des contrats d'assurance Takaful, elle est essentiellement opérée selon l'une des formes suivantes [23] :

- Le modèle wakala : c'est-à-dire gérer pour le compte des autres, et dans lequel la société d'assurance est mandatée par le souscripteur/assuré à gérer les contrats d'assurance contre la perception d'une commission établie au début de l'exercice et qui couvre les frais de gestion. Le surplus éventuel de chaque exercice revient de droit aux souscripteurs.
- Le modèle mudaraba : basé sur un contrat de participation aux bénéfices, et consiste généralement au partage des profits réalisés, selon un partage préétabli par exemple 70 % pour les fonds et 30 % pour l'opérateur, après déduction des frais de gestion.
- Le modèle Mixte : qui tend à être le plus utilisé actuellement et qui consiste à utiliser le modèle wakala pour la souscription et mudaraba pour les investissements liés au fonds.

1.2.6 Département d'accueil

L'assurance Zitouna Takaful comporte six directions à savoir la direction générale, la direction de système d'information (DSI), la direction technique, la direction comptabilité et finance, la direction commerciale et réseau et la direction Entreprise Risk Management (ERM). Cette dernière accueille le sujet de notre projet. La direction ERM comporte trois départements : le Département Gestion des Risques, le Département Conformité, et le Département Contrôle de Gestion et le département. Ce dernier est chargé de la gestion et suivi de budget, reporting interne, reporting réglementaire (pour le Comité Générale d'Assurance (CGA) et la Fédération Tunisienne des Sociétés d'Assurances (FTUSA)), les tableaux de bord, l'étude ponctuelle pour la Direction Générale et l'élaboration du business plan.

1.3 Problématique et objectif

Le business-plan consiste essentiellement à fixer des objectifs, sur un an et plus, que l'entreprise veut achever. il fournit des informations très détaillées qui servent à aider les preneurs de décisions à bien orienter les activités de la compagnie, en particulier pour une compagnie d'assurance, à savoir, les raisons d'une forte sinistralité aide à trouver des solutions. En effet, s'il existe un déficit inhabituel dans une branche déterminée, le business-plan pourra donner des explications. En fait, savoir l'objectif détaillé par garantie pourra localiser l'origine de l'anomalie et permettre la possibilité de la corriger. Par exemple, l'entreprise pourra changer ses conditions de souscriptions, sa stratégie de vente, faire une augmentation des provisions techniques etc. De plus, le business plan permet de simuler des scénarios. Il inclut des courbes, des graphiques et des indicateurs qui illustrent la situation financière de la compagnie. Donc, il permet de visualiser des différentes possibilités ; optimistes et pessimistes avec un niveau de détails fins et sur une période bien déterminée.

Les estimations sur lesquelles se base le business plan actuel de Zitouna Takaful ne sont pas modélisées. Elles sont basées sur l'historique de la compagnie et le savoir faire des collaborateurs. D'où la naissance du besoin de modéliser le business plan avec plus de détails afin d'analyser d'une façon plus fine les résultats de l'entreprise et de répondre à son évolution. Notre projet consiste à satisfaire ce besoin. Nous sommes chargés par la modélisation du business-plan de la compagnie Zitouna Takaful sur une période de cinq ans (de 2020 jusqu'à 2024). Il supposé qu'il va être détaillé de façon à afficher :

- le chiffre d'affaires
- la sinistralité
- la réassurance.
- le budget
- la trésorerie

Vu la crise sanitaire, le stage a été suspendu pendant la période de confinement. De ce fait, nous arrivons à modéliser seulement le chiffre d'affaires et une partie de la sinistralité.

Quant à la prévision, notre projet va s'intéresser à donner des prévisions du chiffre d'affaires par une méthode statistique. Dans le but d'obtenir des résultats pertinents, nous allons fixer l'horizon de notre prévision à un an, en effet, nous n'avons pas assez d'historique qui nous habilita de faire une prévision de cinq ans.

1.4 Conclusion

Dans ce premier chapitre, nous avons présenté le secteur des assurances en Tunisie ainsi que la compagnie et le département d'accueil. Nous avons aussi présenté notre projet en général. Dans le chapitre suivant, nous allons définir de plus le business plan et quelques notions d'une compagnie d'assurance.

Chapitre 2

Définition et notions de bases d'une compagnie d'assurance

Resumé :

Nous définissons dans ce chapitre le business-plan, sa composition et son rôle. Nous expliquons ensuite son importance pour une compagnie d'assurance. Nous introduisons finalement les différentes notions utiles pour notre projet.

2.1 Les éléments du business plan

2.1.1 Le business-plan

Nous commençons par la définition du business-plan dans le domaine de l'entrepreneuriat puis nous expliquons son importance pour une compagnie d'assurance.

Le business-plan en général

Le Business Plan, est un terme emprunté de la littérature anglo-saxonne et il est appelé en littérature francophone par plan d'affaires [14]. C'est un outil indispensable dans le domaine de l'entrepreneuriat. Son élaboration se fait généralement dans la phase de la création d'un projet. Il aide l'entrepreneur à réfléchir afin de construire son projet et lui permet de trouver des financements. Il sert à prévoir la rentabilité et les perspectives à un horizon de temps donné pour pouvoir attirer des tiers. Ces derniers ont besoin d'une compréhension globale du projet pour qu'ils puissent le juger. D'où la nécessité de l'élaborer avec attention pour pouvoir les convaincre. Son importance à l'intérieur de l'entreprise se concrétise dans son rôle en tant qu'un outil d'aide à la décision et en tant qu'un outil de gestion. En fait, il permet de veiller aux objectifs du projet et au plan de développement de l'entreprise.

Il contient deux parties. Une première partie rédactionnelle comporte la présentation du projet, ses objectifs ainsi qu'une étude de l'environnement concurrentiel [8]. La deuxième partie est la partie financière [9]. Elle contient les prévisions du chiffre d'affaires, le bilan prévisionnel et les principaux indicateurs de l'activité de l'entreprise.

Le Business-Plan, alors, sert à contrôler les stratégies de la compagnie. Ici, nous mentionnons qu'il existe une différence entre le plan d'affaires et le plan stratégique. Ce dernier, également appelé plan marketing, est seulement à usage interne. Il consiste à développer et à détailler la stratégie commerciale. Le Business-plan englobe l'idée du plan stratégique et il couvre de plus d'autres aspects tels que le marketing, le budget, les ressources humaines etc.

Le business-plan pour une compagnie d'assurance

Comme tout autre projet, la création d'une compagnie d'assurance débute par l'élaboration du business-plan. Elle l'utilise comme nous avons déjà mentionné dans le premier paragraphe du chapitre; en externe pour des objectifs de financement et en interne pour contrôler son activité. Le business plan ne sert pas uniquement au démarrage de la compagnie, pour cela, elle doit l'actualiser chaque année et elle doit comparer les réalisations par rapport aux objectifs fixés. L'importance de ce dernier se manifeste essentiellement dans le suivi du plan stratégique et dans sa capacité de fournir des chiffres en détail fin. De ce fait, il est considéré comme un outil très puissant d'aide à la décision. Il permet d'avoir une vue globale sur l'avenir de la compagnie de plusieurs indicateurs financiers. Ce qui permet de prendre des décisions décisives concernant les plans stratégiques. Et il permet également de présenter des tableaux de bords mensuels.

Dans ce qui suit, Nous parlerons des définitions des différents éléments du business plan de l'assurance.

2.1.2 Chiffre d'affaires

Le chiffre d'affaires, pour une compagnie d'assurance, est la somme des primes émises. La prime, appelée en Takaful « contribution », quant à elle, se définit par « la rémunération que perçoit l'assureur et/ou le réassureur en contrepartie du risque assuré conformément aux termes du contrat [23]. Le montant de la prime ou de la cotisation intègre les différents éléments suivants :

- La prime pure qui correspond à la valeur probable des prestations pécuniaires qui seront versées aux assurés. Elle intègre notamment la probabilité de survenance du sinistre et l'intensité du sinistre probable (son coût moyen) et la valeur actuelle probable.
- Les chargements qui visent couvrir les frais de gestion et d'acquisition engagés par l'entreprise d'assurance et/ou de réassurance.
- Les taxes d'assurance et les diverses contributions qui sont collectées par l'assureur et reversées au profit de l'Etat.

L'ensemble constitué par la prime pure et les chargements de gestion des risques est dénommé prime de risque [18]. La prime de risque majorée des autres chargements est appelée prime commerciale, et lorsqu'elle est augmentée des taxes, prime totale [18].

2.1.3 Sinistre

Le sinistre lié à l'activité d'assurance vie est défini par «la réalisation du risque, objet du contrat, ou de l'arrivée à échéance du contrat, ou du rachat de contrat, qui serait de nature à entraîner les garanties de l'assureur. Les sinistres ou prestations payés correspondent à l'indemnisation prévue au contrat dans le cas de la réalisation du risque» [19].

Le sinistre lié à l'activité d'assurance non vie est défini par «la réalisation du risque objet du contrat et de nature à entraîner les garanties de l'assureur. Les sinistres payés correspondent à l'indemnisation des dommages garantis subis par les assurés, ou que les assurés ont eux-mêmes fait subir à des tiers» [19].

2.1.4 Réassurance

La réassurance est l'assurance des assureurs. Les compagnies d'assurance achètent des traités pour s'assurer contre les grands risques souscrits [17].

Il existe deux types de couverture : les traités ou encore appelés obligatoires et les facultatives.

La réassurance obligatoire (Le traité)

la réassurance est obligatoire lorsque la négociation porte sur une branche d'assurance, l'assureur s'engageant à céder tous les risques relevant de cette branche et le réassureur s'engageant à les accepter [17].

La réassurance facultative

la réassurance est facultative lorsque la négociation se fait risque par risque en laissant une totale liberté de souscription au réassureur ainsi qu'à l'assureur [2].

2.1.5 Budget

Le budget prévisionnel est une étape indispensable dans le développement de l'entreprise. Il est basé sur une estimation réaliste des dépenses et des recettes et permet d'anticiper le futur résultat. Il s'agit d'un tableau financier regroupant l'ensemble des dépenses (ou « charges ») et des recettes (ou « produits ») liées au projet pour l'année en cours ou à venir. L'année concernée est appelée « exercice ». Elle correspond généralement à une année civile et fiscale [5]. Le budget prévisionnel permet d'anticiper l'ensemble des sorties et entrées d'argent, et ainsi d'assurer que l'entreprise est rentable et suffisamment capitalisée.

2.1.6 Trésorerie

Il s'agit d'un tableau récapitulant les encaissements et les décaissements prévisionnels de l'entreprise. Le plan de trésorerie est détaillé mois par mois pour offrir une visibilité aux décideurs sur les flux de trésorerie prévisibles à court terme. Ceci permet d'estimer les dépenses et les recettes futures et de dégager la situation finale en veillant à l'équilibrer [3]. Le plan de trésorerie permet de ressortir avec le solde prévisionnel mensuel, ce qui permet de faire face aux dépenses d'un mois sur l'autre.

L'utilisation d'un plan de trésorerie présente plusieurs intérêts pour l'entrepreneur [13] :

- S'assurer que l'entreprise pourra faire face aux décaissements programmés sur les jours, semaines ou mois à venir
- Evaluer la capacité de l'entreprise à financer son développement
- Evaluer la capacité de l'entreprise à financer son développement
- En cas d'activité saisonnière, vérifier que la trésorerie permettra de faire face aux périodes creuses

2.1.7 Ratios

Il existe plusieurs ratios permettant d'évaluer l'activité d'une compagnie d'assurance. Les deux principaux ratios sont le ratio de solvabilité et le ratio de représentation.

Ratio de solvabilité

Ratio de solvabilité ou Taux de couverture de la marge de solvabilité (ou marge de solvabilité sur fonds propres) : il est égal au rapport entre la marge de solvabilité effective de l'assureur et l'exigence minimale de marge de solvabilité (calculée sur la base de ses engagements et des risques qu'il doit couvrir) [10].

Taux de représentation

Ce ratio mesure la capacité d'une entreprise à honorer ses engagements (provisions) vis à vis des participants (assurés).

Il est calculé par la formule suivante :

$$\text{Taux de représentation} = \frac{\text{placements}}{\text{provisions}}$$

2.2 Quelques concepts de base

Quittance

Une quittance d'assurance était un document faisant état du paiement effectif de la prime. Avec les nouveaux moyens de paiement tel que le prélèvement automatique, nombreux sont les assureurs qui associent facture et quittance, cette dernière étant alors accompagnée de la mention « sous réserve d'encaissement » [15].

Une quittance d'assurance indique que la prime correspondant au risque assuré a été payée. La différence peut paraître négligeable si l'on considère que le paiement de la prime pour un risque donné atteste que celui-ci est couvert. Ce serait le cas si les assureurs ne fournissaient la quittance qu'une fois la cotisation effectivement encaissée.

Affaire nouvelle (prime au comptant)

Une nouvelle affaire (contrat d'assurances ou « police ») donne lieu à une prime payable (au comptant). Ceci génère un « avis de comptant » ou « avis comptant », une demande de paiement adressée au preneur d'assurance. Généralement, à cet « avis comptant » est joint un formulaire de virement où les données de l'assureur sont pré-encodées. En principe suit le paiement par le preneur d'assurances. Et puis l'assureur délivre une quittance comme preuve de réception du paiement [6].

Terme

La nouvelle affaire arrive par après à sa « date d'échéance », soit après une première année, soit dans le cas de paiement fractionné après un semestre, trimestre ou même mensuellement. Ceci donne chaque fois lieu à un prime (-terme) payable, et donc à un « avis d'échéance terme » ou « avis terme » ; une suivante demande de paiement adressée au preneur d'assurance. A cet « avis terme » est à nouveau joint un formulaire de virement où les données de l'assureur sont pré-encodées. En principe suit le paiement par le preneur d'assurances. Et puis l'assureur délivre une quittance comme preuve de réception du paiement [7].

Les primes émises

Ce sont les primes ayant fait l'objet d'une émission. On distingue généralement en assurance non-vie les primes comptant, relatives aux affaires nouvelles et les primes terme, pour les affaires en tacite reconduction ou pour celles dont les primes sont fractionnées c'est à dire dont le paiement est effectué selon un échéancier prédéterminé indépendant de la période de garantie contractuelle. En assurance vie, on distingue les primes uniques, les primes périodiques et les versements libres [18].

Les annulations de primes

Les annulations de primes sont les quittances qui ont été annulées par l'entreprise d'assurance et/ou de réassurance du fait notamment du non-paiement des primes, de primes émises à tort par l'entreprise ou d'avenants avant l'échéance modifiant les conditions du contrat [24].

Les primes cédées

Les primes cédées correspondent aux primes qui sont reversées par la cédante au cessionnaire (assureur ou réassureur) en contrepartie des risques qu'il a accepté de garantir [18].

Les ristournes de primes ou de cotisations

Il s'agit du remboursement à l'assuré d'une partie de la cotisation ou de la prime qu'il a versée. Il peut s'agir notamment des remboursements effectués pour la modification ou la suspension volontaire des garanties du contrat [20].

Provision

Une provision mathématique est une provision nécessaire à la couverture d'un engagement [22]. C'est le montant qu'un assureur doit détenir dans ses comptes pour garantir son engagement vis-à-vis des souscripteurs de contrats [22]. Son calcul est relativement encadré par la réglementation et intègre la mortalité, les intérêts financiers, les cotisations futures que le souscripteur s'est éventuellement engagé à verser en contrepartie des engagements de l'assureur [22].

Il existe de nombreuses sortes de provisions. Pour des risques du type non vie relatifs aux biens, et pour les risques d'assurance de personnes dits « courts » (rapidité des paiements : assurance décès, arrêt de travail et frais de soin), les deux principales provisions sont les provisions pour sinistres connus et les provisions pour sinistres inconnus [26] :

- les provisions pour sinistres connus correspondent aux sinistres dont l'assureur a eu connaissance, mais qu'il n'a pas encore pu régler à son assuré
- les provisions pour sinistres inconnus (ou IBNR : Incurred But Not Reported) correspondent aux sinistres qui sont déjà survenus, mais dont lesquels l'assureur n'a pas encore connaissance. Cette provision est en général établie de manière statistique par des actuaires.

IBNR

En assurance, les sinistres encourus mais non déclarés (IBNR) sont le montant dû par un assureur à tous les demandeurs valables qui ont eu un sinistre couvert mais ne l'ont pas encore déclaré. Comme l'assureur ne connaît ni le nombre de ces pertes, ni la gravité de chaque perte, l'IBNR est nécessairement une estimation. La somme des pertes IBNR et des pertes déclarées donne une estimation du total des engagements éventuels que l'assureur couvrira, connus sous le nom de pertes ultimes [1].

2.3 Conclusion

Nous avons introduit qu'est-ce qu'un business-plan ainsi que son rôle dans le domaine d'entrepreneuriat. Nous avons parlé de son importance pour une compagnie d'assurance. Nous avons défini les différentes nouvelles notions modalisées. Malheureusement nous n'avons pas pu arriver, à modéliser tous ces éléments vu que le stage a été suspendu. Nous parlerons, dans la suite, de la modélisation du chiffre d'affaires de Zitouna Takaful.

Chapitre 3

Formulation des modèles pour la prévision du chiffre d'affaires

Resumé :

Ce chapitre est consacré à l'étude mathématique des modèles prévisionnels pour l'évaluation du chiffre d'affaires. Nous allons considérer ces éléments comme une suite de valeurs numériques représentant l'évolution de la prime hors taxes (chiffre d'affaires) au cours du temps. Ensuite, nous allons présenter une méthode approchée pour la généralisation d'une série temporelle en se basant sur deux approches.

Enfin, nous allons expliquer ces modèles de point de vue mathématique.

3.1 Série temporelle

Une série temporelle est une série de points de données indexés (ou listés ou représentés graphiquement) dans l'ordre chronologique. Le plus souvent, une série temporelle est une séquence prise à des points successifs équidistants dans le temps. La prévision des séries chronologiques est un domaine important de la prévision dans lequel les observations passées de la même variable sont collectées et analysées pour élaborer un modèle décrivant la relation sous-jacente. Le modèle est ensuite utilisé pour extrapoler la série chronologique dans le l'avenir. Cette approche de modélisation est particulièrement utile lorsque peu de connaissances sont disponibles sur le processus de génération des données sous-jacentes ou lorsqu'il n'y a pas d'explication satisfaisante du modèle qui met en relation la variable de prédiction avec d'autres variables explicatives. Beaucoup d'effort a été consacré au cours des dernières décennies au développement et à l'amélioration des modèles de prévision de séries chronologiques.

3.2 Méthodes de prévision des series temporelles

L'un des modèles de séries chronologiques les plus importants et les plus répandus est le modèle autorégressif de moyenne mobile intégrée (ARIMA). La popularité du modèle ARIMA est en raison de ses propriétés statistiques ainsi que de la méthodologie bien connue de Box-Jenkins dans le processus de construction de modèles. ARIMA est une méthode de prévision pour les données de séries chronologiques univariées.

Comme son nom l'indique, elle prend en charge à la fois des éléments de moyenne mobile et autorégressive. L'élément intégré fait référence à la différenciation permettant à la méthode de prendre en charge des données de séries chronologiques avec une tendance.

Ce qui ne va pas avec ARIMA

Dans une série chronologique, la saisonnalité est un schéma régulier de changements qui se répète sur S périodes de temps, où S définit le nombre de périodes de temps jusqu'à ce que le schéma se répète à nouveau. Bien que le modèle ARIMA puisse traiter des données avec une tendance, il ne prend pas en charge les séries chronologiques avec une composante saisonnière. En fait, ARIMA attend des données qui ne sont pas saisonnières ou dont la composante saisonnière a été supprimée.

Qu'est-ce que le SARIMA ?

Une extension d'ARIMA qui prend en charge la modélisation directe de la composante saisonnière de la série est appelée SARIMA (Seasonal Autoregressive Integrated Moving Average). Il ajoute trois nouveaux hyperparamètres pour spécifier l'autorégression (AR), la différence (I) et la moyenne mobile (MA) pour la composante saisonnière de la série, ainsi qu'un paramètre supplémentaire pour la période de la saisonnalité.

En outre, il existe plusieurs approches différentes de la modélisation des séries chronologiques. Les modèles statistiques traditionnels, notamment la moyenne mobile, le lissage exponentiel et SARIMA, sont linéaires en ce sens que les prévisions des valeurs futures sont contraintes d'être des fonctions linéaires du passé des observations.

En raison de leur relative simplicité de compréhension et de mise en œuvre, ces modèles ont été les principaux axes de recherche et outils appliqués au cours des dernières années quelques décennies.

Bien que ces soient tout à fait flexibles en ce sens qu'ils peuvent représenter plusieurs types de séries chronologiques différentes, c'est-à-dire purement autorégressives saisonnières (AR saisonnier), la moyenne mobile pure saisonnière (MA saisonnier) et les séries combinées AR saisonnier et MA saisonnier (SARMA), leurs principale limitation est la forme linéaire présumée du modèle. C'est-à-dire qu'une structure de corrélation linéaire est supposée parmi les valeurs de la série temporelle et donc qu'aucune valeur non linéaire peuvent être capturés par le modèles SARIMA et ARIMA. L'approximation des modèles linéaires à un problème complexe du monde réel n'est pas toujours satisfaisant.

Pour surmonter la restriction des modèles linéaires et pour tenir compte de certains non linéarités observés dans des problèmes réels, plusieurs classes de modèles non linéaires ont été proposées dans la littérature. Il s'agit notamment du modèle conditionnel autorégressif hétéroscédastique (ARCH).

Bien qu'une certaine amélioration ait été constatée avec ces modèles non linéaires, le gain de leur utilisation pour les problèmes de prévision générale est limité.

Récemment, les techniques d'apprentissage automatique et, plus important encore, les algorithmes d'apprentissage profond (Deep Learning) ont introduit de nouvelles approches aux problèmes de prédiction où les relations entre les variables sont modélisées dans une hiérarchie profonde et stratifiée.

Les algorithmes d'apprentissage approfondi tels que le réseau neuronal récurrent (RNN) et la mémoire à long terme et à court terme (LSTM) ont suscité beaucoup d'intérêt ces dernières années, avec leurs applications dans de nombreuses disciplines, dont finance.

Les méthodes d'apprentissage approfondi sont capables d'identifier la structure et le modèle des données, comme la non-linéarité et la complexité des prévisions en séries chronologiques. Elles sont flexibles avec les modèles non linéaires. LSTM a été notamment utilisé dans la prévision de séries chronologiques et dans les données économiques et finances.

Dans la suite, nous nous concentrons sur les principes de base et le processus de modélisation de SARIMA et RNN.

