



Département : Génie Industriel

Entreprise : Société Générale

## Mémoire de projet de fin d'études

Pour l'obtention du diplôme d'ingénieur d'Etat en Génie Industriel

Option Data Science et Intelligence Artificielle

---

# Utilisation de l'IA pour la conception d'un outil de Stress-Test de la trésorerie de la Société Générale Algérie

---

Réalisé par :

**Mme.NAMANI Nawal**

**M.BENSEFIA Walid**

Sous la direction de :

**M.AIT OUARABI Chems-Eddine (SGA)**

**M.ARKI Oussama (ENP)**

Présenté et soutenu publiquement le 26 juin 2023

### Composition du jury :

Président	M.HAMRI OKBA,	MCA	ENP
Examineur	M.FOURAR LAIDI HAKIM,	MCA	ENP
Promoteur	M.ARKI OUSSAMA,	MCB	ENP

ENP : 2023





Département : Génie Industriel

Entreprise : Société Générale

## Mémoire de projet de fin d'études

Pour l'obtention du diplôme d'ingénieur d'Etat en Génie Industriel

Option Data Science et Intelligence Artificielle

---

# Utilisation de l'IA pour la conception d'un outil de Stress-Test de la trésorerie de la Société Générale Algérie

---

Réalisé par :

**Mme.NAMANI Nawal**

**M.BENSEFIA Walid**

Sous la direction de :

**M.AIT-OUARABI Chems-Eddine (SGA)**

**M.ARKI Oussama (ENP)**

Présenté et soutenu publiquement le 26 juin 2023

### Composition du jury :

Président	M.HAMRI OKBA,	MCA	ENP
Examineur	M.FOURAR LAIDI HAKIM,	MCA	ENP
Promoteur	M.ARKI OUSSAMA,	MCB	ENP

ENP : 2023

# *Dédicaces*

*Je dédie ce travail,*

*À maman et papa, pour leur amour inconditionnel, leur soutien et encouragements, vous avez toujours été ma plus grande motivation et source d'inspiration.*

*À la plus merveilleuse des tantes qui est aussi tendre qu'une mère, Khalto.*

*À ma grande sœur et mon futur petit neveu.*

*À Yassou, ma meilleure amie, ma sœur et mon bras droit.*

*À mon jumeau, frère et surtout soutien moral, Raouf.*

*Au groupe Vrais 10, et plus particulièrement mes amis Souhaib et Amira, vous avez fait de ma prépa une période inoubliable.*

*Aux Daïtas, ce fut un plaisir d'avoir partagé trois belles années avec vous. Mention spéciale pour Kimo, la personne la plus serviable que je n'ai jamais connu.*

*À mes magnifiques rencontres, Imene, Celia, Amel, Lyna pour tous les fou-rires et moments inoubliables. Je vous aime !*

*À mon binôme du primaire, collègue, lycée et avec qui j'ai partagé ce travail, Walid. Quelle belle façon de clore ce chapitre de notre vie !*

*À tous les membres d'IEC et du CAP, j'ai adoré organiser pleins d'événements avec vous, toujours dans la bonne humeur.*

*Le meilleur pour la fin, à la personne qui m'a toujours poussé vers le haut et encouragé à donner le meilleur de moi-même, Yacine.*

***Nawel***

# *Dédicaces*

*Je souhaite sincèrement dédier ce travail,*

*À la personne la plus chère à mon cœur, ma mère, cette sacrée femme qui fait constamment des sacrifices pour me voir réaliser mes ambitions et atteindre mes buts.*

*À mon père, à ce brave héros que j'admire beaucoup, occupant une place précieuse dans mon cœur et qui a été le socle solide de mon éducation, instaurant les valeurs qui m'ont guidé tout au long de mon parcours.*

*À mes adorables frère et sœur, Adoula et Lyna, que j'aime plus que tout et qui ont attentivement suivi chaque étape de mon évolution avec une admiration qui me touche profondément. Mon rêve le plus cher est de vous voir réussir en vous soutenant de tout mon cœur.*

*À mes cousins et cousines, tantes et oncles, spécialement Kaissa et Akila.*

*À la mémoire mes grands-pères et de ma grand-mère maternelle, avec qui j'ai passé des étés inoubliables, et en honneur de ma grand-mère paternelle encore avec nous, que Dieu la garde.*

*À mes chers amis qui ont été présents à chaque étape de ma vie.*

*À l'ensemble de mes camarades de la promotion Data Science & IA.*

*À ma binôme du primaire, collègue, lycée et avec qui j'ai partagé ce travail, Nawel. Une très belle manière de conclure ce chapitre de nos parcours.*

*Aux membres de IEC et du CAP avec lesquels j'ai passé de très beaux et inoubliables moments.*

*À mes anciens professeurs du lycée et de l'école "Pas à Pas", en particulier à ceux qui m'ont soutenu de près. À Tata Chafika et en hommage à madame Haddad.*

*Et finalement, à tous ceux qui ont vu en moi un potentiel, qui ont cru en mes compétences et en ma réussite.*

**Walid**

# Remerciements

Nous adressons nos vifs remerciements à notre encadrant de la SGA, M. Chems-Eddine Ait Ouarabi, responsable ALM, pour son encadrement rigoureux, ses conseils avisés, son engagement inébranlable et sa disponibilité inestimable tout au long de notre projet. Il a été bien plus qu'un simple encadrant, il a été un mentor pour nous, et nous avons énormément tiré de son expertise et de ses connaissances. Malgré son calendrier chargé, il a su trouver du temps pour nous guider et nous soutenir, ce qui a grandement contribué à la réussite de notre projet. Nous lui en sommes profondément reconnaissants.

Nous remercions M. Oussama ARKI, notre promoteur, pour son encadrement et ses conseils.

Nous exprimons notre gratitude à l'ensemble de l'équipe ALM, dont Mr Thamer AHMEDI, pour son soutien constant qui a été une source d'encouragement et de motivation tout au long de ce projet et M. Ouail GUEROUI pour son accueil chaleureux et son intégration bienveillante dès le début de notre parcours.

Nous souhaitons exprimer notre sincère gratitude et reconnaissance envers l'ensemble du personnel de la SGA, et plus particulièrement envers ceux de la direction financière. Ils nous ont permis de travailler dans un environnement agréable. Une mention spéciale pour M. Mohamed SELMANI, responsable CDG, dont l'importante requête nous a permis d'extraire les données nécessaires à notre travail.

Nous remercions par avance les membres du jury, qui nous ont fait l'honneur d'évaluer notre travail.

Nous remercions par la même occasion l'ensemble des enseignants du département Génie Industriel qui nous ont accompagné tout au long de notre formation.

# ملخص

منذ الأزمة المالية لعام 2008 ، تم وضع إصلاحات جديدة تهدف إلى توفير إطار تنظيمي أكثر حماية ، لا سيما من خلال إنشاء أدوات تحليل المخاطر بما في ذلك اختبارات الإجهاد المصرفية "اختبارات الإجهاد" التي تهدف إلى تقييم قوة البنوك في مواجهة الصدمات المالية التي يمكن أن تسبب أزمة سيولة. بالإضافة إلى تطوير سيناريو الإجهاد، فإن الخطوة الأكثر أهمية هي وضع توقعات لأهم البنود النقدية للبنك من أجل التأكيد عليها ثم إجراء اختبار التحمل النهائي

الهدف من مشروع نهاية الدراسات هذا هو اقتراح نموذج تنبؤ للتدفق النقدي لـ SGA باستخدام خوارزميات التعلم الآلي ، ثم حقنه بسيناريوهات الإجهاد بناءً على افتراضات الاقتصاد الكلي والاقتصاد الجزئي. لتقليل مخاطر السيولة المستقبلية بحيث يمكن للبنك الوفاء بالتزاماته.