3.2.1 Méthodologie traditionnelle (modèle SARIMA)

Le modèle saisonnier ARIMA intègre des facteurs non saisonniers et saisonniers dans un modèle multiplicatif. Une notation abrégée du modèle est [25] :

$$ARIMA(p, d, q)(P, D, Q)S$$

avec p = ordre AR non saisonnier, d = différenciation non saisonnière, q = ordre MA non saisonnier, P = ordre AR saisonnier, D = différenciation saisonnière, Q = ordre MA saisonnier, et S = période de répétition du schéma saisonnier [25].

Sans opérations de différenciation, le modèle pourrait être formé de manière plus formelle comme suit :

$$\Phi(B^S)\phi(B)(x_t - \mu) = \Theta(B^S)\theta(B)\epsilon_t \quad (1)$$

$$\begin{aligned} \text{AR : } \phi(B) &= 1 - \phi_1(B) - \dots - \phi_p(B^p) & \text{AR saisonnier : } \Phi(B^S) &= 1 - \Phi_1(B^S) - \dots - \Phi_P(B^P S) \\ \text{MA : } \theta(B) &= 1 - \theta_1(B) - \dots - \theta_q(B^q) & \text{MA saisonnier : } \Theta(B^S) &= 1 - \Theta_1(B^S) - \dots - \Theta_Q(B^Q S) \end{aligned}$$

où x_t et ϵ_t sont respectivement la valeur réelle et l'erreur aléatoire à la période t ; ϕ_i ($i=1,2,\dots,p$), Φ_i ($i=1,2,\dots,P$), θ_j et Θ_j ($0,1,2,\dots,Q$) sont des paramètres du modèle ; B est l'opérateur de backshift (c'est-à-dire $B^k x_t = x_{t-k}$ for $k \geq 1$) ; p , P , q , et Q sont des entiers et souvent appelés ordres du modèle. Les erreurs aléatoires, ϵ_t , sont supposées être distribuées de manière indépendante et identique avec une moyenne de zéro et une variance constante de σ^2 . L'équation (1) implique plusieurs cas particuliers importants de la famille de modèles ARIMA. Si $q = 0$, alors (1) devient un modèle AR saisonnier d'ordre p . Lorsque $p = 0$, le modèle se réduit à un modèle saisonnier d'ordre MA q . Une des tâches principales de la construction du modèle SARIMA est déterminer le modèle approprié (p,P,q,Q) [25].

Décomposition des séries temporelles

Le tracé des données des séries chronologiques est une première étape importante dans l'analyse de leurs différentes composantes. Au-delà, cependant, nous avons besoin d'un moyen plus formel pour identifier et supprimer des caractéristiques telles qu'une tendance ou une variation saisonnière. Le modèle de décomposition réduit une série chronologique en trois composantes : tendance, effets saisonniers et erreurs aléatoires. En retour, nous cherchons à modéliser les erreurs aléatoires comme une forme de processus stationnaire.

Ainsi, une série chronologique utilisant un modèle additif peut être considérée comme :

$$x_t = m_t + s_t + e_t$$

où, au temps t , m_t est la tendance, s_t est l'effet saisonnier, et e_t est une erreur aléatoire que l'on suppose généralement avoir une moyenne nulle et être corrélée dans le temps. Ainsi, en estimant et en soustrayant à la fois m_t et s_t à partir de x_t nous espérons avoir une série chronologique de résidus stationnaires e_t

alors qu'un modèle multiplicatif serait : $x_t = m_t * s_t * e_t$

Un modèle additif serait utilisé lorsque les variations autour de la tendance ne varient pas avec le niveau de la série temporelle, tandis qu'un modèle multiplicatif serait approprié si la tendance est proportionnelle au niveau de la série temporelle [21].

Différenciation

Il peut être nécessaire d'examiner des données différenciées lorsque nous avons une saisonnalité. La saisonnalité rend généralement la série non stationnaire parce que les valeurs moyennes à certains moments particuliers de la période saisonnière (les mois, par exemple) peuvent être différentes des valeurs moyennes à d'autres moments [21].

Différenciation saisonnière

La différenciation saisonnière est définie comme une différence entre une valeur et une valeur avec décalage qui est un multiple de S . Avec $S = 12$, ce qui peut se produire avec des données mensuelles, une différence saisonnière est :

$$(1 - B^{12})x_t = x_t - x_{t-12}$$

La différenciation saisonnière supprime la tendance saisonnière et peut également éliminer un type de marche aléatoire saisonnière de non-stationnarité [25].

Différenciation non saisonnière

Si les données présentent une tendance, nous pouvons également avoir besoin d'une différenciation non saisonnière. Souvent (pas toujours), une première différence (non saisonnière) éliminera la tendance. En d'autres termes, en présence de tendance nous utilisons [25] :

$$(1 - B)^d x_t = x_t - x_{t-d}$$

Différenciation de tendance et de saisonnalité

Lorsque la tendance et la saisonnalité sont toutes les deux présentes, il peut être nécessaire d'appliquer à la fois une première différence non saisonnière et une différence saisonnière. Autrement dit, nous pourrions avoir besoin d'examiner le ACF et le PACF de [25] :

$$(1 - B^S)^D (1 - B)^d x_t = (x_t - x_{t-d}) - (x_{t-S} - x_{t-S-D})$$

Corrélation dans et entre les séries chronologiques

Les concepts de covariance et de corrélation sont très importants dans l'analyse des séries chronologiques. En particulier, nous pouvons examiner la structure de corrélation des données originales ou des erreurs aléatoires d'un modèle de décomposition pour nous aider à identifier la ou les formes possibles de modèle(s) (non)stationnaire(s) pour le processus stochastique [21].

Fonction d'autocorrélation (ACF)

L'autocorrélation est la corrélation d'une variable avec elle-même à différents décalages temporels. Rappelons que nous avons défini la fonction d'autocovariance de l'échantillon (ACVF), c_k , pour certains lag k comme [21] : $c_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$.

L'autocovariance de l'échantillon de x_t au retard 0, c_0 , est égal à la variance d'échantillon de x_t calculée avec un dénominateur de n . La fonction d'autocorrélation de l'échantillon (FAC) est définie comme [21] :

$$r_k = \frac{c_k}{c_0} = Cor(x_t, x_{t+k})$$

Rappelons également qu'un intervalle de confiance d'environ 95% sur l'ACF peut être estimé

$$\frac{-1}{n} \pm \frac{2}{\sqrt{n}}$$

où n est le nombre de points de données utilisés dans le calcul du FAC.

Il est important de rappeler deux choses ici.

Premièrement, bien que l'intervalle de confiance soit couramment tracé et interprété comme une ligne horizontale sur tous les décalages temporels, l'intervalle lui-même s'accroît en fait à mesure que le décalage augmente parce que le nombre de points de données n utilisée pour estimer la corrélation diminue de 1 pour chaque augmentation entière du retard.

Deuxièmement, il faut être prudent dans l'interprétation de la "signification" de la corrélation à divers retards car il faut s'attendre, a priori, à ce qu'environ 1 corrélation sur 20 soit significative sur la base du seul hasard.

Fonction d'autocorrélation partielle (PACF)

La fonction d'autocorrélation partielle (PACF) mesure la corrélation linéaire d'une série x_t et d'une version décalée de celle-ci x_{t+k} , la dépendance linéaire de $x_{t-1}, x_{t-2}, \dots, x_{t-(k-1)}$ étant supprimée. Rappelez-vous de la conférence que nous définissons le PACF comme [21] :

$$f_k = \begin{cases} Cor(x_1, x_0) = r_1 & \text{si } k = 1; \\ Cor(x_k - x_k^{k-1}, x_0 - x_0^{k-1}) & \text{si } k \geq 2; \end{cases}$$

Bruit blanc

Une série chronologique w_t est une série de bruits blancs discrets (BB) si le w_1, w_2, \dots, w_t sont indépendants et identiquement distribués (iid) avec une moyenne de zéro. Nous supposons que les $w_t \sim N(0, q)$ et nous nous référons donc aux séries chronologiques w_t comme un bruit blanc gaussien. Si notre modèle de série temporelle a réussi à éliminer toute l'autocorrélation en série dans la série temporelle avec les tendances, les effets saisonniers, etc., alors les résidus du modèle ($e_t = y_t - \hat{y}_t$) sera une séquence BB ayant les propriétés suivantes pour sa moyenne (\bar{e}), la covariance (c_k), et l'autocorrélation (r_k) [21] :

$$f_k = \begin{cases} x = 0 \\ c_k = Cov(e_t, e_{t+k}) & si \quad k = 1; \\ Cor(x_k - x_k^{k-1}, x_0 - x_0^{k-1}) & si \quad k \geq 2; \end{cases}$$

Tests de Portmanteau pour l'autocorrélation

En plus d'examiner la courbe ACF, nous pouvons également effectuer un test d'autocorrélation plus formel en considérant tout un ensemble de valeurs r_k comme un groupe, plutôt que de traiter chacune d'entre elles séparément. Rappelons que r_k est l'autocorrélation pour le décalage k . Lorsque

nous examinons le graphique ACF pour voir si chaque pic se situe dans les limites requises, nous effectuons implicitement de multiples tests d'hypothèse, chacun avec une faible probabilité de donner un faux positif. Lorsqu'un nombre suffisant de ces tests sera effectué, il est probable qu'au moins

un d'entre eux donnera un faux positif, et nous pourrions donc conclure que les résidus présentent encore une certaine autocorrélation, alors qu'en fait ils n'en présentent pas. Afin de surmonter ce

problème, nous testons si les h premières autocorrélations sont significativement différentes de ce que l'on attendrait d'un processus de bruit blanc. Un test pour un groupe d'autocorrélations est appelé test de **portmanteau**. L'un de ces tests est le test de Box-Pierce, basé sur la statistique suivante :

$$Q = T \sum_{k=1}^h r_k^2,$$

où h est le retard maximum considéré et T est le nombre d'observations. Si chaque r_k est proche de zéro, alors Q sera petit. Si certaines valeurs de r_k sont grandes (positives ou négatives), alors Q sera grand. Nous utiliserons $h=10$ pour les données non saisonnières et $h=2m$ pour les données

saisonnières, où m est la période de saisonnalité. Cependant, le test n'est pas bon lorsque h est grand, donc si ces valeurs sont plus grandes que $T/5$, alors nous utilisons $h=T/5$.

Un test connexe (et plus précis) est le test Ljung-Box, basé sur :

$$Q^* = T(T+2) \sum_{k=1}^h (T-k)^{-1} r_k^2$$

les grandes valeurs de Q suggèrent aussi que les autocorrélations ne proviennent pas d'une série de bruit blanc [4].

3.2.2 Approche deep learning

"Long Short-Term Memory" (LSTM) est une sorte de Réseau neuronal récurrent (RNN) ayant la capacité se souvenir des valeurs des étapes précédentes dans le but d'utilisation future. Avant d'approfondir le LSTM, il est nécessaire de avoir un aperçu de ce à quoi ressemble un réseau de neurones [16].

Artificial Neural Network (ANN) :

Un réseau de neurones est constitué d'au moins trois couches, à savoir 1) une couche d'entrée, 2) des couches cachées, et 3) une couche de sortie. Les caractéristiques de l'ensemble de données ("features") détermine le nombre de nœuds dans la couche d'entrée. Ces nœuds sont reliés par des liens appelés "synapses" aux nœuds créés dans la ou les couches cachées. Les liens synapses portent un poids pour chaque nœud de la couche d'entrée. Les poids jouent essentiellement le rôle d'un décideur qui décide quel signal, ou quelle entrée, peut passer et lequel ne peut pas passer. Les poids indiquent également la force ou l'étendue de la couche cachée. Un réseau de neurones apprend essentiellement en ajustant le poids de chaque synopsis [16].

Dans les couches cachées, les nœuds appliquent une fonction d'activation (par exemple, sigmoïde ou tangente hyperbolique (\tanh)) sur la somme pondérée des entrées pour transformer les entrées en sorties, ou en valeurs prédites. La couche de sortie génère un vecteur de probabilités pour les différentes sorties et sélectionne celle qui présente le taux d'erreur ou le coût le plus faible, c'est-à-dire qui minimise les différences entre les valeurs attendues et les valeurs prédites, également appelées coût, à l'aide d'une fonction appelée SoftMax.

Les affectations au vecteur de pondération et donc les erreurs obtenus grâce à la formation en réseau pour la première fois pourraient pas être le meilleur. Pour trouver les valeurs les plus optimales pour les erreurs, les erreurs sont "retransmises" dans le réseau à partir de la vers les couches cachées et, par conséquent, la les poids sont ajustés. La procédure est répétée, c'est-à-dire les époques, plusieurs fois avec les mêmes observations et les poids sont réajusté jusqu'à ce qu'il y ait une amélioration de la et ensuite dans le coût. Lorsque la fonction de coût est minimisée, le modèle est formé. La

relation entre la sortie (y_t) et les entrées ($y_{t-1} ; y_{t-2}; \dots; y_{t-p}$) ont la valeur mathématique suivante représentation :

$$y_t = \alpha_0 + \sum_{j=1}^q (\alpha_j g(\beta_{0j} + \sum_{i=1}^p (\beta_{ij} y_{t-i}))) + \epsilon_t$$

où α_j ($j = 0, 1, 2, \dots, q$) et β_{ij} ($i = 0, 1, 2, \dots, p, j = 1, 2, \dots, q$) sont les paramètres du modèle souvent appelés poids de connexion ; p est le nombre de nœuds d'entrée et q est le nombre de nœuds cachés. La fonction logistique est souvent utilisée comme couche cachée fonction de transfert, c'est-à-dire,

$$g(x) = \frac{1}{1 + \exp(-x)}$$

Ainsi, le modèle ANN réalise en fait une cartographie fonctionnelle non linéaire à partir les observations passées ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) à la valeur future y_t , c'est-à-dire

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w) + \epsilon_t$$

où w est un vecteur de tous les paramètres et f est une fonction déterminée par le réseau la structure et les poids de connexion. Ainsi, le réseau de neurones est équivalent à un réseau non linéaire modèle autorégressif. Notez que l'expression implique un noeud de sortie dans la sortie qui est généralement utilisée pour les prévisions à un pas d'avance.

Recurrent Neural Network (RNN) :

Un RNN est un cas particulier de réseau neuronal où l'objectif est de prévoir la prochaine étape dans la séquence de les observations relatives aux étapes précédentes observées dans la séquence. En

fait, l'idée derrière les RNN est d'utiliser d'observations séquentielles et de tirer les leçons des étapes précédentes pour prévoir les tendances futures. Par conséquent, les données des premières étapes Il faut se souvenir des étapes suivantes. Dans RNN, les couches cachées servent de stockage interne pour le stockage des les informations saisies lors des premières étapes de la lecture séquentielle données. Les RNN sont appelés "récurrents" parce qu'ils effectuent les même tâche pour chaque élément de la séquence, avec la caractéristique d'utiliser des informations capturées plus tôt pour prédire des données séquentielles futures invisibles. Le défi majeur avec un RNN typique est que ces réseaux ne se souviennent quelques étapes antérieures dans la séquence et ne sont donc pas adaptées à la mémorisation de séquences de données plus longues. Ce défi Le problème est résolu en utilisant la "ligne de mémoire" introduite dans le "Long Short-Term Memory" (LSTM) [16].

Long Short-Term Memory (LSTM) :

LSTM est un type spécial de RNN avec des caractéristiques supplémentaires pour mémoriser la séquence de données. La mémorisation de la tendance antérieure des données est possible grâce à certaines portes ainsi qu'à une ligne de mémoire incorporée dans un LSTM typique. Chaque LSTM est un ensemble de cellules, ou modules de système, où les flux de données sont capturés et stockés. Les cellules ressemblent à une ligne de transport (la ligne supérieure de chaque cellule) qui se connecte d'un module à l'autre, transportant les données du passé et les rassemblant pour le présent. En raison de l'utilisation de certaines portes dans chaque cellule, les données de chaque cellule peuvent être éliminées, filtrées ou ajoutées pour les cellules suivantes. Ainsi, les "gates"(les portes), qui sont basées sur le neurone sigmoïdal permettent aux cellules de laisser passer les données ou de les éliminer, au choix. Chaque couche sigmoïde donne des nombres dans la plage de zéro et un, représentant la quantité de chaque segment de données qui doit passer dans chaque cellule. Plus précisément, une estimation de la valeur zéro implique de "ne rien laisser passer", tandis qu'une estimation de la valeur un indique de "tout laisser passer". Trois types de portes sont impliqués dans chaque LSTM dans le but de contrôler l'état de chaque cellule [16] :

- **"Forget Gate"** produit un nombre entre 0 et 1, où Le chiffre 1 indique "garder complètement ceci" ; tandis que le chiffre 0 implique "ignorez complètement cela".
- **"Memory Gate"** choisit les nouvelles données qui doivent être stockées dans la cellule. D'abord, une couche sigmoïde, appelée "porte d'entrée layer" choisit les valeurs qui seront modifiées. Ensuite, La couche de "tanh" fait un vecteur de nouvelles valeurs candidates qui pourrait être ajouté.
- **"Memory Gate"** décide la sortie de chaque cellule. La valeur produite sera basée sur l'état de la cellule le ainsi que sur les données filtrées et nouvellement ajoutées.

3.2.3 Métrique d'évaluation

Le "Root-Mean-Square Error" (erreur quadratique moyenne) (RMSE) est une mesure fréquemment utilisée pour évaluer la précision des prédictions obtenues par un modèle. Elle mesure les différences ou les résidus entre les valeurs réelles et prédites. La mesure compare les erreurs de prédiction de différents modèles pour une donnée particulière et non entre des ensembles de données. La formule de calcul de la RMSE est la suivante [16] :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}$$

Où N est le nombre total d'observations, x_i est le valeur réelle ; alors que \hat{x}_i est la valeur prédite. Le principal avantage de l'utilisation de la RMSE est qu'elle pénalise les erreurs importantes. Elle permet également d'échelonner les scores dans les mêmes unités que les valeurs prévues.

3.3 Algorithmes

Algorithme 1 : SARIMA

Inputs : series

Outputs : RMSE of the forecasted data

```
1 Split data into : 80% training and 20% testing data size  $\leftarrow \text{length}(\text{series}) * 0.80$ 
2 train  $\leftarrow \text{series}[0 \dots \text{size}]$ 
3 test  $\leftarrow \text{series}[\text{size} \dots \text{length}(\text{size})]$ 
4 history  $\leftarrow \text{train}$  predictions  $\leftarrow \text{empty}$ 
5 foreach  $t$  in  $\text{range}(\text{length}(\text{test}))$  do
6   model  $\leftarrow \text{SARIMA}(\text{history}, \text{order} = (p, d, q), \text{seasonal\_order} = (P, D, Q, S))$ 
7   model fit  $\leftarrow \text{model.fit}()$ 
8   hat  $\leftarrow \text{model\_fit.forecast}()$ 
9   predictions.append(hat)
10  observed  $\leftarrow \text{test}[t]$ 
11  history.append(observed)
12 end
13 MSE = mean_squared_error(test, predictions)
14 RMSE = sqrt(MSE)
15 Return RMSE
```

[16]

Algorithm 2 : LSTM

Inputs : Time series

Outputs : RMSE of the forecasted data

```
1 Split data into : 80% training and 20% testing data size  $\leftarrow \text{length}(\text{series}) * 0.80$ 
2  $\text{train} \leftarrow \text{series}[0 \dots \text{size}]$ 
3  $\text{test} \leftarrow \text{series}[\text{size} \dots \text{length}(\text{size})]$ 
4 Set the random seed to a fixed value : set random.seed(7)
5 Fit an LSTM model to training data
6 Procedure fit_lstm(train, epoch, neurons)
7  $X \leftarrow \text{train}$ 
8  $y \leftarrow \text{train} - X$ 
9 model = Sequential()
10 model.add(LSTM(neurons), stateful=True))
11 model.compile(loss='mean_squared_error', optimizer='adam')
12 foreach  $i$  in  $\text{range}(\text{epoch})$  do
13 |   model.fit(X, y, epochs=1, shuffle=False)
14 |   model.reset_states()
15 end
16 return model
17 Make a one-step forecast
18 Procedure forecast_lstm(model, X)
19  $\hat{y} \leftarrow \text{model.predict}(X)$ 
20 return  $\hat{y}$ 
21  $\text{epoch} \leftarrow \text{nb\_epochs}$ 
22  $\text{neurons} \leftarrow \text{nb\_neurons}$ 
23  $\text{predictions} \leftarrow \text{empty}$ 
24 Fit the lstm model
25  $\text{lstm\_model} = \text{fit\_lstm}(\text{train}, \text{epoch}, \text{neurons})$ 
26 Forecast the training dataset
27  $\text{lstm\_model.predict}(\text{train})$ 
28 Walk-forward validation on the test data
29 foreach  $i$  in  $\text{range}(\text{length}(\text{test}))$  do
30 |    $X_{\text{test}}[i]$ 
31 |    $\hat{y} \leftarrow \text{forecast\_lstm}(\text{lstm\_model}, X)$ 
32 |    $\text{predictions.append}(\hat{y})$ 
33 |    $\text{expected} \leftarrow \text{test}[i]$ 
34 end
35  $\text{MSE} \leftarrow \text{mean\_squared\_error}(\text{expected}, \text{predictions})$ 
36 Return (RMSE  $\leftarrow \text{sqrt}(\text{MSE})$ )
```

[16]

3.4 Conclusion

Dans ce chapitre, nous nous sommes intéressés aux principales notions de deux approches différentes qu'on va les utiliser dans la modélisation du chiffre d'affaires. Nous avons présenté aussi une vue sur les modèles Concernant l'élément sinistralité, nous allons présenter une vue sur une méthode pour la modélisation d'une partie de cet élément dans le chapitre 5.

Chapitre 4

Application des modèles : modélisation du chiffre d'affaires

Resumé : Les produits de Zitouna Takaful sont classifiés en 6 branches :

- Automobile
- Maladie
- Risque Divers et Spéciaux (RDS)
- Family (assurance vie)
- Transport
- Incendie

Nous allons considérer que le chiffre d'affaires est la prime hors taxes. Dans une première étape, nous allons modéliser le chiffre d'affaires global. Ensuite, nous allons effectuer notre modélisation sur les six branches (comptant puis terme).

Nous allons étudier le chiffre global avec SARIMA et LSTM.