---

**كلمات مفتاحية :** السيولة ، اختبار الضغط ، التعلم الآلي ، البنك ، التنبؤ ، إدارة الأصول والخصوم.

---

# Abstract

Since the financial crisis of 2008, new reforms have been put in place to provide a more protective regulatory framework, notably through the implementation of risk analysis tools, including bank stress tests, aimed at assessing the robustness of banks to financial shocks that may cause a liquidity crisis. In addition to the elaboration of a stress scenario, the most important step is to establish a forecast of the most significant items of the bank's cash flow in order to stress it and then have the final stress test. The objective of this Final Year Project is first to propose a forecasting model for SGA's cash flow thanks to Machine Learning algorithms, then to inject stress scenarios based on macroeconomic and microeconomic hypotheses to minimize the future liquidity risk so that the bank can meet its commitments.

---

**Keywords :** Liquidity, Stress-test, Machine Learning, Banking, Forecasting, ALM.

---



# Résumé

Depuis la crise financière de 2008, de nouvelles réformes ont été mises en place visant à fournir un cadre réglementaire plus protecteur notamment par la mise en place d'outils d'analyse de risque dont les tests de résistance bancaire "Les stress-tests" visant à évaluer la robustesse des banques face aux chocs financiers qui peuvent causer une crise de liquidité. En plus de l'élaboration d'un scénario de stress, l'étape la plus importante est d'établir un prévisionnel des postes les plus significatifs de la trésorerie de la banque afin de le stresser pour ensuite avoir le stress-test final.

L'objectif de ce PFE est en premier lieu de proposer un modèle de prévision pour la trésorerie de la SGA grâce à des algorithmes de Machine Learning, ensuite de lui injecter des scénarios de stress basés sur des hypothèses macroéconomiques et microéconomiques pour minimiser le risque futur de liquidité afin que la banque puisse faire face à ses engagements.

---

**Mots clés :** Liquidité, Stress-test, Machine Learning, Banque, Prévision,ALM.

---

# Table des matières

<b>Table des figures</b>	
<b>Liste des tableaux</b>	
<b>Liste des sigles et acronymes</b>	
<b>Introduction générale</b>	15
<b>I État des lieux</b>	16
<b>Introduction</b>	17
<b>1 Organisme d'accueil</b>	18
1.1 Groupe Société Générale	18
1.2 Société Générale Algérie	19
1.3 Direction financière	20
1.4 Département ALM	20
1.5 Domaine de l'ALM	20
1.5.1 Passifs et Actifs	20
1.5.2 Activités de l'ALM	21
<b>2 Étude de l'existant</b>	23
2.1 Le prévisionnel de la trésorerie au niveau de la SGA :	23
2.2 Stress-test de la trésorerie :	23
2.2.1 Définition :	23
2.2.2 Stress-test de la trésorerie au sein de la SGA :	24
2.3 Diagnostic externe :	24
2.3.1 Analyse SWOT	24
2.3.2 Analyse PESTEL	25
2.4 Enoncé de la problématique	26
2.5 Méthodologie de travail	26
<b>II État de l'art</b>	28
<b>Introduction</b>	29
<b>3 Liquidité et risque bancaire</b>	30
3.1 Définition du risque bancaire	30
3.2 Les différents risques bancaires	30