Vu que nous n'avons pas assez d'historique pour utiliser l'approche de deep learning pour toutes les branches, nous allons étudier seulement le chiffre d'affaires global avec les deux approches. Concernant l'étude par branche, nous allons utiliser le modèle SARIMA. Nous allons utiliser le langage Python. Pour l'affichage de certains plot, nous allons utiliser le langage R

4.1 Chiffre d'affaires global

4.2 SARIMA

Le but est de sélectionner le meilleur modèle SARIMA pour la série chiffre d'affaires. Nous allons utiliser les données mensuelles du Septembre 2012 jusqu'au Décembre 2019.

Nous commençons alors par importer les données et les représenter dans un graphique.

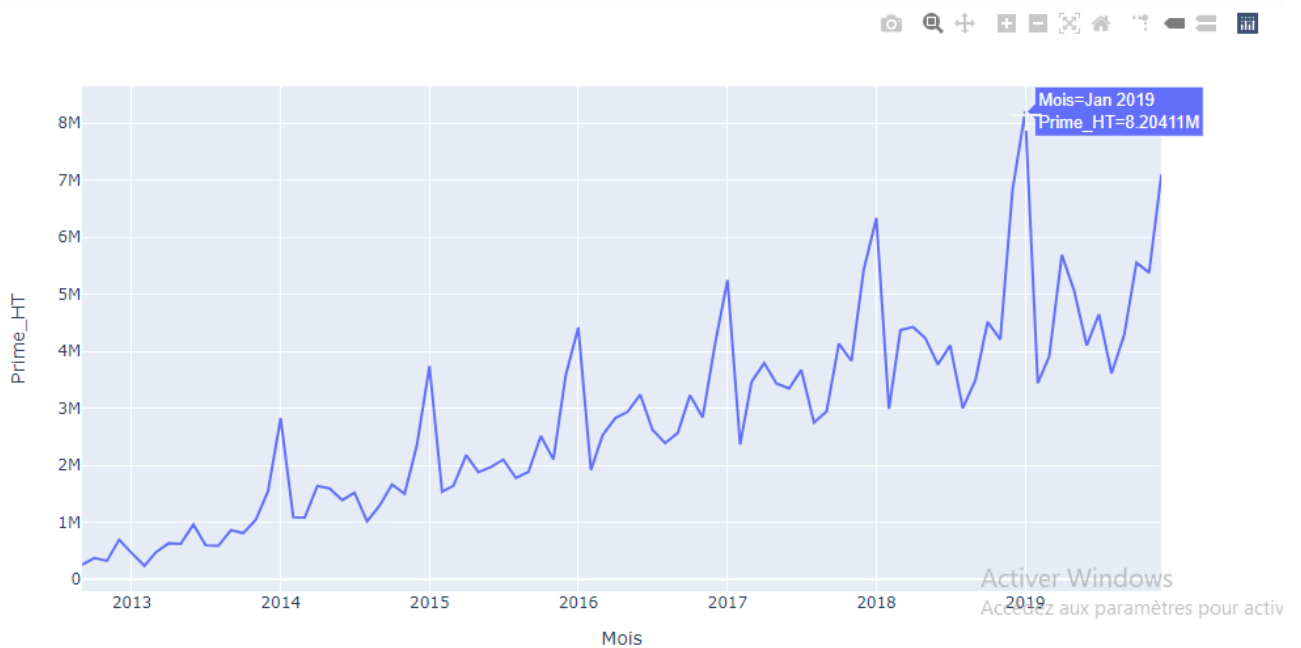


FIGURE 4.1 – Chiffre d'affaires global

Nous pouvons rapidement remarquer qu'il y a un comportement de tendance et une sorte de saisonnalité. Nous allons donc utiliser `autoarima`, du package `pyramid.arima`, commande pour vérifier rapidement si ce type de données peut être adapté en utilisant les modèles SARIMA.

Si l'AIC estimé est non négatif et si la probabilité logarithmique est négative, nous pouvons conclure que les modèles SARIMA peuvent être utiles pour ces données. Nous utiliserons également la commande `autoarima` pour avoir une idée de la valeurs de d , D et T

```

SARIMAX Results
=====
Dep. Variable:          y      No. Observations:      88
Model:          SARIMAX(2, 1, 0)x(1, 0, 0, 12)      Log Likelihood      -1292.788
Date:          Wed, 08 Jul 2020      AIC      2595.575
Time:          20:37:05      BIC      2607.905
Sample:          0      HQIC      2600.540
                  - 88
Covariance Type:          opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
intercept    1.335e+05    1.92e+05      0.697      0.486    -2.42e+05    5.09e+05
ar.L1         -0.4146      0.222     -1.865      0.062     -0.850      0.021
ar.L2         -0.4983      0.310     -1.609      0.108     -1.105      0.109
ar.S.L12       0.7274      0.132      5.506      0.000      0.468      0.986
sigma2        8.69e+11      0.041    2.13e+13      0.000    8.69e+11    8.69e+11
=====
Ljung-Box (Q):          59.90      Jarque-Bera (JB):          17.10
Prob(Q):          0.02      Prob(JB):          0.00
Heteroskedasticity (H):    1.83      Skew:          0.52
Prob(H) (two-sided):    0.11      Kurtosis:          4.91
=====

```

FIGURE 4.2 – "auto_arima"

Nous pouvons d'abord remarquer que nous pouvons continuer à utiliser les modèles SARIMA qui correspondent à ces données. Nous n'utiliserons pas les modèles estimés avec `auto.arima` car nous ne sommes pas sur que nous avons sélectionner le meilleur modèle en utilisant la commande `auto_arima`. Nous différencierons la série jusqu'à ce que nous obtenions la stationnarité. Une fois cette étape terminée, nous lancerons une sélection que nous fournisse le meilleur modèle en ajustant les données que nous pouvions utiliser pour faire les prédictions.

Nous allons tout d'abord représenter le processus $(1 - B)(1 - B^{12})X_t$

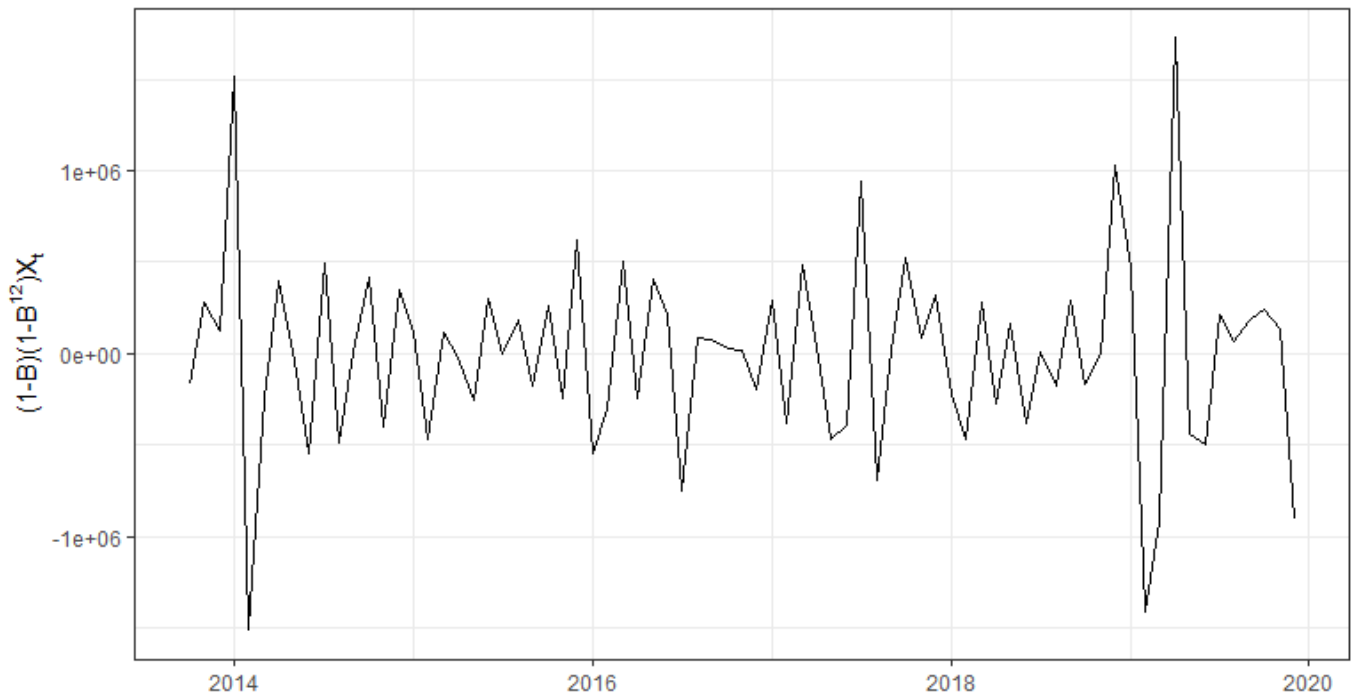


FIGURE 4.3 – $(1 - B)(1 - B^{12})X_t$

Nous pouvons remarquer, d'après le graphique précédent, que $(1 - B)(1 - B^{12})X_t$ a un comportement stationnaire. Nous pouvons confirmer cela en utilisant le test ADF et le test KPSS également. Les deux tests peuvent être effectués en utilisant respectivement les commandes `ur.df` et `ur.kpss`.

```
Entrée [61]: # ADF test on random numbers
series = df.Prime_HT.diff(12).diff(1).dropna()
result = adfuller(series, autolag='AIC')
print(f'ADF Statistic: {result[0]}')
print(f'p-value: {result[1]}')
for key, value in result[4].items():
    print('Critical Values:')
    print(f'    {key}, {value}')
```

ADF Statistic: -7.954048242257197
p-value: 3.071052705972105e-12
Critical Values:
1%, -3.526004646825607
Critical Values:
5%, -2.9032002348069774
Critical Values:
10%, -2.5889948363419957

FIGURE 4.4 – Augmented Dickey-Fuller

Tout d'abord, je remarque que la régression du test ADF est significative ; une petite valeur p , la valeur de la statistique du test est de -7,9540. Elle est très inférieure à la valeur critique correspondant à la valeur p de 1 %. Ensuite, la p -value du test d'hypothèse :

$$H_0 : \phi = 0 \text{ vs } H_1 : \phi < 0$$

Par conséquent, H_0 est rejeté et nous pouvons en conclure que $(1 - B)(1 - B^{12})$ peut être considéré comme un processus stationnaire.

Effectuons maintenant un test KPSS.

```
Entrée [63]: # KPSS test
from statsmodels.tsa.stattools import kpss
def kpss_test(series, **kw):
    statistic, p_value, n_lags, critical_values = kpss(series, **kw)
    # Format Output
    print(f'KPSS Statistic: {statistic}')
    print(f'p-value: {p_value}')
    print(f'num lags: {n_lags}')
    print('Critical Values:')
    for key, value in critical_values.items():
        print(f'    {key} : {value}')
    print(f'Result: The series is {"not " if p_value < 0.05 else ""}stationary')

kpss_test(series)

KPSS Statistic: 0.11905547697425072
p-value: 0.1
num lags: 12
Critical Values:
 10% : 0.347
  5% : 0.463
 2.5% : 0.574
  1% : 0.739
Result: The series is stationary
```

Activer Windows
Accédez aux paramètres pour activ

FIGURE 4.5 – KPSS test

La valeur de la statistique du test est égale à 0,119; elle est inférieure à la valeur critique correspondant à la valeur p de 10 %. Nous pouvons conclure également que le processus $(1 - B)(1 - B^{12})X_t$ peut être considéré comme un processus stationnaire. Dans la prochaine étape, nous déterminerons les valeurs possibles de p, P, q et Q. Nous utiliserons l'ACF pour déterminer les couples (q, Q) et le PACF pour déterminer les couples (p, P).

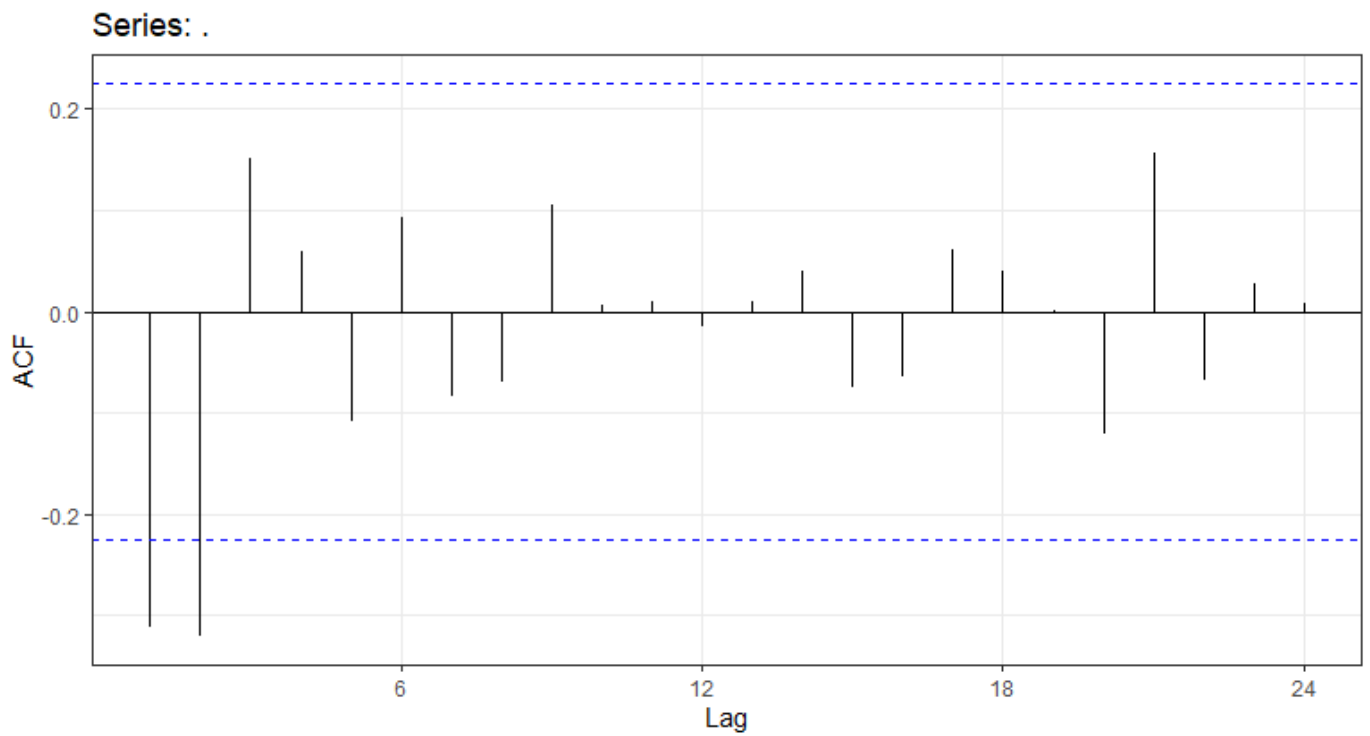


FIGURE 4.6 – ACF

Nous pouvons conclure du graphique précédent que les ACF sont égales à zéro à partir de l'ordre 2 et que ces ACF sont en décroissant jusqu'à zéro. Donc $q + 12Q \leq 2$. Rappelons que $q + 12Q$ est le degré de l'opérateur MA.

Dessignons maintenant le PACF.

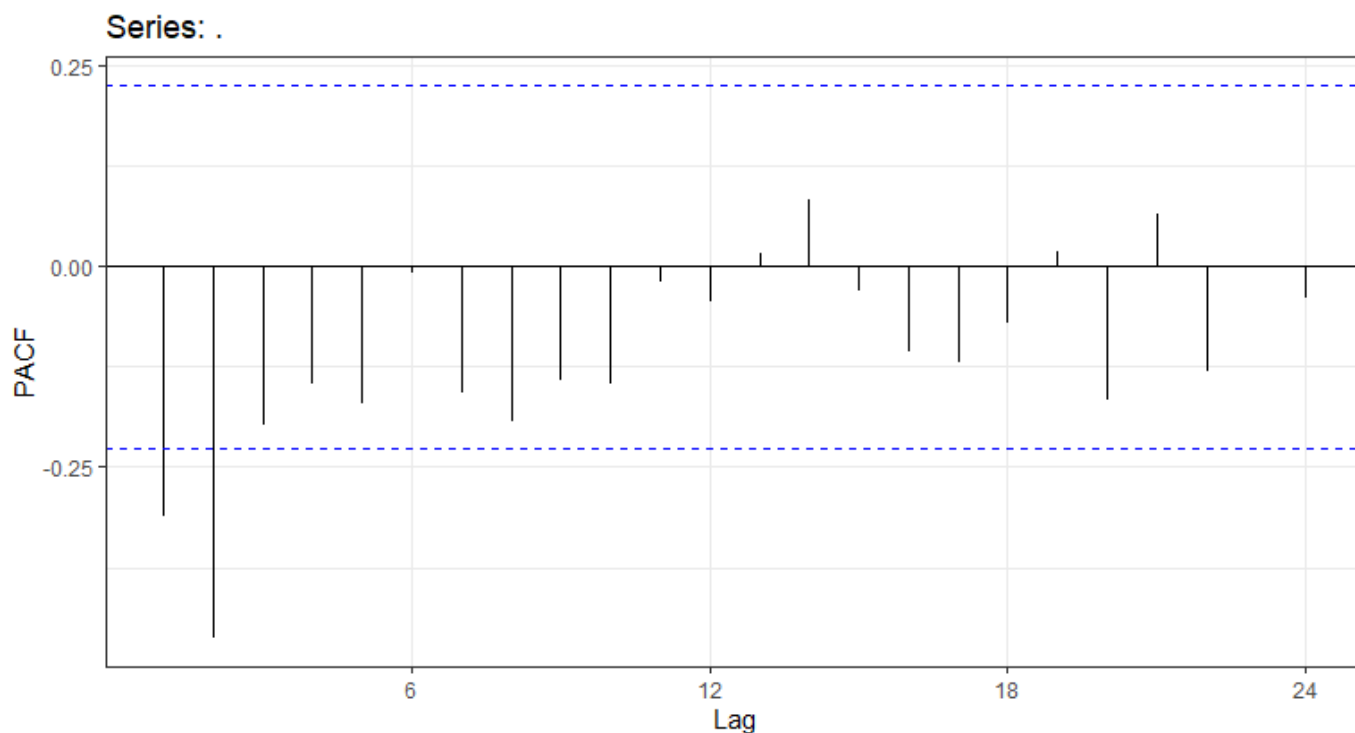


FIGURE 4.7 – PACF

Nous aurons alors la même conclusion que dans l'ACF et nous choisirons les couples $p + 12P$ tels que $p + 12P \leq 2$

Nous allons maintenant lancer la procédure de sélection. Nous allons d'abord construire une matrice avec les valeurs possibles de p , d , q , P , D , Q et T . Chaque ligne correspond à un cas. Au total, nous devons en estimer 9. Nous éliminerons ensuite les modèles dont l'AIC est négatif, puis nous éliminerons les modèles qui vont générer des résidus de bruit non blanc. Construction de la matrice des paramètres p , d , q , P , D , q , et T .

Nous allons maintenant vérifier les résidus, nous allons utiliser un test qui élimine les modèles dont résidus ne sont pas des bruits blancs. Nous allons utiliser un test de Portemanteau avec des retards de 1 à 10. Le modèle est rejeté si au moins une p-valeur du dernier test est inférieure au niveau de signification de 5%. Nous utiliserons la commande *boxpierce* du package *pypr.stattest.ljungbox*.

Ainsi, nous allons créer un objet qui nous dira si le test de bruit blanc a réussi ou non.

Puis, nous allons créer un vecteur *aic* qui contient l'AIC des modèles estimés.

Voici un tableau avec tous les résultats précédents :

p	d	q	P	D	Q	T	residuals	aic	model
2	1	2	0	1	0	12 y		2151.5758	ARIMA(2,1,2)(0,1,0)[12]
2	1	1	0	1	0	12 y		2150.2999	ARIMA(2,1,1)(0,1,0)[12]
2	1	0	0	1	0	12 y		2166.8711	ARIMA(2,1,0)(0,1,0)[12]
1	1	2	0	1	0	12 y		2151.7511	ARIMA(1,1,2)(0,1,0)[12]
1	1	1	0	1	0	12 y		2152.6368	ARIMA(1,1,1)(0,1,0)[12]
1	1	0	0	1	0	12 n		2183.8893	ARIMA(1,1,0)(0,1,0)[12]
0	1	2	0	1	0	12 y		2151.985	ARIMA(0,1,2)(0,1,0)[12]
0	1	1	0	1	0	12 y		2151.3291	ARIMA(0,1,1)(0,1,0)[12]
0	1	0	0	1	0	12 n		2189.7929	ARIMA(0,1,0)(0,1,0)[12]

FIGURE 4.8 – Les modèles

Notons que nous obtenons ici des modèles avec un AIC plus petit que le modèle sélectionné par `auto_arima`. Nous allons donc sélectionner les modèles dont l'AIC est inférieur à 2174 et nous les affichons tous.

```
"ARIMA(2,1,2)(0,1,0)[12]" "ARIMA(2,1,1)(0,1,0)[12]" "ARIMA(2,1,0)(0,1,0)[12]"
"ARIMA(1,1,2)(0,1,0)[12]" "ARIMA(1,1,1)(0,1,0)[12]" "ARIMA(0,1,2)(0,1,0)[12]"
"ARIMA(0,1,1)(0,1,0)[12]"
```

FIGURE 4.9 – Les modèles sélectionnés

Nous allons maintenant rejeter les modèles qui contiennent des coefficients non significatifs. Pour ce faire, nous utiliserons la fonction `t` de `scipy.stats` qui effectue un test pour chaque coefficient.

```
0 0 1 0 0 0 1
```

FIGURE 4.10 – Les modèles significatifs

Seuls le troisième et le dernier modèle qui ont alors tous leurs coefficients significatifs. Ces modèles sont :

SARIMAX Results

Dep. Variable:	Prime_HT	No. Observations:	75			
Model:	SARIMAX(2, 1, 0)x(0, 1, 0, 12)	Log Likelihood	-926.788			
Date:	Sat, 27 Jun 2020	AIC	1859.576			
Time:	18:45:32	BIC	1865.958			
Sample:	0	HQIC	1862.082			
	- 75					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.8801	0.122	-7.231	0.000	-1.119	-0.642
ar.L2	-0.5987	0.105	-5.691	0.000	-0.805	-0.393
sigma2	5.788e+11	5.38e-14	1.08e+25	0.000	5.79e+11	5.79e+11
Ljung-Box (Q):	46.18	Jarque-Bera (JB):	1.62			
Prob(Q):	0.23	Prob(JB):	0.44			
Heteroskedasticity (H):	1.80	Skew:	0.06			
Prob(H) (two-sided):	0.18	Kurtosis:	3.78			

FIGURE 4.11 – Modèle 1

SARIMAX Results

Dep. Variable:	Prime_HT		No. Observations:	75		
Model:	SARIMAX(0, 1, 1)x(0, 1, [], 12)		Log Likelihood	-920.949		
Date:	Sat, 27 Jun 2020		AIC	1845.898		
Time:	18:46:15		BIC	1850.152		
Sample:	0		HQIC	1847.568		
			- 75			
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.9959	0.155	-6.427	0.000	-1.300	-0.692
sigma2	4.943e+11	3.06e-13	1.61e+24	0.000	4.94e+11	4.94e+11
Ljung-Box (Q):			42.86	Jarque-Bera (JB): 3.23		
Prob(Q):			0.35	Prob(JB): 0.20		
Heteroskedasticity (H):			1.93	Skew: 0.14		
Prob(H) (two-sided):			0.14	Kurtosis: 4.08		

FIGURE 4.12 – Modèle 2

Le meilleur modèle est alors **SARIMAX(0,1,1)(0,1,0)[12]** ; celui qui a le plus petit Aic.