3.2.1	Risques structurels	30
3.2.2	Les autres risques	31
3.3	Les différents types de liquidités	31
3.4	Coefficient de liquidité :	32
3.5	Supervision bancaire et comité de Bâle	32
3.5.1	Définition de la supervision bancaire :	32
3.5.2	Présentations du comité :	32
3.5.3	Missions de Bâle :	32
3.6	Besoin en liquidité :	33
3.7	Situation d'illiquidité et méthodes de refinancement des banques	33
3.8	Réglementation BA sur le contrôle de risque de liquidité, le ratio des fonds propres et des ressources permanentes :	34
<b>4</b>	<b>L'IA pour le Stress Test</b>	<b>36</b>
4.1	Stress-test	36
4.1.1	Introduction	36
4.1.2	Historique	36
4.1.3	Stress tests au sein des banques	37
4.1.4	Catégories	37
4.1.5	Typologie	38
4.1.6	Approches	38
4.1.7	Stress test de la trésorerie	39
4.2	Machine Learning et Séries Temporelles	40
4.2.1	Intelligence Artificielle en Finance	40
4.2.2	Introduction au Machine Learning	41
4.2.3	Séries Temporelles	44
4.3	Etude bibliographique	47
4.3.1	Tableau comparatif	47
4.3.2	Algorithmes identifiés :	52
<b>III</b>	<b>Conception de l'outil de stress-test</b>	<b>61</b>
	<b>Introduction</b>	<b>62</b>
<b>5</b>	<b>Prévisionnel</b>	<b>63</b>
5.1	Collecte des Données	63
5.1.1	Extraction des données	63
5.1.2	Importation des données	64
5.2	Analyse des Données	67
5.2.1	Statistiques descriptives	67
5.2.2	Vérification de la complétude et des valeurs manquantes	69
5.2.3	Visualisation des données	70
5.2.4	Analyse des valeurs aberrantes	73
5.2.5	Décomposition des séries temporelles	75
5.2.6	Test de stationnarité	77
5.2.7	Autocorrélation et Autocorrélation Partielle	78
5.2.8	Vérification de la normalité	81
5.2.9	Analyse de la volatilité	84
5.3	Préparation des Données et Modélisation	86

5.3.1	Identification des modèles	87
5.3.2	Sélection des modèles	88
5.3.3	Feature engineering	91
5.3.4	Entraînement des modèles et optimisation de l'Apprentissage Au- tomatique	92
5.4	Résultats et Evaluation	94
5.4.1	Résultats	94
5.4.2	Métriques	96
5.4.3	Evaluation	97
5.4.4	Conclusion	98
<b>6</b>	<b>Stress-test</b>	<b>99</b>
6.1	Introduction	99
6.1.1	Scénario de stress pour la Caisse	99
6.1.2	Scénario de stress pour le Trésor public	101
6.2	Résultats du stress	103
6.2.1	Résultat du <i>scénario 1</i> sur la caisse	103
6.2.2	Résultat du <i>scénario 2</i> sur le compte Trésor	103
6.3	Conclusion	104
<b>7</b>	<b>Application utilisateur</b>	<b>105</b>
7.1	Présentation de l'outil	105
7.2	Fonctionnalités	106
	<b>Conclusion générale</b>	<b>111</b>
	<b>Bibliographie</b>	<b>112</b>
	<b>Annexes</b>	<b>115</b>
<b>A</b>	<b>Actifs et Passifs</b>	<b>115</b>
<b>B</b>	<b>Exploration, analyse de données et modélisation : CCP</b>	<b>117</b>
<b>C</b>	<b>Exploration, analyse de données et modélisation : Trésor</b>	<b>122</b>
<b>D</b>	<b>Exploration, analyse de données et modélisation : Scellés BA</b>	<b>128</b>
<b>E</b>	<b>Présentation des outils utilisés</b>	<b>133</b>