4.2.1 Erreur de prédiction (année 2019)

Nous allons, maintenant, comparer les résultats de notre modèle aux observations de 2019.

	observed	forecast
Mois		
2019-01	8.204110e+06	7.011925e+06
2019-02	3.443802e+06	3.798541e+06
2019-03	3.909537e+06	5.139047e+06
2019-04	5.691191e+06	5.058083e+06
2019-05	5.063170e+06	4.936166e+06
2019-06	4.102143e+06	4.485938e+06
2019-07	4.647451e+06	4.779049e+06
2019-08	3.612876e+06	3.664046e+06
2019-09	4.289162e+06	4.146396e+06
2019-10	5.552703e+06	5.184747e+06
2019-11	5.379505e+06	4.918691e+06
2019-12	7.096554e+06	7.590380e+06
Total	6.099220e+07	6.071301e+07

FIGURE 4.13 – Forecast 2019

```
from statsmodels.tools.eval_measures import rmse
rmse_error = rmse(ts[len(ts)-12:len(ts)], Pred_2019)
mse_error = rmse_error**2
mean_value = df['Prime_HT'].mean()

print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

```
MSE Error: 354584969135.74506
RMSE Error: 595470.3763712726
Mean: 2837527.1066022743
```

FIGURE 4.14 – RMSE

Représentons maintenant les données, les valeurs ajustées par le modèle sélectionné et une prévision fournie par celui-ci.

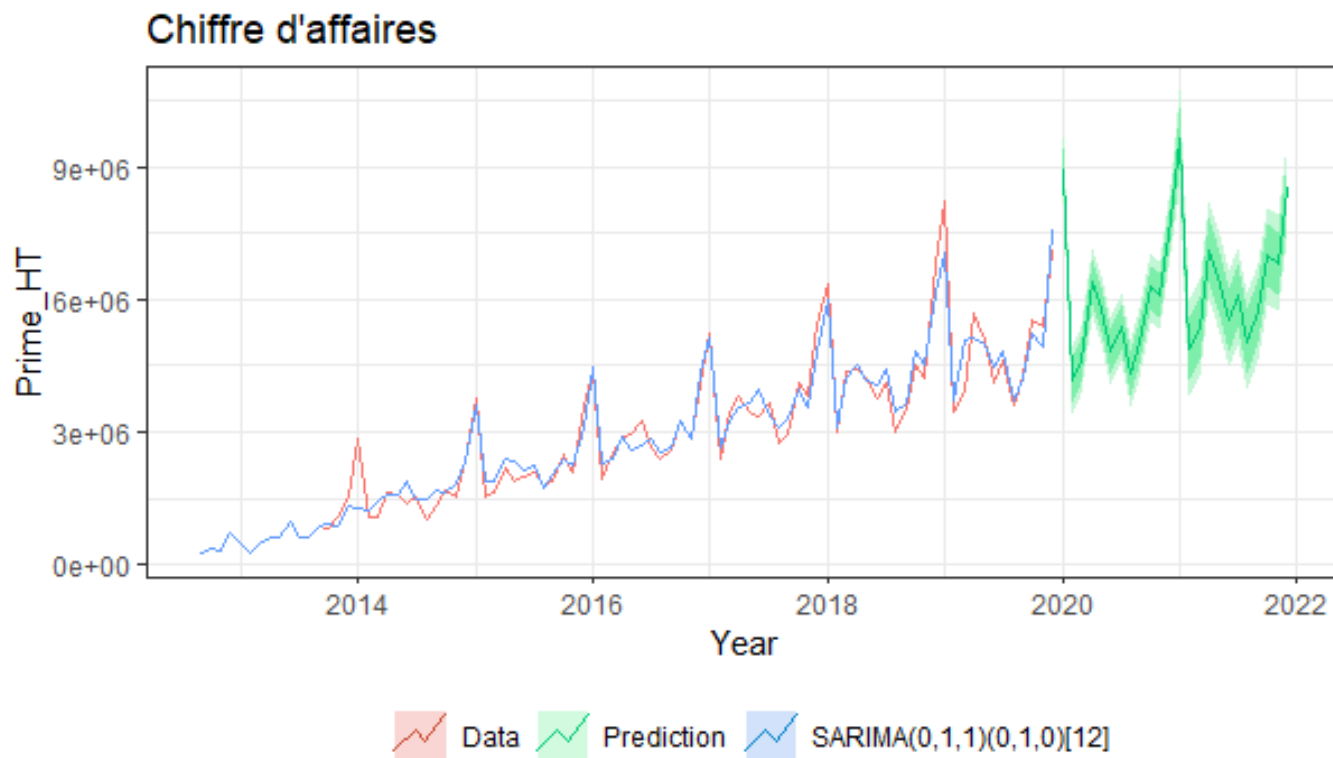


FIGURE 4.15 – Prédiction du chiffre d'affaires global]

	Mois	2020
0	janvier	8.912541e+06
1	fevrier	4.152233e+06
2	mars	4.617968e+06
3	avril	6.399622e+06
4	mai	5.771601e+06
5	juin	4.810575e+06
6	juillet	5.355882e+06
7	aout	4.321307e+06
8	septembre	4.997593e+06
9	octobre	6.261134e+06
10	novembre	6.087936e+06
11	decembre	7.804985e+06

FIGURE 4.16 – Chiffre d'affaires 2020 avec le modèle SARIMA

4.2.2 Analyse des résidus

```
results.plot_diagnostics(figsize=(7,5))  
plt.show()
```

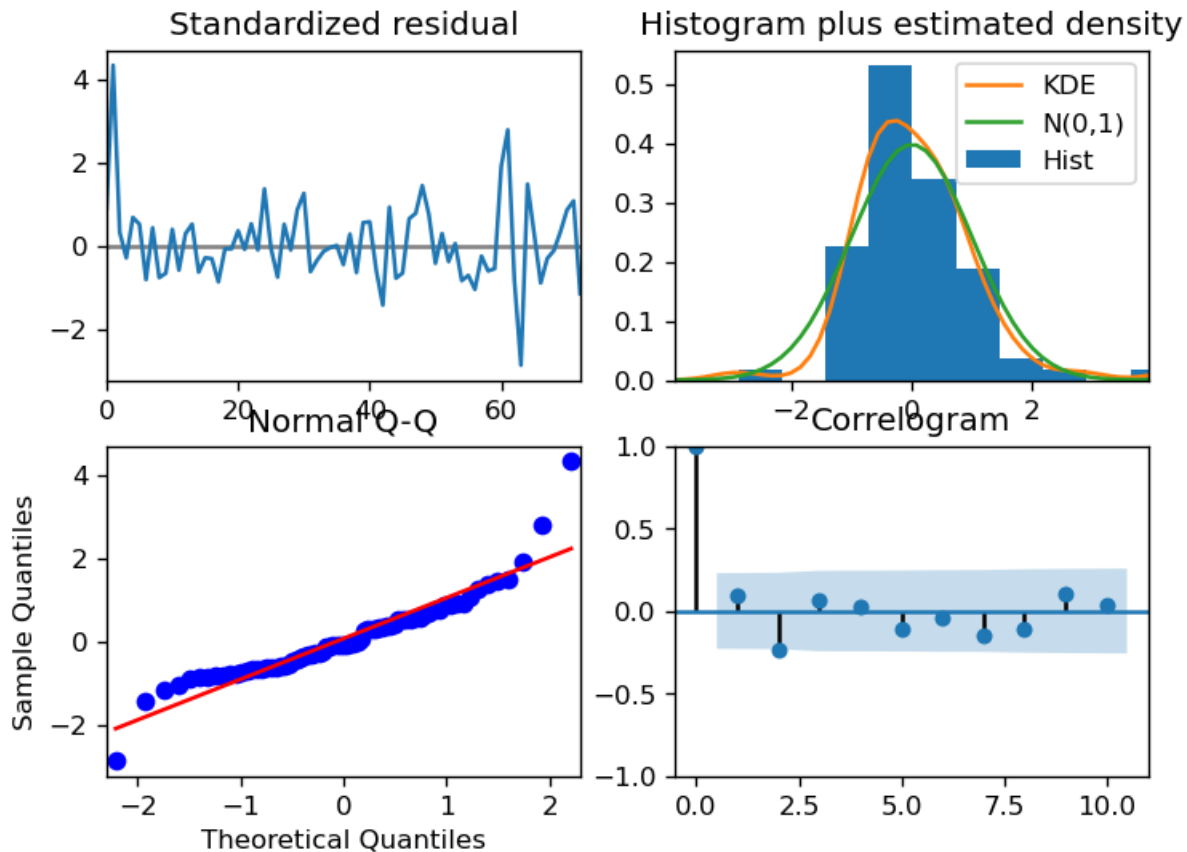


FIGURE 4.17 – Les résidus du modèle SARIMA

- L’histogramme des résidus montre que les résidus (et donc les termes d’erreur) sont normalement distribués.
- La courbe de probabilité normale des résidus est approximativement linéaire, ce qui nous permet de supposer que les termes d’erreur sont normalement distribués.
- Le graphique des résidus nous montre qu’il s’agit d’un processus stationnaire avec une moyenne constante 0 .
- Il est clair que le correlogramme présente une coupure sur la courbe d’autocorrelation après le 1 Lag, ce qui nous affirme la stationnarité des résidus.

4.3 LSTM

4.3.1 Preparation de données :

Tout d'abord, comme pour tout autre problème de machine learning, nous devons diviser nos données en "features" et "labels".

Dans ce cas, notre "feature" est en fait un certain nombre de valeurs dans la série, notre "label" étant la valeur suivante.

Nous appellerons ce nombre de valeurs qui seront traitées comme notre "feature", "l'input" (ou le "window size"), ou nous prenons une fenetre des données et formons un modèle de machine learning pour prédire la valeur suivante.

Ainsi, par exemple, si nous prenons nos données de série chronologiques, disons 12 mois à la fois, nous utiliserons 12 valeurs comme "feature" et la valeur suivante sera le "label". Ensuite, au fil du temps, nous formons un réseau de neurones pour faire correspondre les 12 "feautres" à l'unique "label".



FIGURE 4.18 – Préparation de données

4.3.2 Normalisation des données :

Pour le modèle LSTM, les données du train doivent être mises à l'échelle, ce qui est fait avec **MinMaxScaler** de la bibliothèque **sklearn** :

```
train, test = df[:-12], df[-12:]
scaler = MinMaxScaler()
scaler.fit_transform(train)
train = scaler.transform(train)
test = scaler.transform(test)
```

FIGURE 4.19 – Normalisation des données

4.3.3 Préparation du vecteur d'entrée pour le LSTM :

Nous allons considérer la prédiction de séries temporelles comme une tâche de prédiction de séquences. Les dates ne sont pas vraiment importantes pour l'algorithme "Machine Learning", il tente plutôt d'identifier la dépendance entre les données passées et futures. Il existe de nombreuses façons de transformer les données de séquences brutes pour les adapter à l'entraînement des réseaux de neurones LSTM. Nous allons utiliser le **TimeseriesGenerator** de Keras. Nous allons prédire 12 mois de chiffre d'affaires, pour cela nous utilisons la fenêtre de 12 mois du passé pour identifier la prédiction :

TimeseriesGenerator class

```
tf.keras.preprocessing.sequence.TimeseriesGenerator(
    data,
    targets,
    length,
    sampling_rate=1,
    stride=1,
    start_index=0,
    end_index=None,
    shuffle=False,
    reverse=False,
    batch_size=128,
)
```

FIGURE 4.20 – La classe TimeseriesGenerator

(Voir l'annexe pour plus d'information sur la fonction **TimeseriesGenerator**)

4.3.4 Architecture du modèle :

Le modèle d'architecture utilisé ci-dessous est assez explicite (il s'agit d'une couches LSTM contenant 300 neurones, la sortie de cette couche est transmise à la couche de sortie contenant un seul neurone).

```
from keras.models import Sequential
from keras.layers import Dense, LSTM, Dropout
n_input = 12
n_features = 1
model = Sequential()
model.add(LSTM(300,activation='relu',input_shape=(n_input,n_features),return_sequences=False))
model.add(Dropout(0.15))
model.add(Dense(1))
optimizer = keras.optimizers.Adam(learning_rate=0.001)
model.compile(optimizer=optimizer, loss=tf.keras.losses.Huber(),metrics=["mae"])
print('Model compiled')
model.summary()
```

Model compiled
Model: "sequential_22"

Layer (type)	Output Shape	Param #
lstm_25 (LSTM)	(None, 300)	362400
dropout_25 (Dropout)	(None, 300)	0
dense_22 (Dense)	(None, 1)	301
Total params: 362,701		
Trainable params: 362,701		
Non-trainable params: 0		

FIGURE 4.21 – Architecture du modèle LSTM

4.3.5 Prédiction du chiffre d'affaires de l'année 2019 :

Pour le "train" et le "test", nous allons utiliser les données de notre série du septembre 2012 jusqu'à décembre 2018, l'apprentissage du modèle sera appliqué sur cet ensemble de données. Cela permettra de générer des prévisions pour les dates ultérieures à la dernière date (décembre 2018).

```
generator = TimeseriesGenerator(train, train, length=n_input, batch_size = 6)
model.fit_generator(generator, epochs=70)
```

FIGURE 4.22 – Apprentissage du modèle LSTM

Une fois l'apprentissage est fini, nous pouvons générer une prévision pour les dates futures (2019). L'idée est de générer une prédiction pour une valeur future, en utilisant les 12 dernières valeurs. Nous ajoutons une nouvelle prédiction au tableau, nous supprimons la première entrée du même tableau et prédire l'étape suivante avec un tableau actualisé de 12 valeurs.

Créons un tableau de données avec les mois à venir, nous affecterons les données de prévision calculées à ces mois :

```

pred_list = []
batch = train[-n_input:].reshape(1, n_input,n_features)
for i in range(n_input):
    pred_list.append(model.predict(batch)[0])
    batch = np.append(batch[:,1:,:],[[pred_list[i]]],axis=1)
df_predict = pd.DataFrame(scaler.inverse_transform(pred_list), index = df[-n_input:].index, columns = ['Predictions'])
df_test = pd.concat([df,df_predict],axis = 1)

```

FIGURE 4.23 – Prediction du chiffre d'affaires 2019 avec LSTM

	Prime_HT	Predictions
Mois		
2019-01-01	8204110.115	8.744505e+06
2019-02-01	3443801.770	3.926225e+06
2019-03-01	3909536.873	4.403826e+06
2019-04-01	5691191.300	6.147205e+06
2019-05-01	5063170.012	5.483549e+06
2019-06-01	4102143.404	4.447427e+06
2019-07-01	4647450.997	4.987315e+06
2019-08-01	3612876.058	3.949193e+06
2019-09-01	4289161.690	4.642351e+06
2019-10-01	5552703.358	5.992004e+06
2019-11-01	5379504.584	5.868005e+06
2019-12-01	7096553.617	7.594705e+06

FIGURE 4.24 – Prediction vs valeurs réelles : année 2019

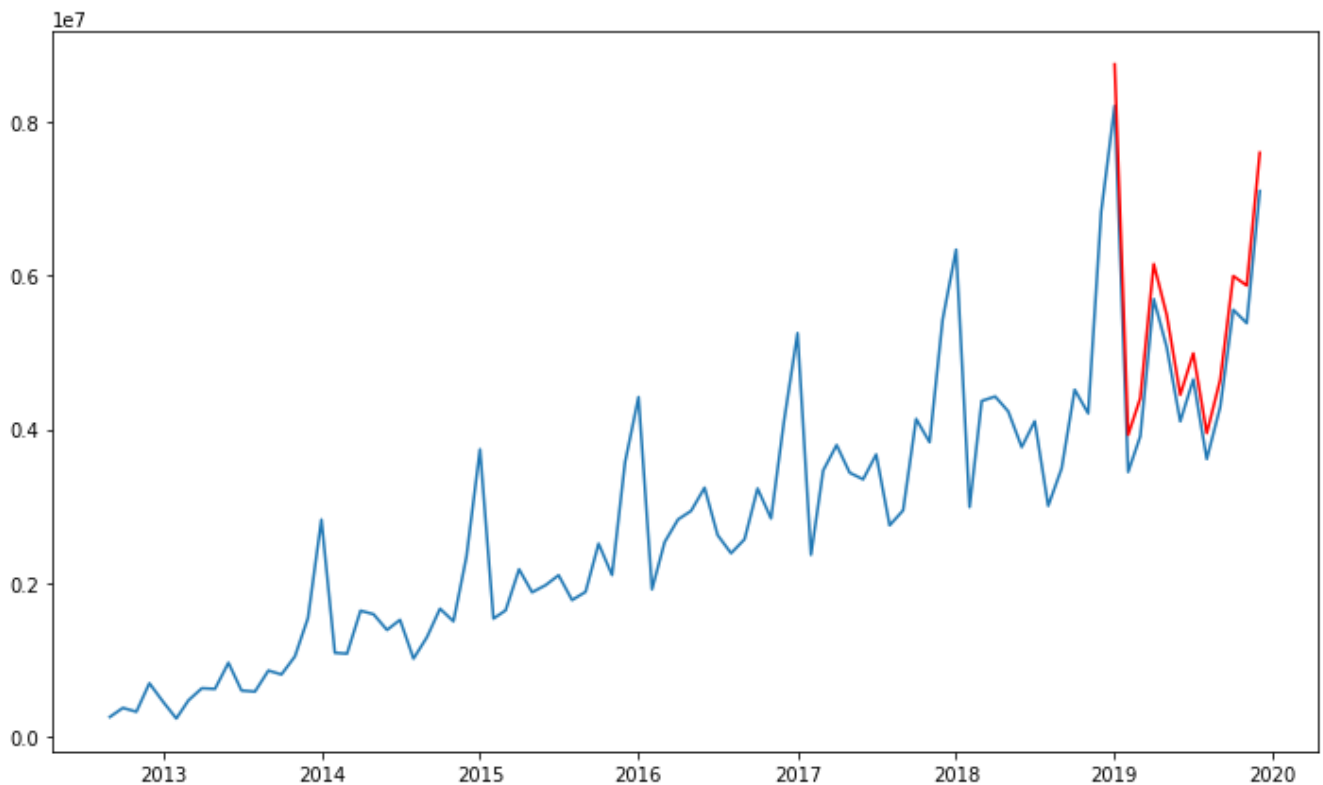


FIGURE 4.25 – Prediction LSTM : année 2019

Le tracé des prédictions par rapport aux chiffres réels 2019 se chevauchent presque à tel point que nous ne pouvons pas distinguer la courbe bleue et la courbe rouge dans le tracé ci-dessous.

Nous allons maintenant calculer le RMSE.

```
from statsmodels.tools.eval_measures import rmse
LSTM_rmse_error = rmse(test_data['Prime_HT'], test_data['LSTM_Predictions'])
LSTM_mse_error = LSTM_rmse_error**2
mean_value = df_test['Prime_HT'][-12:].mean()

print(f'MSE Error: {LSTM_mse_error}\nRMSE Error: {LSTM_rmse_error}\nMean: {mean_value}')
```

```
MSE Error: 192175010086.8951
RMSE Error: 438377.7025430184
Mean: 5082683.648166671
```

FIGURE 4.26 – RMSE du modèle LSTM

4.3.6 Prédiction du chiffre d'affaires de l'année 2020 :

Nous allons effectuer la même démarche que précédemment pour la prédiction du chiffre d'affaires 2020, mais maintenant l'apprentissage des données sera appliqué à l'ensemble de données (du septembre 2012 jusqu'au décembre 2019).

	Prime_HT	Predictions
2019-11-01	5379504.584	NaN
2019-12-01	7096553.617	NaN
2020-01-01	NaN	1.033274e+07
2020-02-01	NaN	3.832027e+06
2020-03-01	NaN	4.334573e+06
2020-04-01	NaN	6.296726e+06
2020-05-01	NaN	5.558880e+06
2020-06-01	NaN	4.435357e+06
2020-07-01	NaN	5.020899e+06
2020-08-01	NaN	4.093645e+06
2020-09-01	NaN	4.943045e+06
2020-10-01	NaN	6.307256e+06
2020-11-01	NaN	6.110589e+06
2020-12-01	NaN	7.585852e+06

FIGURE 4.27 – Prédiction du chiffre d'affaires 2020 avec LSTM

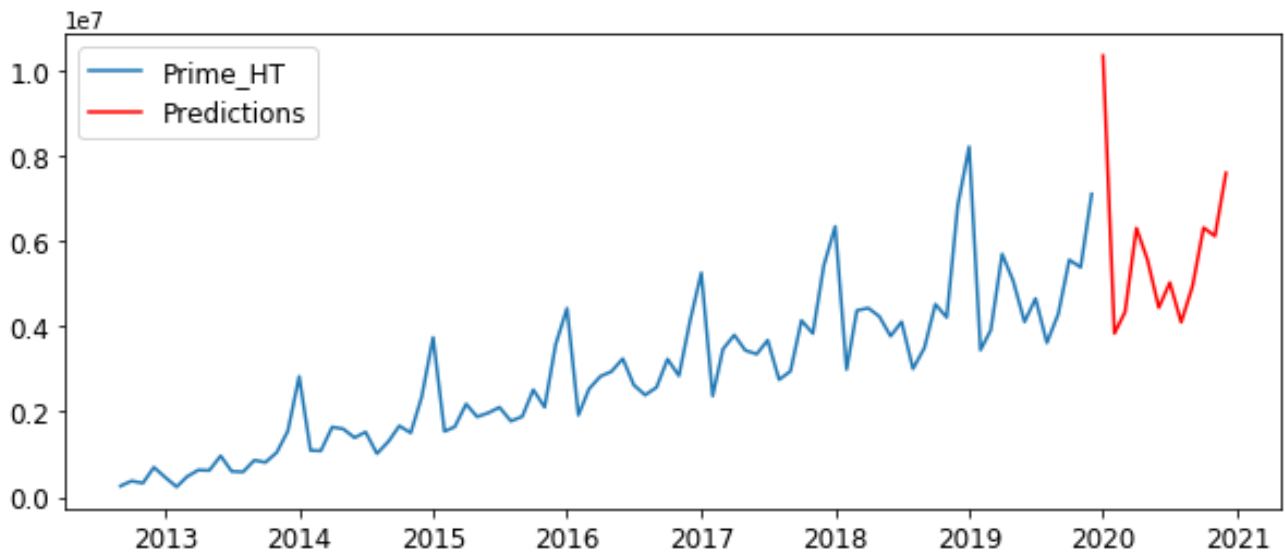


FIGURE 4.28 – LSTM 2020

4.4 Comparaison (SARIMA vs LSTM) :

Nous allons maintenant comparer les résultats des deux modèles avec les valeurs réelles de 2019.

```
rmse_errors = [sarima_rmse_error , LSTM_rmse_error]
mse_errors = [LSTM_mse_error, LSTM_mse_error]
errors = pd.DataFrame({"Models" : ["SARIMA", "LSTM"], "RMSE Errors" : rmse_errors, "MSE Errors" : mse_errors})
plt.figure(figsize=(16,9))
plt.plot_date(test_data.index, test_data["Prime_HT"], linestyle="-")
plt.plot_date(test_data.index, test_data["SARIMA_Predictions"], linestyle="-.")
plt.plot_date(test_data.index, test_data["LSTM_Predictions"], linestyle="--",color='r')
plt.legend()
plt.show()
```

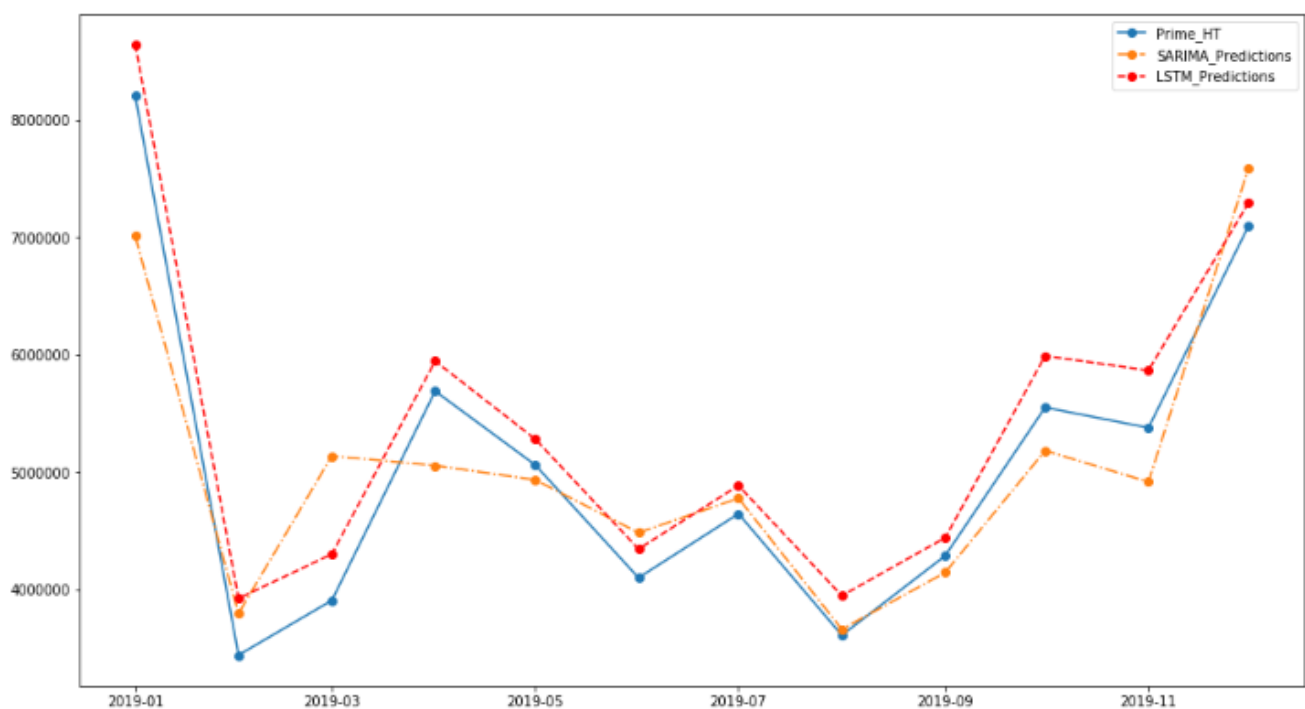


FIGURE 4.29 – Valeurs réelles&Sarima&LSTM : année 2019

```
: print(f"Mean: {test_data['Prime_HT'].mean()}")
errors
```

Mean: 5082683.648166671

	Models	RMSE Errors	MSE Errors
0	SARIMA	595470.320375	3.545849e+11
1	LSTM	438377.702543	1.921750e+11

FIGURE 4.30 – Comparaison Sarima et LSTM avec RMSE

: test_data			
	Prime_HT	SARIMA_Predictions	LSTM_Predictions
Mois			
2019-01-01	8204110.115	7011925	8744505
2019-02-01	3443801.770	3798541	3926225
2019-03-01	3909536.873	5139047	4403826
2019-04-01	5691191.300	5058083	6147205
2019-05-01	5063170.012	4936166	5483549
2019-06-01	4102143.404	4485938	4447427
2019-07-01	4647450.997	4779049	4987315
2019-08-01	3612876.058	3664046	3949193
2019-09-01	4289161.690	4146396	4642351
2019-10-01	5552703.358	5184747	5992004
2019-11-01	5379504.584	4918691	5868005
2019-12-01	7096553.617	7590380	7594705

FIGURE 4.31 – Valeurs réelles vs SARIMA vs LSTM : année 2019

	Réel	SARIMA	LSTM
Total 2019	60992203	60713009	66186310

TABLE 4.1 – Comparaison entre le réel et la prédiction 2019 SARIMA LSTM

D'après ces résultats, nous pouvons conclure que le modèle SARIMA a donné des résultats globales plus proches des valeurs réelles du chiffre d'affaires. Par contre, en se basant sur le RMSE, il est clair que le modèle LSTM est plus performant que le modèle SARIMA (la valeur de RMSE du modèle LSTM est plus petite que celle de SARIMA). En fait, la grande marge de différence entre la valeur réelle et la valeur prédite du mois janvier 2019 et mars 2019 (différence 12 milliards pour chaque mois) a fortement influencer la valeur du RMSE de SARIMA pour être plus élevée que celle de LSTM.

	SARIMA	LSTM
Total 2020	69493377	68851592

TABLE 4.2 – Comparaison entre LSTM et SARIMA 2020

Concernant la prédiction de 2020, les deux modèles ont donné une valeur de chiffre d'affaires d'environ 69 000 000 DT.

Dans ce cas, nous ne pouvant pas préciser laquelle de deux méthode qui s'adapte mieux aux données, en effet, chaque méthode a ses points forts et faibles.

Voici un tableau qui résume les points forts et faibles de deux approches :

Méthodes de prédiction des séries temporelles	Points forts	Points faibles
SARI MA	<ul style="list-style-type: none"> — Convergence garantie — Simple à manipuler — Faible coût en efforts et complexité. 	<ul style="list-style-type: none"> — Suppose des processus linéaires et stationnaires — Peut être coûteux en termes de calcul pour un modèle d'ordre élevé
LSTM	<ul style="list-style-type: none"> — Ne dépend pas du modèle — Ne dépend pas de la linéarité et la stationnarité du processus — Capable d'apprendre les dépendances à long terme — Peut détecter toutes les relations non linéaires possibles et complexes entre les entrées et les sorties 	<ul style="list-style-type: none"> — sélection de paramètres libres généralement calculés empiriquement — Pas de garantie de convergence vers la solution optimale — Pas de garantie de convergence vers la solution optimale — Peut être coûteux en termes de calcul (phase d'apprentissage) — Les réseaux neuronaux ont une nature de boîte noire. par conséquent, les erreurs au sein du réseau complexe sont difficiles à minimiser

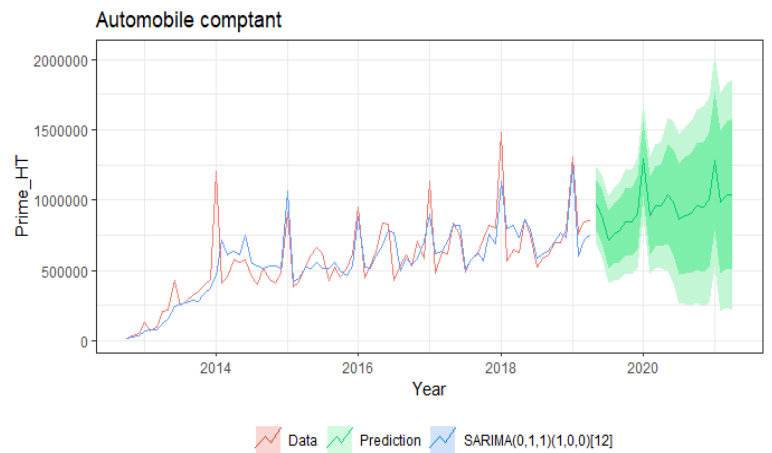
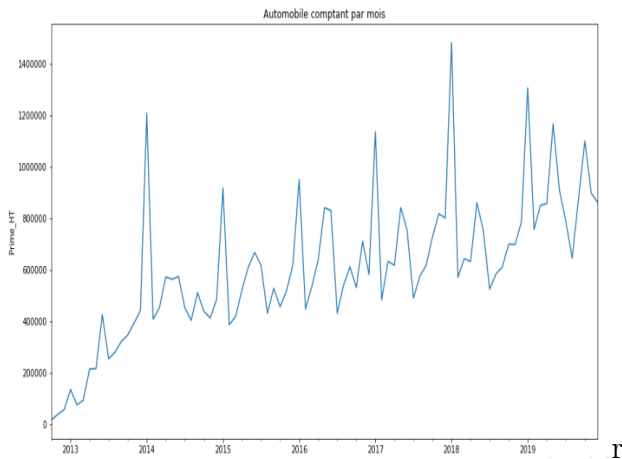
TABLE 4.3 – Points forts et faibles (LSTM & SARIMA)

La méthode d'apprentissage automatique exige un historique "un peu" riche de données. Du coup, nous allons continuer notre modélisation du chiffre d'affaires par branches avec le modèle SARIMA vu l'insuffisance d'historique pour certaines branches (comme la branche Maladie).

4.5 Chiffre d'affaires par branche

4.5.1 Branche Automobile

Comptant



	Mois	observed 2019	forecast 2019
0	janvier	1307100.013	1.318474e+06
1	fevrier	756899.763	5.617401e+05
2	mars	850338.167	6.977552e+05
3	avril	857715.548	7.482502e+05
4	mai	1167154.982	9.789958e+05
5	juin	907996.341	9.635420e+05
6	juillet	788808.788	7.539684e+05
7	aout	645463.520	8.159028e+05
8	septembre	876886.656	7.718476e+05
9	octobre	1100603.158	8.859782e+05
10	novembre	897748.057	9.681796e+05
11	decembre	862532.305	1.012406e+06

	Mois	2020
0	janvier	1.384595e+06
1	fevrier	9.310063e+05
2	mars	1.008038e+06
3	avril	1.014120e+06
4	mai	1.269224e+06
5	juin	1.055571e+06
6	juillet	9.573123e+05
7	aout	8.391374e+05
8	septembre	1.029924e+06
9	octobre	1.214358e+06
10	novembre	1.047123e+06
11	decembre	1.018091e+06

SARIMAX Results

Dep. Variable:	Prime_HT		No. Observations:	87		
Model:	SARIMAX(0, 1, 1)x(1, 0, [], 12)		Log Likelihood	-994.081		
Date:	Mon, 08 Jul 2020		AIC	1994.163		
Time:	20:58:00		BIC	2001.075		
Sample:	0		HQIC	1998.920		
				- 87		
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.6107	0.147	-4.149	0.000	-0.899	-0.322
ar.S.L12	0.8244	0.155	5.302	0.000	0.520	1.129
sigma2	4.717e+10	1.01e-12	4.67e+22	0.000	4.72e+10	4.72e+10
Ljung-Box (Q):	15.72	Jarque-Bera (JB):	208.23			
Prob(Q):	1.00	Prob(JB):	0.00			
Heteroskedasticity (H):	0.48	Skew:	1.61			
Prob(H) (two-sided):	0.07	Kurtosis:	10.58			

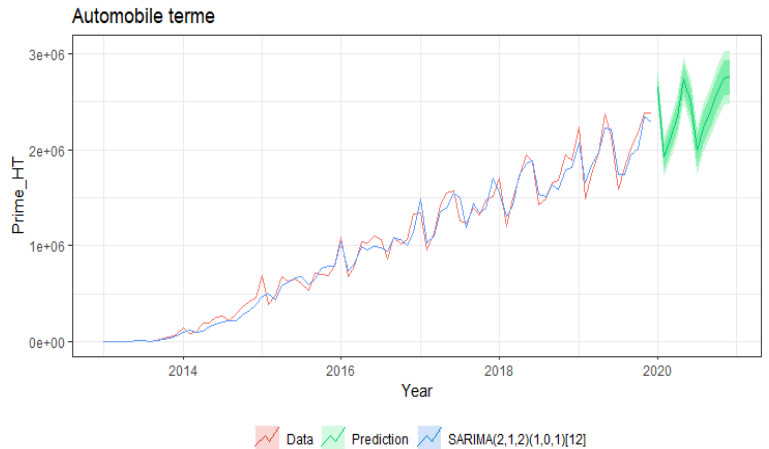
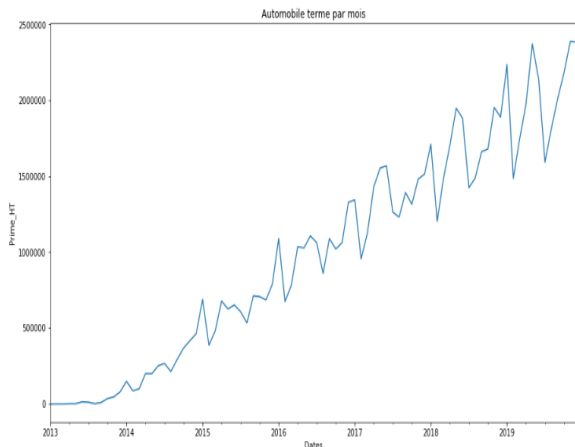
```
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Prime_HT'][-12:].mean()

print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

MSE Error: 18896110323.357533
 RMSE Error: 137463.12350356925
 Mean: 918270.6081666664

figure 1 :Automobile comptant

Terme



	Mois	observed 2019	forecast 2019
0	janvier	2235319.069	2.137433e+06
1	fevrier	1485633.657	1.588495e+06
2	mars	1744266.416	1.885065e+06
3	avril	1972630.548	2.048068e+06
4	mai	2371980.703	2.307416e+06
5	juin	2139136.414	2.243567e+06
6	juillet	1591879.397	1.676376e+06
7	aout	1813526.480	1.726069e+06
8	septembre	2010965.593	1.954706e+06
9	octobre	2181439.195	2.017933e+06
10	novembre	2388881.951	2.402932e+06
11	decembre	2381386.835	2.320819e+06

	Mois	2020
0	janvier	2.749015e+06
1	fevrier	1.875599e+06
2	mars	2.152054e+06
3	avril	2.410592e+06
4	mai	2.880499e+06
5	juin	2.590660e+06
6	juillet	1.971279e+06
7	aout	2.246985e+06
8	septembre	2.469286e+06
9	octobre	2.686108e+06
10	novembre	2.897844e+06
11	decembre	2.903247e+06

SARIMAX Results

Dep. Variable:	Prime_HT	No. Observations:	84
Model:	SARIMAX(2, 1, 2)x(1, 0, [1], 12)	Log Likelihood	-897.160
Date:	Mon, 06 Jul 2020	AIC	1808.320
Time:	21:20:13	BIC	1823.857
Sample:	0	HQIC	1814.476

Covariance Type:	opg
------------------	-----

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.1863	207.122	-0.001	0.999	-406.138	405.765
ar.L2	-0.0100	6.630	-0.002	0.999	-13.004	12.984
ma.L1	-0.4235	207.170	-0.002	0.998	-406.470	405.623
ma.L2	-0.0864	119.786	-0.001	0.999	-234.862	234.689
ar.S.L12	1.1248	0.379	2.968	0.003	0.382	1.868
ma.S.L12	0.1920	0.593	0.324	0.746	-0.970	1.355
sigma2	3.643e+10	2.52e-07	1.45e+17	0.000	3.64e+10	3.64e+10

Ljung-Box (Q):	51.03	Jarque-Bera (JB):	4.62
Prob(Q):	0.11	Prob(JB):	0.10
Heteroskedasticity (H):	1.39	Skew:	-0.61
Prob(H) (two-sided):	0.43	Kurtosis:	3.39

```
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Prime_HT'][-12:].mean()

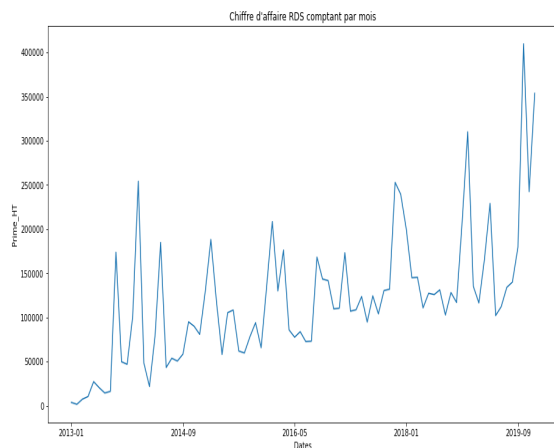
print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

```
MSE Error: 9108747588.076225
RMSE Error: 95439.7589481251
Mean: 2026420.521499995
```

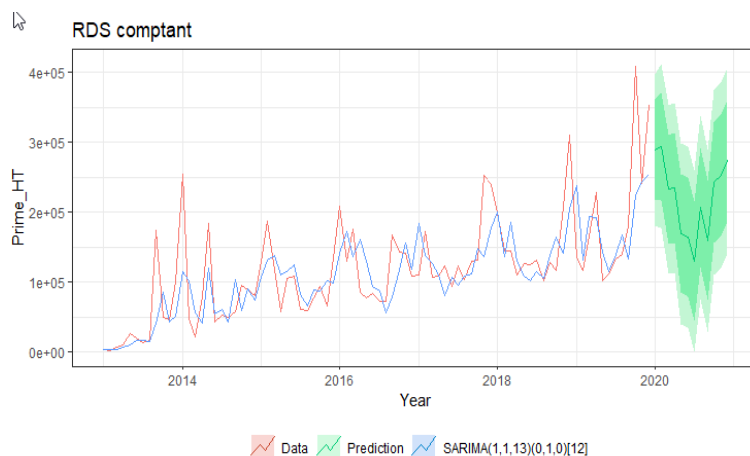
figure 2 : Automobile terme

4.5.2 Branche Risques divers et spéciaux (RDS)

Comptant



	Mois	observed 2019	forecast 2019
0	janvier	135241.219	224495.107158
1	fevrier	116351.594	158215.946534
2	mars	165960.280	157572.045248
3	avril	229014.959	174273.456420
4	mai	102023.140	178628.387435
5	juin	112184.197	133456.451643
6	juillet	134008.478	139518.430805
7	aout	139952.907	143725.521441
8	septembre	179609.381	128822.482940
9	octobre	409706.670	149816.305793
10	novembre	242383.187	289554.242775
11	decembre	353627.975	268321.008434



	Mois	2020
0	janvier	184112.965459
1	fevrier	183015.441632
2	mars	178719.082221
3	avril	292511.792229
4	mai	170019.379350
5	juin	214134.048241
6	juillet	171197.104767
7	aout	241419.432639
8	septembre	199638.975524
9	octobre	412986.094598
10	novembre	289507.319341
11	decembre	374654.981536

SARIMAX Results

Dep. Variable:	Total général		No. Observations:	84		
Model:	SARIMAX(1, 1, 13)x(0, 1, [], 12)		Log Likelihood	-703.128		
Date:	Sun, 05 Jul 2020		AIC	1438.258		
Time:	12:20:41		BIC	1468.901		
Sample:	0		HQIC	1448.168		
	- 84					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.0890	0.719	0.124	0.901	-1.320	1.497
ma.L1	-1.1572	0.577	-2.005	0.045	-2.289	-0.028
ma.L2	-0.0186	0.850	-0.022	0.983	-1.684	1.647
ma.L3	-0.2985	0.545	-0.548	0.584	-1.386	0.789
ma.L4	0.4019	0.614	0.655	0.513	-0.801	1.605
ma.L5	-0.1122	0.658	-0.170	0.865	-1.403	1.178
ma.L6	0.3277	0.591	0.555	0.579	-0.830	1.485
ma.L7	-0.2705	0.699	-0.387	0.699	-1.640	1.099
ma.L8	0.1902	0.621	0.306	0.759	-1.026	1.407
ma.L9	-0.5070	0.496	-1.021	0.307	-1.480	0.466
ma.L10	0.5248	0.636	0.825	0.410	-0.722	1.771
ma.L11	-0.3921	0.700	-0.560	0.575	-1.764	0.980
ma.L12	-0.3979	0.641	-0.621	0.535	-1.655	0.859
ma.L13	0.6764	0.488	1.387	0.166	-0.280	1.632
sigma2	3.529e+09	1.62e-10	2.17e+19	0.000	3.53e+09	3.53e+09
Ljung-Box (Q):	21.14	Jarque-Bera (JB):	4.78			
Prob(Q):	0.99	Prob(JB):	0.09			
Heteroskedasticity (H):	2.24	Skew:	0.54			
Prob(H) (two-sided):	0.09	Kurtosis:	3.92			

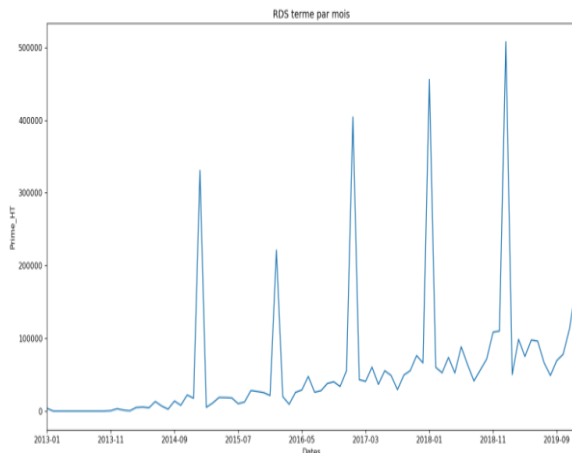
```
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Total général'][-12:].mean()

print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

MSE Error: 8231336464.398665
 RMSE Error: 90726.7130695181
 Mean: 135241.21899999998

figure 3 RDS comptant

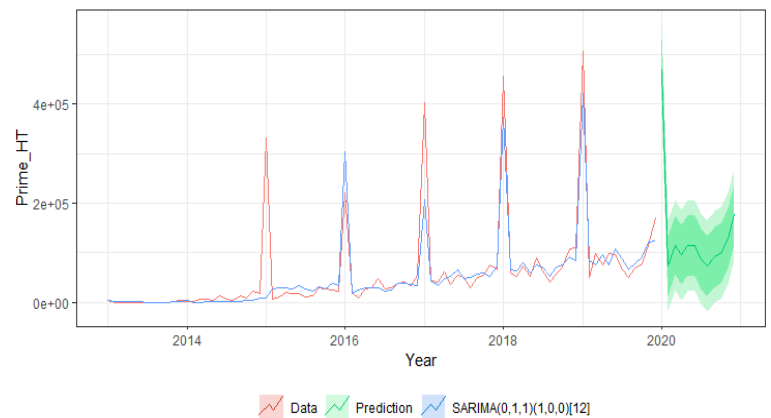
RDS terme



:

	Mois	observed 2019	forecast 2019
0	janvier	507671.048	494671.701867
1	fevrier	50359.710	80081.323971
2	mars	98746.693	71235.157566
3	avril	75236.587	94301.401225
4	mai	97725.212	71253.102042
5	juin	96207.833	109945.868933
6	juillet	66624.543	83632.871405
7	aout	49035.404	59554.161999
8	septembre	69606.719	75136.633527
9	octobre	78341.782	90769.727554
10	novembre	113587.206	129061.686737
11	decembre	170223.064	130104.085304

RDS terme



	Mois	2020
0	janvier	550110.221789
1	fevrier	68128.139624
2	mars	119125.473121
3	avril	94347.058723
4	mai	118048.885907
5	juin	116449.648300
6	juillet	85270.417595
7	aout	66732.390798
8	septembre	88413.474149
9	octobre	97619.770828
10	novembre	134766.592774
11	decembre	194457.806386

SARIMAX Results

Dep. Variable:	Total général	No. Observations:	72
Model:	SARIMAX(0, 1, 1)x(1, 0, [], 12)	Log Likelihood	-740.600
Date:	Fri, 19 Jun 2020	AIC	1487.200
Time:	15:25:41	BIC	1493.432
Sample:	0	HQIC	1489.633

- 72

Covariance Type:	opg
------------------	-----

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.9732	0.584	-1.667	0.095	-2.117	0.171
ar.S.L12	1.0539	0.491	2.145	0.032	0.091	2.017
sigma2	1.174e+10	7.86e-12	1.49e+21	0.000	1.17e+10	1.17e+10

Ljung-Box (Q):	22.06	Jarque-Bera (JB):	708.36
Prob(Q):	0.99	Prob(JB):	0.00
Heteroskedasticity (H):	0.33	Skew:	0.82
Prob(H) (two-sided):	0.02	Kurtosis:	19.90

```
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Total général'][-12:].mean()

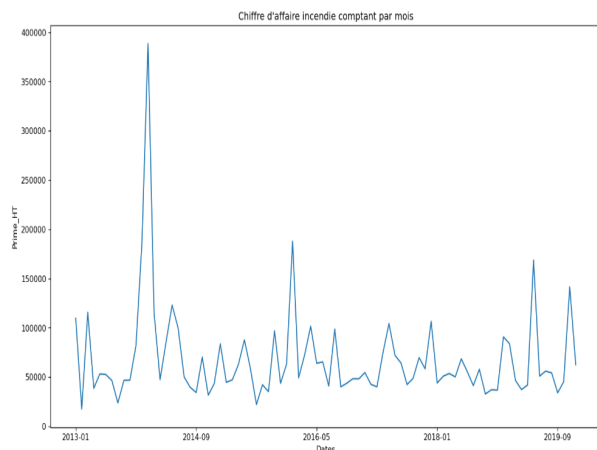
print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

MSE Error: 458014049.43933296
RMSE Error: 21401.262800109085
Mean: 122780.48341666667

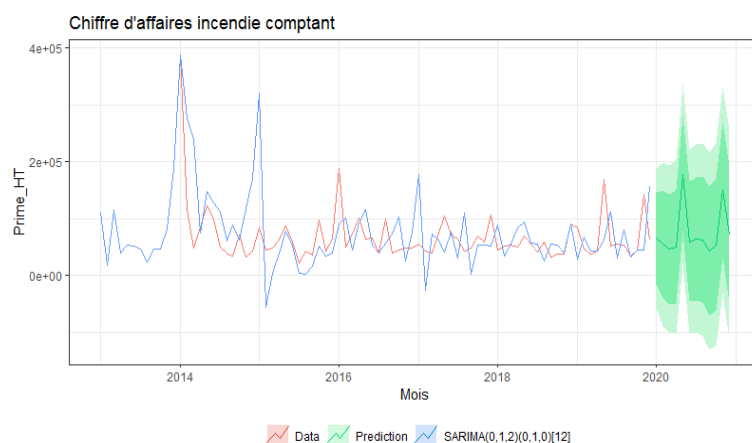
figure 4 : RDS terme

4.5.3 Branche Incendie

Comptant



	Mois	observed 2019	forecast 2019
0	janvier	83958.200	28187.826610
1	fevrier	46528.623	63459.463809
2	mars	37331.667	46414.636680
3	avril	41981.257	41705.421979
4	mai	168744.881	62011.763465
5	juin	50944.356	103574.363419
6	juillet	56123.143	41159.166900
7	aout	54278.304	75979.555304
8	septembre	34009.176	36685.136999
9	octobre	45385.816	43970.185175
10	novembre	141733.977	44812.350443
11	decembre	62492.279	148410.286166



	Mois	2020
0	janvier	78030.629063
1	fevrier	57782.706393
2	mars	48585.750393
3	avril	53235.340393
4	mai	179998.964393
5	juin	62198.439393
6	juillet	67377.226393
7	aout	65532.387393
8	septembre	45263.259393
9	octobre	56639.899393
10	novembre	152988.060393
11	decembre	73746.362393

SARIMAX Results

Dep. Variable:	Total général	No. Observations:	84
Model:	SARIMAX(0, 1, 2)x(0, 1, [], 12)	Log Likelihood	-838.355
Date:	Mon, 06 Jul 2020	AIC	1682.709
Time:	12:43:07	BIC	1689.368
Sample:	0	HQIC	1685.347

- 84

Covariance Type:	opg
------------------	-----

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.4869	0.095	-5.102	0.000	-0.674	-0.300
ma.L2	-0.2000	0.138	-1.450	0.147	-0.470	0.070
sigma2	3.146e+09	1.39e+11	2.26e+20	0.000	3.15e+09	3.15e+09

Ljung-Box (Q):	28.94	Jarque-Bera (JB):	65.44
Prob(Q):	0.90	Prob(JB):	0.00
Heteroskedasticity (H):	0.34	Skew:	-1.22
Prob(H) (two-sided):	0.01	Kurtosis:	7.14

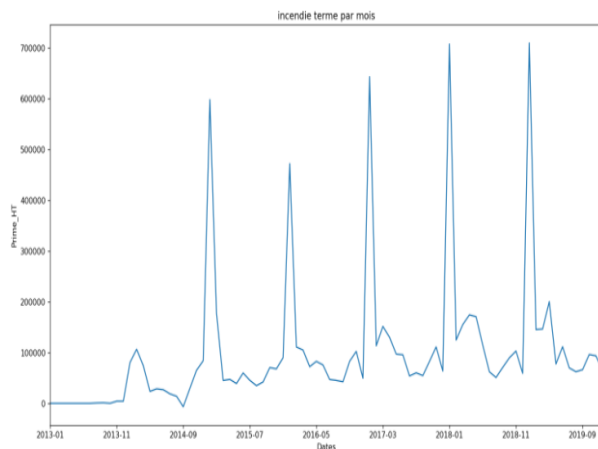
```
rmse_error = rmse(ts[36:48], output)
mse_error = rmse_error**2
mean_value = df['Total général'].mean()

print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

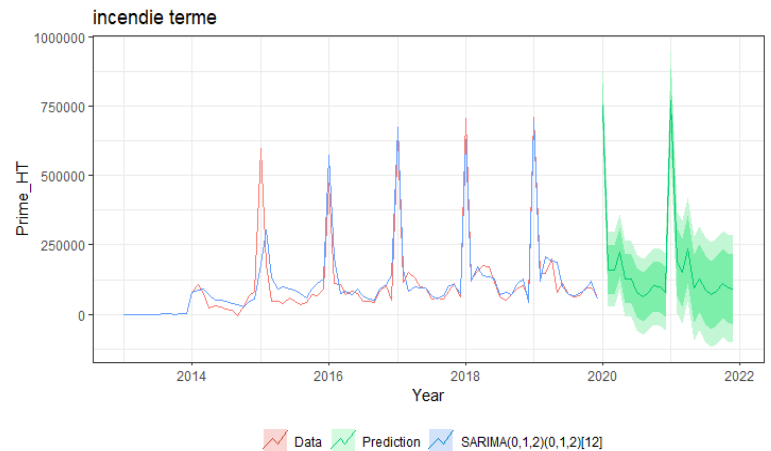
MSE Error: 3505112665.3087616
 RMSE Error: 59203.99197105514
 Mean: 68128.09923809522

figure 5 : incendie comptant

Terme



	Mois	observed 2019	forecast 2019
0	janvier	710064.844	714039.050851
1	fevrier	145284.538	124910.488663
2	mars	146508.997	178222.803236
3	avril	200521.318	189673.663069
4	mai	77565.385	170617.805207
5	juin	111761.720	124266.956837
6	juillet	69871.882	60119.937273
7	aout	62279.472	53721.936200
8	septembre	66343.597	68003.936121
9	octobre	96279.619	87992.252272
10	novembre	93781.572	108117.743034
11	decembre	65369.605	58138.343731



	Mois	2020
0	janvier	705778.557238
1	fevrier	153183.152367
2	mars	154381.194172
3	avril	207228.221722
4	mai	86925.012039
5	juin	120383.575278
6	juillet	79397.493179
7	aout	71968.886279
8	septembre	75945.329466
9	octobre	105235.494193
10	novembre	102791.341522
11	decembre	74992.350940

SARIMAX Results

Dep. Variable:	Total général	No. Observations:	84
Model:	SARIMAX(0, 1, 2)x(0, 1, 2, 12)	Log Likelihood	-539.470
Date:	Mon, 13 Jul 2020	AIC	1088.939
Time:	18:06:06	BIC	1097.860
Sample:	0	HQIC	1092.248

- 84

Covariance Type:	opg
------------------	-----

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.9838	1.495	-0.658	0.511	-3.915	1.947
ma.L2	0.1105	1.247	0.089	0.929	-2.334	2.555
ma.S.L12	-0.2128	0.673	-0.316	0.752	-1.533	1.107
ma.S.L24	0.2205	0.385	0.573	0.567	-0.534	0.975
sigma2	6.188e+09	1.02e-10	6.08e+19	0.000	6.19e+09	6.19e+09

Ljung-Box (Q):	31.04	Jarque-Bera (JB):	5.65
Prob(Q):	0.84	Prob(JB):	0.06
Heteroskedasticity (H):	0.89	Skew:	-0.23
Prob(H) (two-sided):	0.82	Kurtosis:	4.69

```
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Total général'][-12:].mean()

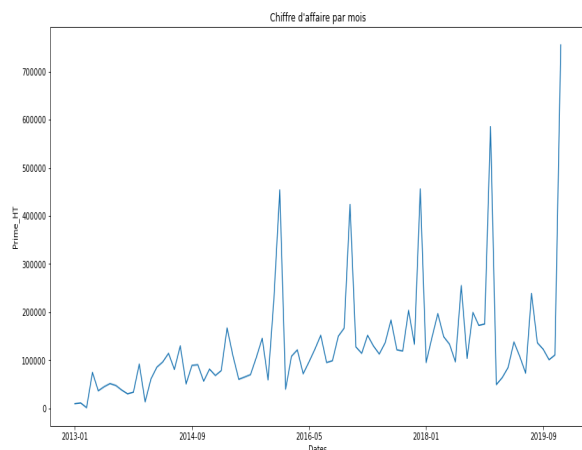
print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

MSE Error: 905587761.8565918
 RMSE Error: 30092.985259967012
 Mean: 153802.7124166665

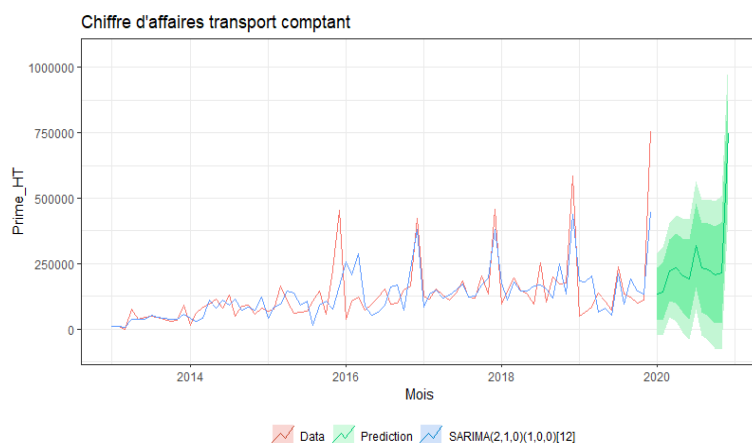
figure 6 : incendie terme

4.5.4 Branche Transport

Comptant



	Mois	observed 2019	forecast 2019
0	janvier	49352.552	132232.306713
1	fevrier	63756.161	164621.389084
2	mars	84804.017	191032.724417
3	avril	137884.768	65409.128046
4	mai	107567.075	74122.362316
5	juin	73122.288	38344.983845
6	juillet	238776.738	247472.325395
7	aout	136175.972	75865.951206
8	septembre	122742.307	208235.192523
9	octobre	100926.934	136519.645892
10	novembre	110737.901	140931.202580
11	decembre	755200.353	551862.224798



	Mois	2020
0	janvier	59456.387479
1	fevrier	67729.361054
2	mars	133522.135894
3	avril	170350.306612
4	mai	133967.206545
5	juin	105585.415297
6	juillet	282311.232199
7	aout	169388.910479
8	septembre	156679.829384
9	octobre	132186.660734
10	novembre	142597.655188
11	decembre	845729.542983

SARIMAX Results

Dep. Variable:	Total général	No. Observations:	84
Model:	SARIMAX(2, 1, 0)x(1, 0, 0, 12)	Log Likelihood	-886.971
Date:	Mon, 06 Jul 2020	AIC	1781.942
Time:	12:11:22	BIC	1790.878
Sample:	0	HQIC	1785.487
			- 84

Covariance Type:

opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.5519	0.174	-3.176	0.001	-0.892	-0.211
ar.L2	-0.3488	0.156	-2.231	0.026	-0.655	-0.042
ar.S.L12	1.0904	0.164	6.635	0.000	0.768	1.412
sigma2	1.538e+10	1.72e-12	8.94e+21	0.000	1.54e+10	1.54e+10

Ljung-Box (Q):	37.85	Jarque-Bera (JB):	12.32
Prob(Q):	0.57	Prob(JB):	0.00
Heteroskedasticity (H):	0.62	Skew:	0.45
Prob(H) (two-sided):	0.27	Kurtosis:	4.87

```
from statsmodels.tools.eval_measures import rmse
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Total général'][-12:].mean()

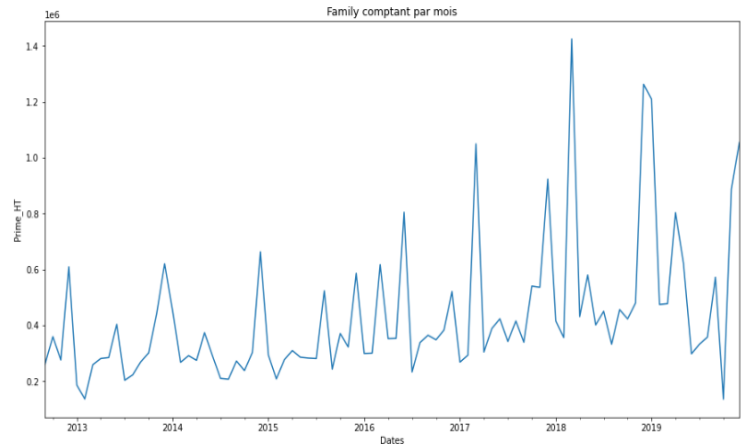
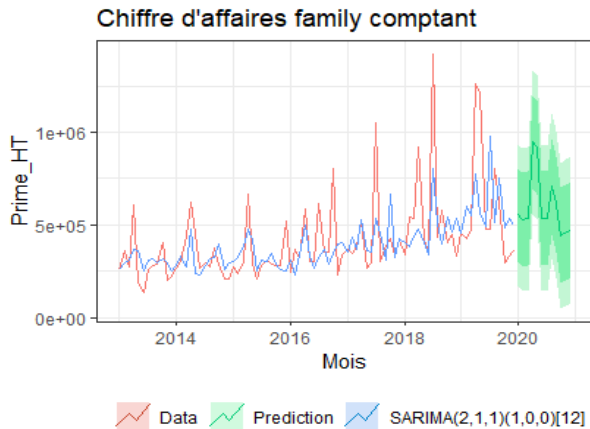
print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

MSE Error: 7537918886.5586605
 RMSE Error: 86821.1430848423
 Mean: 165087.2555

figure 7 : transport comptant

4.5.5 Branche Family (assurance vie)

Comptant



	Mois	observed 2019	forecast 2019
0	janvier	1209389.261	5.101006e+05
1	fevrier	474503.487	4.312562e+05
2	mars	477298.003	1.123174e+06
3	avril	803515.908	4.471468e+05
4	mai	621930.068	6.731658e+05
5	juin	297826.068	4.385687e+05
6	juillet	332300.142	4.888716e+05
7	août	357676.531	4.236554e+05
8	septembre	572207.223	5.169149e+05
9	octobre	135472.416	4.904094e+05
10	novembre	887570.147	4.963198e+05
11	decembre	1054417.614	1.127369e+06

	Mois	2020
0	janvier	1.003995e+06
1	fevrier	5.183782e+05
2	mars	5.233046e+05
3	avril	7.511914e+05
4	mai	6.222664e+05
5	juin	3.931571e+05
6	juillet	4.176045e+05
7	août	4.355427e+05
8	septembre	5.873519e+05
9	octobre	2.782934e+05
10	novembre	8.105239e+05
11	decembre	9.285951e+05

```
import statsmodels.api as sm
import numpy as np
#ts=np.array(series.Prime_HT)
model = sm.tsa.statespace.SARIMAX(ts,
                                  order=(2, 1, 1),
                                  seasonal_order=(1, 0, 0, 12),
                                  enforce_stationarity=False,
                                  enforce_invertibility=False)
results = model.fit()
print(results.summary().tables[1])
results.summary()
```

```
-----
coef    std err          z      P>|z|    [0.025    0.975]
-----
ar.L1      0.0528      0.100      0.530      0.596     -0.143      0.248
ar.L2     -0.0906      0.187     -0.486      0.627     -0.456      0.275
ma.L1     -1.0393      0.063    -16.505      0.000     -1.163     -0.916
ar.S.L12   0.7077      0.145      4.865      0.000      0.423      0.993
sigma2     6.079e+10    1.22e-12    4.98e+22    0.000    6.08e+10    6.08e+10
-----
```

SARIMAX Results

Dep. Variable:	Prime_HT	No. Observations:	88
Model:	SARIMAX(2, 1, 1)x(1, 0, 0, 12)	Log Likelihood:	-907.974
Date:	Mon, 08 Jul 2020	AIC:	2005.947
Time:	09:25:13	BIC:	2017.400
Sample:		HQIC:	2010.511
			- 88

Covariance Type: opg

```

coef    std err          z      P>|z|    [0.025    0.975]
-----
ar.L1      0.0528      0.100      0.530      0.596     -0.143      0.248
ar.L2     -0.0906      0.187     -0.486      0.627     -0.456      0.275
ma.L1     -1.0393      0.063    -16.505      0.000     -1.163     -0.916
ar.S.L12   0.7077      0.145      4.865      0.000      0.423      0.993
sigma2     6.079e+10    1.22e-12    4.98e+22    0.000    6.08e+10    6.08e+10

Ljung-Box (Q): 35.30 Jarque-Bera (JB): 27.83
Prob(Q): 0.68 Prob(JB): 0.00
Heteroskedasticity (H): 6.05 Skew: 0.85
Prob(H) (two-sided): 0.00 Kurtosis: 5.50
```

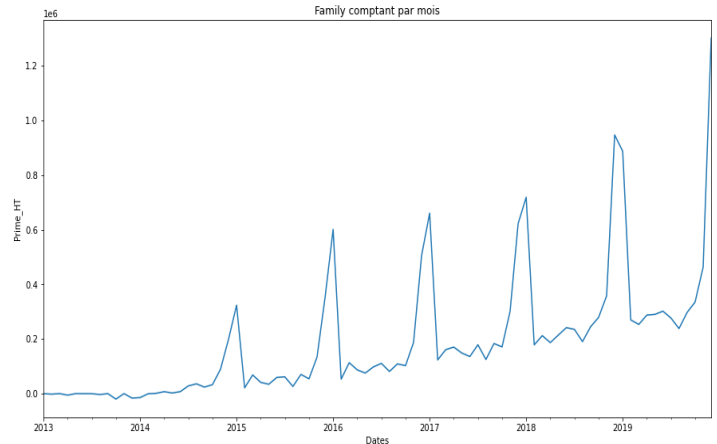
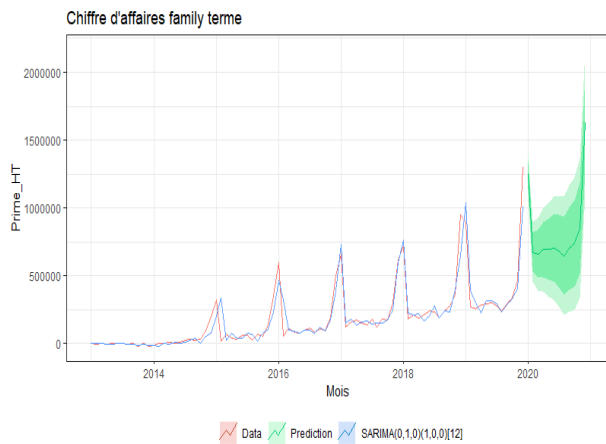
```
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Prime_HT'][-12:].mean()

print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

MSE Error: 259214111234.338
 RMSE Error: 509130.7408066596
 Mean: 602008.9056666667

figure 8 : family comptant

Terme



	Mois	observed 2019	forecast 2019
0	janvier	887031.321	1.061036e+06
1	fevrier	269695.942	2.457394e+05
2	mars	253214.500	3.098966e+05
3	avril	287337.791	2.229336e+05
4	mai	289816.580	3.202338e+05
5	juin	301551.614	3.220586e+05
6	juillet	276656.129	2.940437e+05
7	août	238148.239	2.233224e+05
8	septembre	296798.021	3.023230e+05
9	octobre	335186.664	3.376506e+05
10	novembre	462882.708	4.289518e+05
11	decembre	1301826.565	1.161949e+06

	Mois	2020
0	janvier	1.230984e+06
1	fevrier	4.982291e+05
2	mars	4.786662e+05
3	avril	5.191693e+05
4	mai	5.221116e+05
5	juin	5.360406e+05
6	juillet	5.064906e+05
7	août	4.607831e+05
8	septembre	5.303983e+05
9	octobre	5.759642e+05
10	novembre	7.275348e+05
11	decembre	1.723331e+06

SARIMAX Results

Dep. Variable:	Prime_HT	No. Observations:	84
Model:	SARIMAX(0, 1, 0)x(1, 0, 0, 12)	Log Likelihood	-896.748
Date:	Mon, 06 Jul 2020	AIC	1797.495
Time:	10:42:49	BIC	1802.021
Sample:	0	HQIC	1799.295

- 84

Covariance Type:	opg
------------------	-----

	coef	std err	z	P> z	[0.025	0.975]
ar.S.L12	1.1870	0.035	33.536	0.000	1.118	1.256
sigma2	5.47e+09	4.78e-13	1.14e+22	0.000	5.47e+09	5.47e+09

Ljung-Box (Q):	33.50	Jarque-Bera (JB):	84.41
Prob(Q):	0.76	Prob(JB):	0.00
Heteroskedasticity (H):	0.84	Skew:	-1.09
Prob(H) (two-sided):	0.68	Kurtosis:	7.87

```
rmse_error = rmse(ts[len(ts)-12:len(ts)], output)
mse_error = rmse_error**2
mean_value = df['Prime_HT'][-12:].mean()

print(f'MSE Error: {mse_error}\nRMSE Error: {rmse_error}\nMean: {mean_value}')
```

MSE Error: 5069501248.154512
 RMSE Error: 71200.43011214548
 Mean: 433345.5061666665

figure 9 : family terme

Voici un tableau récapitulatif des résultats de la modélisation SARIMA par branche :

Branche	Automobile		RDS	
	C	T	C	T
Total 2019	11019247.298	24317046.258	2320063.987	1473365.801
Prédiction 2020	12768498.96	29833168.07	2911916.617	1733469.88
Taux d'évolution	15.87%	22.68%	25.51%	17.65%

Branche	Incendie		Family		Transport	
	C	T	C	T	C	T
Total 2019	823511.678	1845632.55	6542594.679	5200146.074	1981047.066	-
Prédiction 2020	941379.025	1938210.61	6864527.450	8617698.6	2386389.692	-
Taux d'évolution	14.31%	5%	4.9%	65.7%	21.12%	-

TABLE 4.13 – Evolution chiffre d'affaires par branche (2019-2020)

Dans le tableau précédent, nous désignons par C Comptant (les affaires nouvelles) et T Terme.

Pour la branche Automobile, la différence du taux d'évolution prédit (Terme - Comptant) tend vers 7% . C'est une différence légère. L'intérêt de lancer une campagne de Marketing pour attirer des nouveaux clients de ce type d'assurance est si important que celui de garder les contactés déjà. Donc nous proposons de penser à des offres qui comprennent les deux (comptant et terme).

La branche Family (assurance vie) montre un taux d'évolution prédit qui s'élève à 65.7% en terme et 4.9% en comptant. Quelque soit en terme ou en comptant, cette évolution s'explique par l'orientation vers l'employabilité organisée chez les sociétés qui impose ce type d'assurance vie. De ce fait, nous pouvons faire le lien entre l'évolution en affaires nouvelles (comptant) et le taux d'employabilité du pays en quelque sorte.

Concernant les autres branches, elles peuvent être analysées de la même façon dans l'intérêt de lancer une campagne de Marketing.

4.6 Conclusion

Dans ce chapitre, nous avons appliqué deux méthodes pour modéliser le chiffre d'affaires globale ainsi que par branches sauf la branche Maladie et la branche Transport (terme) à cause d'un problème d'insuffisance de données. Nous passons maintenant à l'autre élément du business plan : la sinistralité.

Chapitre 5

Sinistralité : modélisation de l'IBNR

Résumé :

Dans ce chapitre, nous allons modéliser un élément important dans le calcul de provisionnement et la sinistralité : l'IBNR. Nous faisons le calcul de l'IBNR seulement pour la branche Automobile, en effet, c'est la branche la plus riche au niveau d'historique de données. Nous appliquons une méthode fréquemment utilisée dans le domaine d'assurance pour le calcul de l'IBNR : c'est la méthode "Chain Ladder".

5.1 Présentation des données

Notre étude se base sur des triangles de règlements d'historiques comptables, issus de quatre branches d'activité, nets de réassurance et en run-off. Ils ont été retraités de l'inflation de manière à manipuler des sinistres en euros constants. La périodicité des données considérées est annuelle. Nous

utiliserons les notations suivantes :

- i l'année de survenance des sinistres
- j l'année de développement
- $(X_{i,j})$ les paiements incrémentaux
- $(C_{i,j})$ les paiements cumulés : $C_{i,j} = \sum_{k=1}^j X_{i,k}$
- R_i les provisions par année de survenance : $R_i = \sum_{j=n-i+1}^n X_{i,j}, i = 1 \dots n$
- R le montant total des provisions pour une branche : $R = \sum_{i=1}^n R_i$

Triangle de liquidation :

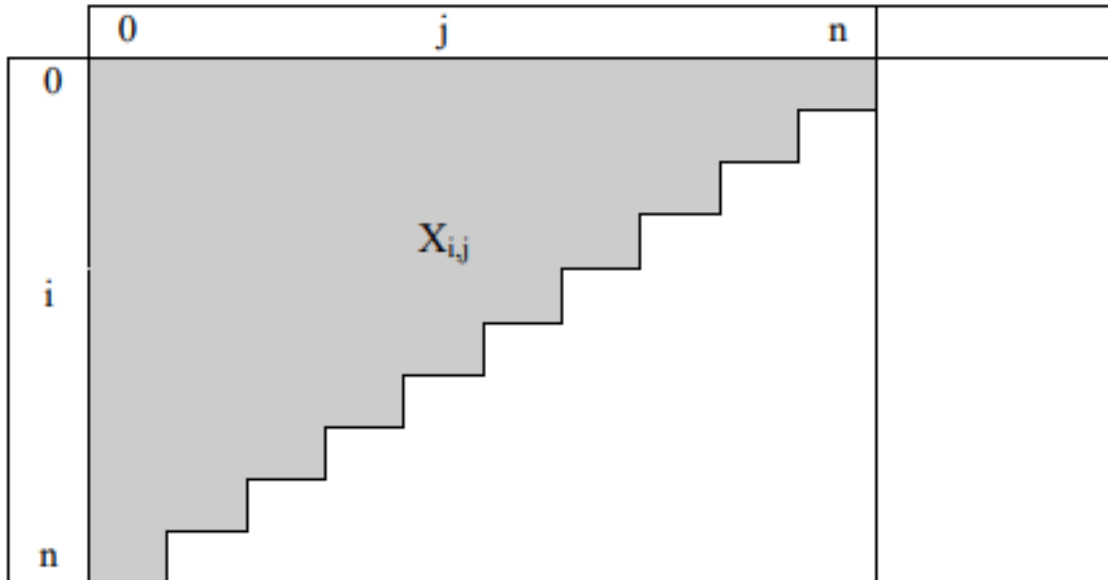


FIGURE 5.1 – Triangle de liquidation

La partie grisée représente les données historiques. Nous cherchons à estimer les paiements futurs concernant les sinistres survenus avant l'année n , c'est-à-dire la partie du tableau non grisée. Il existe de nombreuses méthodes permettant d'estimer le triangle inférieur.

Pour estimer la partie rectangulaire droite du tableau (nommée queue du triangle), nous avons mis en œuvre la méthode **"Chain-Ladder"**.

5.2 Méthode de chain ladder

La méthode Chain Ladder est une méthode déterministe fréquemment utilisée car facile à mettre en œuvre. Elle s'applique à des triangles de paiements cumulés ou des triangles de charges.

Soit le facteur de développement individuel $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$

La méthode de Chain Ladder repose sur l'hypothèse forte suivante :

Pour j allant de 1 à n , les facteurs de développement $f_{i,j}$ sont indépendants de l'année de survenance i .

Nous allons alors considérer des coefficients de passage, d'une année à l'autre, commun pour les années de survénance, et dont l'estimation est donnée par :

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j+1} C_{i,j+1}}{\sum_{i=0}^{n-j+1} C_{i,j}} \quad j = 0 \dots n$$

Grâce à ces facteurs, nous pouvons estimer :

- les charges ultimes par exercice de survénance $C_{i,n} = C_{i,n-1} * \prod_{j=n-i}^{n-1} \hat{f}_j$
- les provisions par exercice de survénance $R_i = (\hat{C}_{i,n} - C_{i,n-i}^*)$
- les provisions totales $\hat{R} = \sum_{i=1}^n \hat{R}_i$

5.2.1 Triangle Dommages Automobile

Le triangle suivant montre des données de l'IBNR de Zitouna Takaful (2012-2019) :

```
> rownames(RAA) <- seq(2012, 2019,1)
> colnames(RAA) <- seq(0, 7, 1)
> RAA
```

origin	dev							
	0	1	2	3	4	5	6	7
2012	1938.926	13491.55	13491.55	13491.55	13491.55	13491.55	13491.55	13491.55
2013	1547464.022	1790109.81	1803708.22	1866248.26	1866248.26	1866248.26	1866248.26	NA
2014	5023305.955	5947811.80	6044486.16	6143178.96	6148723.46	6148723.46	NA	NA
2015	7383709.246	8745208.14	8953786.60	9017777.43	9051277.43	NA	NA	NA
2016	12700107.390	14894203.25	15152424.18	15211674.85	NA	NA	NA	NA
2017	18348428.490	21341828.30	21646950.80	NA	NA	NA	NA	NA
2018	24525058.900	27602615.39	NA	NA	NA	NA	NA	NA
2019	28045136.570	NA	NA	NA	NA	NA	NA	NA

FIGURE 5.2 – Triangle IBNR

Ce triangle indique les valeurs connues des pertes de chaque année d'origine et des évaluations par la suite. Par exemple, les valeurs connues des pertes provenant de la Les périodes d'exposition en 2017 sont 18348428.490, 21341828.30 et 21646950.80 à la fin des années 2017, 2018 et 2019, respectivement. La dernière diagonale - c'est-à-dire le vecteur 13491.55, 1866248.26, . . . 28045136.570 du haut à droite au bas à gauche - montre l'évaluation la plus récente disponible. Les titres des colonnes - 0, 1, . . . 7 - contiennent les âges (en années) des observations dans la colonne relative au début de la période d'exposition. Par exemple, pour l'année d'origine 2017, l'âge de la valeur 21646950.80, évaluée au 31 décembre 2019, est de trois ans. L'objectif d'un exercice de réservation est de prévoir l'évolution

future des sinistres dans le coin inférieur droit du triangle et les développements potentiels au-delà développement à l'âge de 8 ans. Finalement, toutes les demandes pour une période d'origine donnée

seront réglées, mais il n'est pas toujours évident de juger cela prendra combien d'années, voire de décennies.

5.2.2 Traçage de triangles

Pour un ensemble de données de classe triangle, le package ChainLadder fournit des méthodes de traçage par défaut pour donner un aperçu graphique des données :

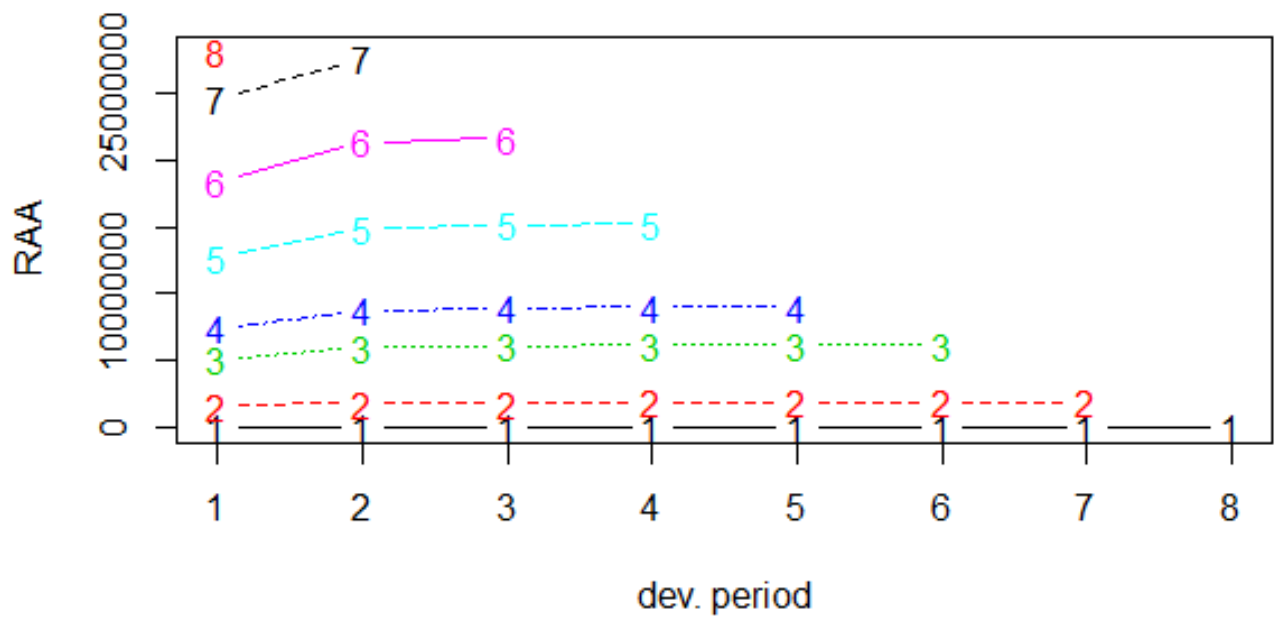


FIGURE 5.3 – Tableau d'évolution des sinistres du triangle Automobile, avec une ligne par période d'origine

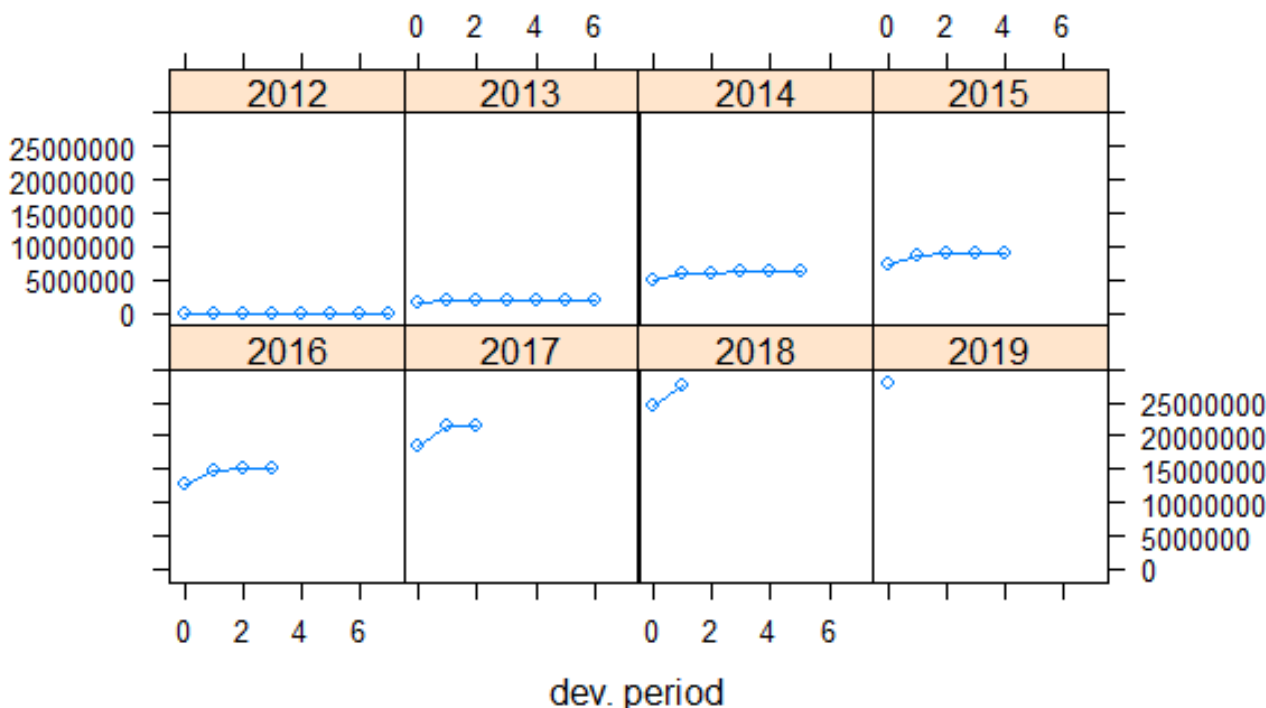


FIGURE 5.4 – Tableau d'évolution des sinistres du triangle automobile, avec des panneaux individuels pour chaque période d'origine

Nous remarquons, d'après les deux graphiques précédents, que le triangle automobile présente l'évolution des demandes pour les années d'origine 2012 à 2019 sous une forme cumulative.

5.2.3 Application de la méthode "chain Ladder"(triangle automobile)

Dans un premier temps, les ratios de liaison âge-âge sont calculés comme étant les ratios de développement moyens pondérés en volume d'un triangle de développement des pertes cumulées d'une période de développement à l'autre.

Souvent, il ne convient pas de supposer que l'année d'origine la plus ancienne est pleinement développée. Une approche typique consiste à extrapoler les ratios de développement, par exemple en supposant un modèle log-linéaire.

```

> dev.period <- 1:(n1-1)
> plot(log(f1-1) ~ dev.period, main="Log-linear extrapolation of age-to-age factors")
> tail.model <- lm(log(f1-1) ~ dev.period)
> abline(tail.model)
> co <- coef(tail.model)
> tail <- exp(co[1] + c(n:(n + 100)) * co[2]) + 1
> f.tail <- prod(tail)
> f.tail
[1] 1.000014

```

FIGURE 5.5 – Extrapolation log-linéaire des facteurs d'âge

Les facteurs d'âge nous permettent de tracer les schémas de développement attendus des sinistres.

```

> dev.period <- 1:(n1-1)
> plot(log(f1-1) ~ dev.period, main="Log-linear extrapolation of age-to-age factors")
> tail.model <- lm(log(f1-1) ~ dev.period)
> abline(tail.model)
> co <- coef(tail.model)
> tail <- exp(co[1] + c(n:(n + 100)) * co[2]) + 1
> f.tail <- prod(tail)
> f.tail
[1] 1.000014

```

FIGURE 5.6 – schéma de developpement des sinistres attendus

Les ratios de liaison sont ensuite appliqués au dernier montant cumulé connu des sinistres pour prévoir la prochaine période de développement. La quadrature du triangle automobile est calculée ci-dessous, où une dernière colonne est ajoutée à droite pour tenir compte du développement prévu au-delà de l'âge le plus ancien (10) du triangle en raison de la queue facteur (1.000014) étant supérieur à l'unité.


```

> f <- c(f, f.tail)
> fullRAA <- cbind(RAA, Ult = rep(0, 8))
> for(k in 1:n){
+   fullRAA[(n-k+1):n, k+1] <- fullRAA[(n-k+1):n,k]*f[k]
+ }
> round(fullRAA)

```

	0	1	2	3	4	5	6	7	Ult
2012	1939	13492	13492	13492	13492	13492	13492	13492	13492
2013	1547464	1790110	1803708	1866248	1866248	1866248	1866248	1866248	1866274
2014	5023306	5947812	6044486	6143179	6148723	6148723	6148723	6148723	6148809
2015	7383709	8745208	8953787	9017777	9051277	9051277	9051277	9051277	9051403
2016	12700107	14894203	15152424	15211675	15246529	15246529	15246529	15246529	15246740
2017	18348428	21341828	21646951	21839582	21889622	21889622	21889622	21889622	21889925
2018	24525059	27602615	28064395	28314133	28379008	28379008	28379008	28379008	28379402
2019	28045137	32403468	32945564	33238739	33314897	33314897	33314897	33314897	33315359

FIGURE 5.7 – triangle intégral

La perte totale estimée selon cette méthode est d'environ 6325200 DT :

```

> sum(fullRAA[,9] - getLatestCumulative(RAA))
[1] 6325286

```

FIGURE 5.8 – Perte totale

5.3 Conclusion

Dans ce chapitre, nous avons modéliser une partie de l'élément sinistre du business plan. L'objectif était de modéliser toute la partie de la sinistralité ainsi que les autres éléments du business plan mais nous nous sommes arrêtés à ce niveau à cause d'une contrainte du temps.

Conclusion générale et perspectives

Dans ce projet, nous nous sommes intéressés à la modélisation de quelques éléments du Business Plan (le chiffre d'affaires et une partie de la sinistralité).

Cette modélisation apporte une valeur ajoutée pour l'analyse de ces éléments car elle permet de prédire le future de ces éléments de base du Business plan pour avoir une idée globale de l'activité future de l'entreprise.

Pour ce faire, nous avons introduit les notions des séries temporelles, chose qui n'a jamais été utilisée dans ce contexte vu que toutes les estimations sur lesquelles se base le Business Plan de Zitouna Takaful sont basées sur l'historique de la compagnie et le savoir faire des collaborateurs.

Il existe plusieurs approches pour la prévision des séries chronologiques. Parmi ces approches nous avons utiliser deux approche : l'approche de Box Jenkins et l'approche d'apprentissage en profondeur, en particulier dans la famille RNN. Nous avons proposé la méthode la plus répandue pour chaque approche pour évaluer le chiffre d'affaires : l'une est basée sur l'autoregressivité et la moyenne mobile (SARiMA : approche Box Jenkins) et l'autre est basée sur l'apprentissage automatique (LSTM famille RNN).

La prévision des séries chronologiques peut s'avérer complexe et compliquée avec l'approche d'apprentissage automatique : bien qu'elles soient pertinentes et performantes, elles présentent, néanmoins, certaines limites telles que les temps de calcul importants, la complexité de la modélisation, et la convergence des algorithmes de résolutions en ce qui concerne les approches par décomposition. Par contre, une technique simple et efficace : le modèle SARIMA, peut offrir l'avantage de bon résultats pour un faible coût en efforts et complexité.

Pour l'évaluation des sinistres rétroactifs, nous avons utiliser une méthode fréquemment utilisée dans le domaine d'assurance : la méthode "Chian Ladder". La facilité de son implémentation et la pertinence de ses résultats l' a rendu la plus utilisée dans le domaine de calcule de provisionnement.

A l'issu de la crise de Coronavirus, des perspectives peuvent être envisagées : une modélisation statistique incluant l'hypothèse des chocs économiques dans le but d'étudier l'impact de cette crise sur le chiffre d'affaires et l'activité de l'entreprise.

Annexe A

Structure du business-plan

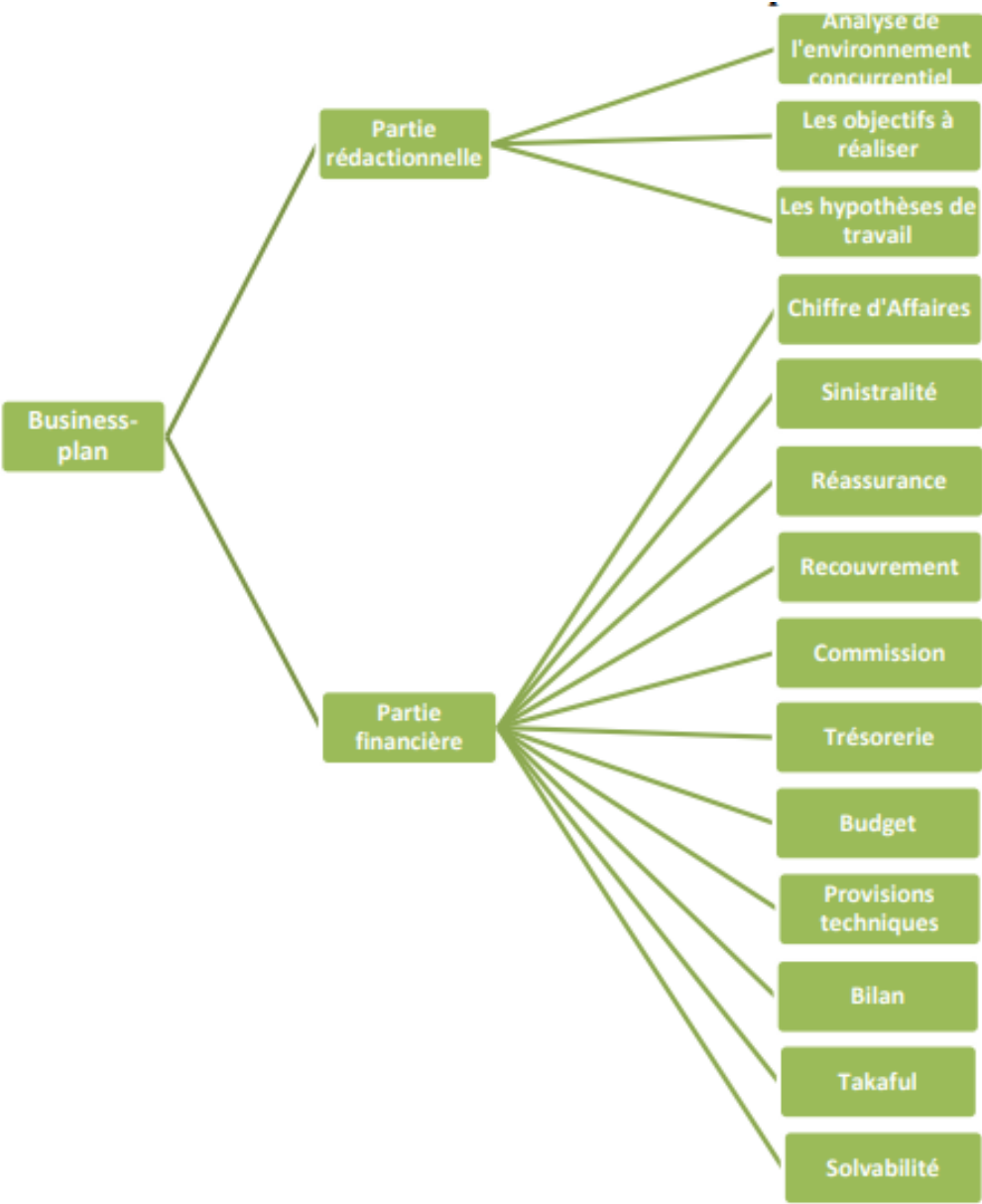


FIGURE A.1 – Structure du Business Plan

Annexe B

Structure du Chiffre d'affaires

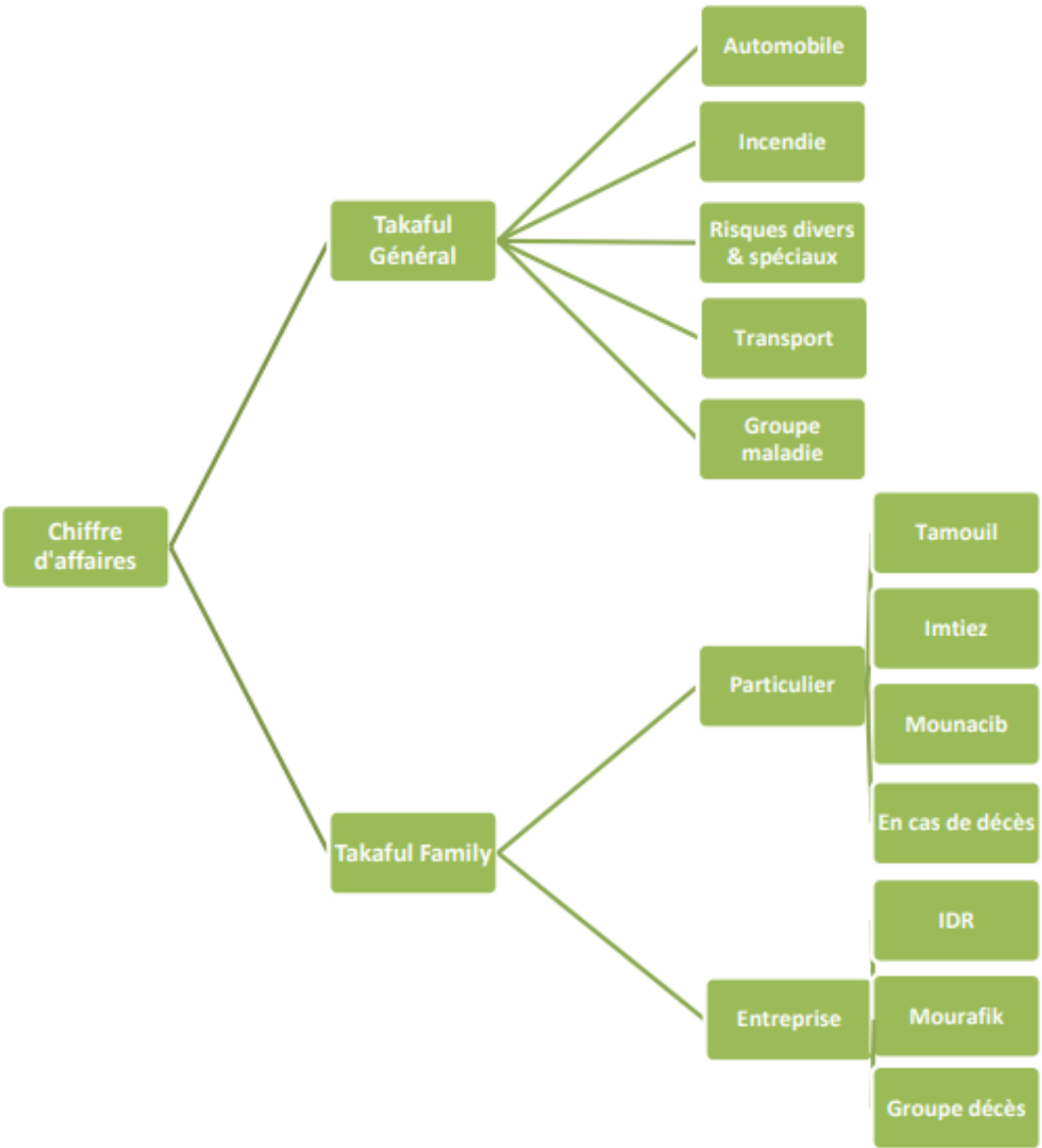


FIGURE B.1 – Structure du Chiffre d'affaires

Annexe C

LSTM

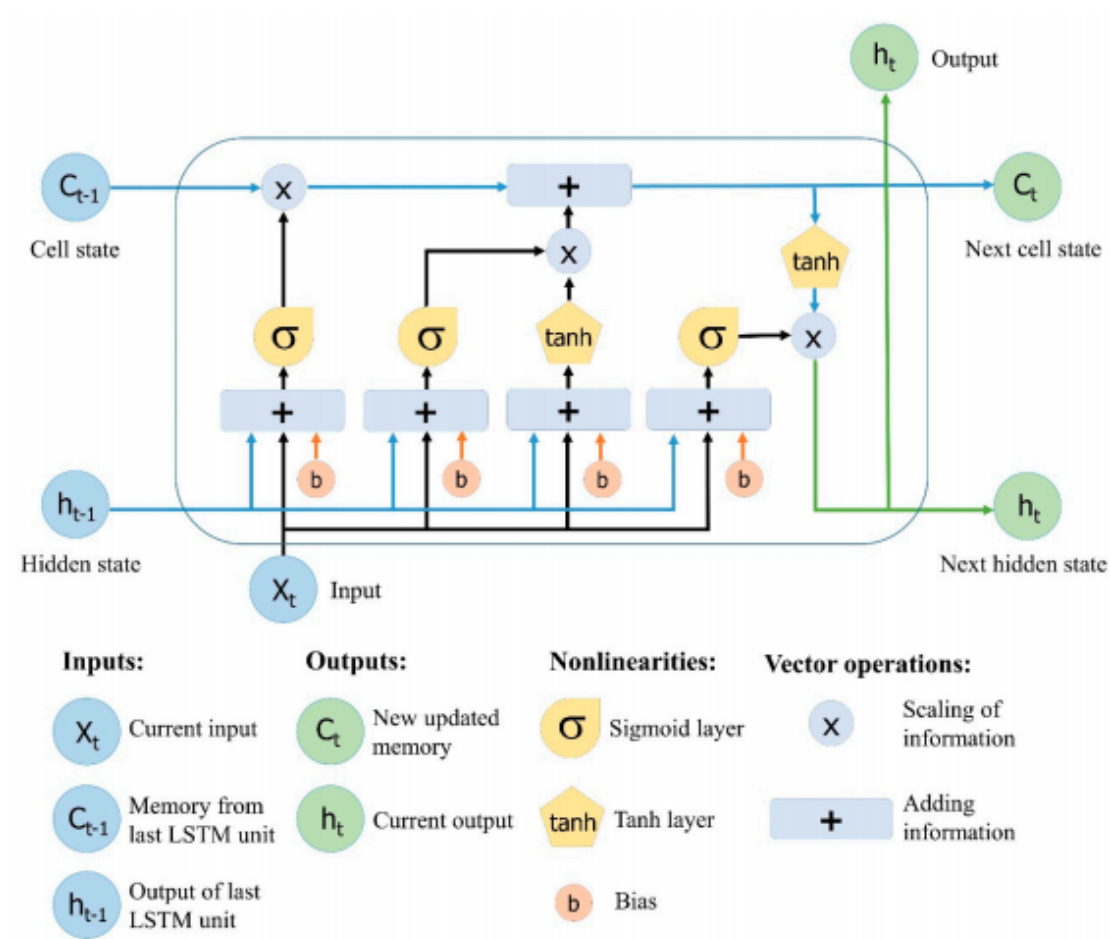


FIGURE C.1 – Structure de LSTM

Annexe D

La classe TimeseriesGenerator

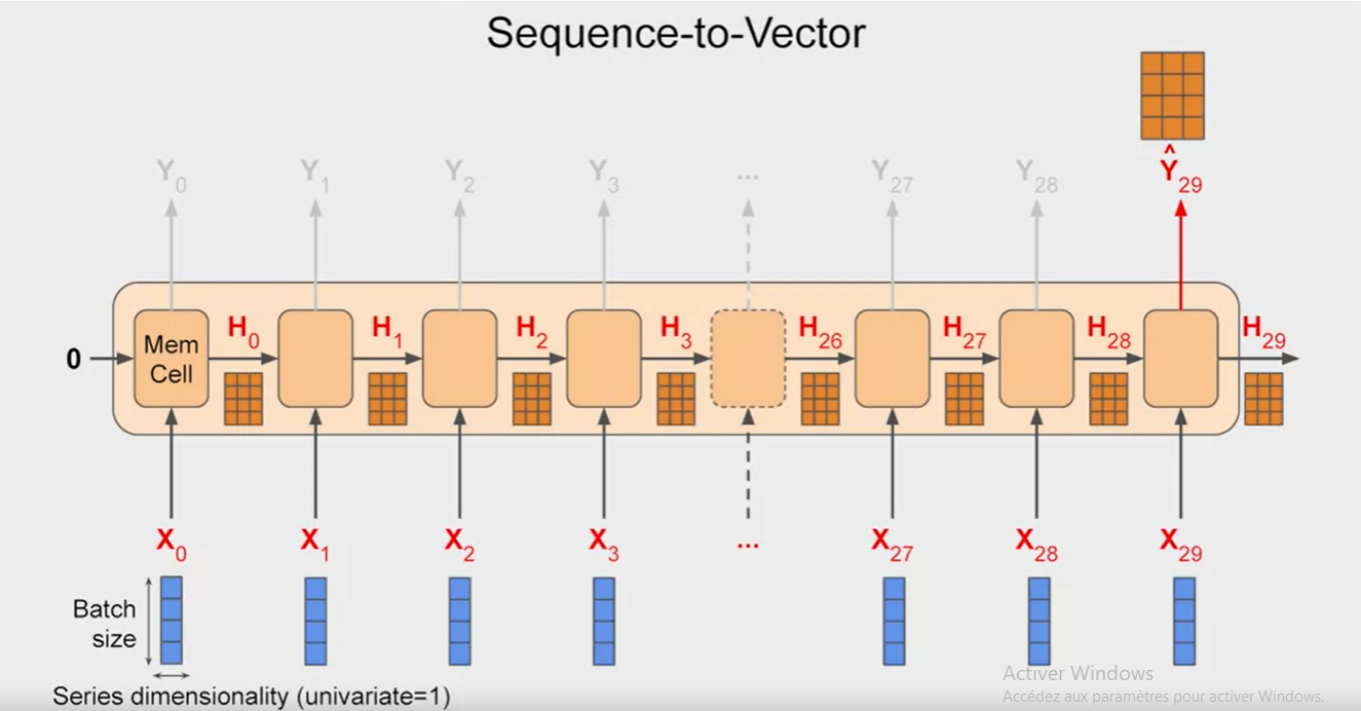


FIGURE D.1 – LSTM

TimeseriesGenerator class

```
tf.keras.preprocessing.sequence.TimeseriesGenerator(  
    data,  
    targets,  
    length,  
    sampling_rate=1,  
    stride=1,  
    start_index=0,  
    end_index=None,  
    shuffle=False,  
    reverse=False,  
    batch_size=128,  
)
```

FIGURE D.2 – La classe TimeseriesGenerator

Utility class for generating batches of temporal data.

This class takes in a sequence of data-points gathered at equal intervals, along with time series parameters such as stride, length of history, etc., to produce batches for training/validation.

Arguments

- **data**: Indexable generator (such as list or Numpy array) containing consecutive data points (timesteps). The data should be at 2D, and axis 0 is expected to be the time dimension.
- **targets**: Targets corresponding to timesteps in `data`. It should have same length as `data`.
- **length**: Length of the output sequences (in number of timesteps).
- **sampling_rate**: Period between successive individual timesteps within sequences. For rate `r`, timesteps `data[i]`, `data[i-r]`, ... `data[i - length]` are used for create a sample sequence.
- **stride**: Period between successive output sequences. For stride `s`, consecutive output samples would be centered around `data[i]`, `data[i+s]`, `data[i+2*s]`, etc.
- **start_index**: Data points earlier than `start_index` will not be used in the output sequences. This is useful to reserve part of the data for test or validation.
- **end_index**: Data points later than `end_index` will not be used in the output sequences. This is useful to reserve part of the data for test or validation.
- **shuffle**: Whether to shuffle output samples, or instead draw them in chronological order.
- **reverse**: Boolean: if `true`, timesteps in each output sample will be in reverse chronological order.
- **batch_size**: Number of timeseries samples in each batch (except maybe the last one).

Returns

A `Sequence` instance.

FIGURE D.3 – Paramètres de la classe TimeseriesGenerator

Bibliographie

- [1] les sinistres encourus mais non déclarés Assurance. *Incurred But Not Reported*. https://en.wikipedia.org/wiki/Incurred_but_not_reported. 2016.
- [2] assurance&mutuelle. *Principe de fonctionnement de la réassurance*. <http://www.assurance-et-mutuelle.com/assurance/reassurance.html>. 2020.
- [3] BetterStudy. *PLAN DE TRÉSORERIE : DÉFINITION ET COMMENT L'ÉTABLIR* . <https://betterstudy.ch/plan-tresorerie/>.
- [4] Greta M. Ljung George E. P. Box. *Ljung–Box test* . <https://en.wikipedia.org/wiki/Ljung2020>.
- [5] *Le budget prévisionnel*. Le budget prévisionnel : mode d'emploi. <https://www.thebusinessplanshop.com/fr/blog/budget-previsionnel>. 2019.
- [6] Assuralia – CMS. RECOMMANDATION 09 AVIS D'ECHEANCE – DEMANDE DE PAIEMENT. https://www.telebib2.org/WhiteBook/FR/RB_09_v23_20150106_CMS.pdf. 2015.
- [7] Assuralia – CMS. RECOMMANDATION 09 AVIS D'ECHEANCE – DEMANDE DE PAIEMENT. https://www.telebib2.org/WhiteBook/FR/RB_09_v23_20150106_CMS.pdf. 2015.
- [8] Compta-Facile. La partie rédactionnelle du Business Plan. <https://www.compta-facile.com/business-plan-partie-redactionnelle/>. 2019.
- [9] Compta-Facile. La partie rédactionnelle du Business Plan. <https://www.compta-facile.com/business-plan-partie-redactionnelle/>. 2019.
- [10] *Ratio de solvabilité*. Taux de couverture de la marge de solvabilité (ou marge de solvabilité sur fonds propres). <https://astyrian.fr/taux-de-couverture-de-la-marge-de-solvabilite-ratio-de-solvabilite-definition>. 2019.
- [11] Comité Général des Assurances. Le Marché Tunisien des Assurances. rapport annuel du Comité Général des Assurances (CGA) 2013 . http://www.cga.gov.tn/fileadmin/contenus/pdf/rapport_annuel_FR_2013.pdf. 2013.
- [12] Comité Général des Assurances (CGA). Rapport annuel 2018 : Le marché Tunisien des assurances. http://www.cga.gov.tn/fileadmin/contenus/pdf/Rapport_CGA_FR-ANG_-_2018_final.pdf. 2018.
- [13] *Le Coin des Entrepreneurs*. Le plan de trésorerie prévisionnel . <https://www.lecoindesentrepreneurs.fr/le-plan-de-tresorerie/>. 2019.
- [14] *LExpertcomptable*. Business plan français vs business plan anglo-saxon. l-expert-comptable.com/a/534154-business-plan-francais-vs-business-plan-anglo-saxon.html. 2020.
- [15] *Mieux comprendre l'assurance L'assurance, Guide*. Qu'est-ce que la quittance d'assurance? <https://www.lesfurets.com/assurance/guide/quest-ce-que-la-quittance-dassurance>. 2017.
- [16] Sima Siامي-Namini Neda Tavakoli Akbar Siامي Namin. A Comparison of ARIMA and LSTM in Forecasting Time Series . <https://ieeexplore.ieee.org/abstract/document/8614252>. 2018.

- [17] *Procomptable*. Le contrôle interne et l'organisation comptable dans les entreprises d'assurance et / ou de réassurance. https://www.procomptable.com/normes/nc_27.htm. 2001.
- [18] *Procomptable*. Les revenus dans les entreprises d'assurances et / ou de réassurance. https://www.procomptable.com/normes/nc_28.htm. 2001.
- [19] *Procomptable*. Les charges techniques dans les entreprises d'assurance et / ou de réassurance. http://www.oect.org.tn/tiny_mce/plugins/filemanager/files/Reglementation/SystemeComptableTu
- [20] *procomptable*. Les revenus dans les entreprises d'assurances et / ou de réassurance . https://www.procomptable.com/normes/nc_28.html. 2001.
- [21] D. Scheuerell and E. J. Ward. *Applied Time Series Analysis* . <https://nwfsctimeseries.github.io/atsa-labs/index.html>. 2020.
- [22] *spac actuaires*. *Provision mathématique* . <https://www.spac-actuaire.fr/lexique/provision-mathematique/>. 2020.
- [23] Assurance Zitouna Takaful. *Compagnie d'Assurance en Tunisie qui opère selon le modèle Takaful*. <http://www.zitounatakaful.com.tn/Fr/>. 2011.
- [24] *Système Comptable Tunisien*. Les revenus dans les entreprises d'assurances et / ou de réassurance. http://www.oect.org.tn/tiny_mce/plugins/filemanager/files/Reglementation/SystemeComptable
- [25] *The Pennsylvania State University*. Seasonal ARIMA models . <https://online.stat.psu.edu/stat510/lesson/4/4.1>. 2020.
- [26] *wikipedia*. Provision technique . https://fr.wikipedia.org/wiki/Provision_technique. 2018.