# Table des figures

1.1 Organigramme de la SGA	19
1.2 Organigramme de la direction financière de la SGA	20
1.3 Organigramme de l'ALM	20
1.4 Illustration du bilan de gestion	21
2.1 Cycle de vie de la méthodologie CRISP-DM	27
4.1 Schéma synthétique du projet	62
5.1 Indexation des données	65
5.2 Ré-échantillonnage des données par semaine	65
5.3 Visualisation du DataFrame de la caisse	66
5.4 Données utilisées pour calculer les soldes :	67
5.5 Vérifications de la complétude et des valeurs manquantes	69
5.6 Mouvements hebdomadaires de la Caisse	70
5.7 Histogramme des Mouvements annuels de la Caisse	70
5.8 Evolution des Mouvements avec Moyenne Mobile et Variance	71
5.9 Soldes hebdomadaires de la Caisse	71
5.10 Histogramme des Soldes annuels de la Caisse	72
5.11 Evolution des Soldes avec Moyenne Mobile et Variance	72
5.12 Boite à Moustaches (BoxPlot) des mouvements de la caisse	74
5.13 Boite à Moustaches (BoxPlot) des soldes de la caisse	74
5.14 Décomposition de la série temporelle des mouvements de la caisse	76
5.15 Décomposition de la série temporelle du solde de la caisse	76
5.16 ADF Test sur les Mouvements de la Caisse	78
5.17 Fonction d'autocorrelation des mouvements de la caisse	79
5.18 Fonction d'autocorrelation partielle des mouvements de la caisse	80
5.19 Fonction d'autocorrelation du solde de la caisse	80
5.20 Fonction d'autocorrelation partielle du solde de la caisse	81
5.21 QQ Plot des mouvements de la caisse	82
5.22 Distribution des mouvements hebdomadaires de la caisse	82
5.23 QQ Plot du solde de la caisse	83
5.24 Distribution du solde hebdomadaires de la caisse	84
5.25 Analyse de la volatilité des mouvements hebdomadaires de la caisse	85
5.26 Analyse de la volatilité du solde hebdomadaire de la caisse	86
5.27 Fonctionnement général d'un RNN	88
5.28 Structure générale d'une cellule LSTM	89
5.29 Tête du DataFrame des données d'entraînement de la caisse	92
5.30 Modélisation - Compilation et Adaptation	93
5.31 Modélisation - Optimisation	94
5.32 Prévisions du modèle RNN pour les soldes hebdomadaires de la caisse	95

5.33	Prévisions du modèle LSTM pour les soldes hebdomadaires de la caisse	95
5.34	Prévisions du modèle Prophet pour les soldes hebdomadaires de la caisse	96
6.1	Le prévisionnel BAU et le prévisionnel stressé de la caisse	103
6.2	Le prévisionnel BAU et le prévisionnel stressé du Trésor	104
7.1	Structure de l'outil de prévision	106
7.2	Choix de la date de prévision	107
7.3	Choix du type de poste à prédire	107
7.4	Choix du type de prévision	108
7.5	Prédictions générées	108
7.6	Graphique des prédictions	109
B.1	Evolution du solde CCP	118
B.2	Histogramme du solde CCP	118
B.3	Evolution du solde avec moyenne mobile et variance	119
B.4	Décomposition de la série temporelle du solde CCP	120
B.5	Autocorrélation de la série temporelle du solde CCP	120
B.6	Autocorrélation partielle de la série temporelle du solde CCP	121
B.7	Prévisions (LSTM) et valeurs réelles du solde CCP	121
C.1	Évolution du compte Trésor à travers le temps	123
C.2	Histogramme annuel du Trésor	123
C.3	Décomposition de la série avant différenciation	125
C.4	Décomposition de la série après différenciation	125
C.5	Autocorrélation de la série temporelle du compte Trésor	126
C.6	Autocorrélation partielle de la série temporelle du compte Trésor	126
C.7	Série temporelle original et prédictions	127
D.1	Évolution des Scellés BA à travers le temps	129
D.2	Autocorrélation de la série temporelle des scellés BA	130
D.3	Autocorrélation partielle de la série temporelle des scellés BA	130
D.4	Architecture d'un BI-LSTM	131
D.5	Série temporelle original et prédictions	132

# Liste des tableaux

3.2	Quotités des actifs à court terme	35
3.1	Quotités des passifs à court terme	35
5.2	Postes significatifs et leurs définitions	64
5.3	Statistiques descriptives pour la série temporelle des Mouvements de la caisse	68
5.4	Statistiques descriptives pour la série temporelle du Solde de la caisse	68
5.5	Evaluation des modèles appliqués à la caisse	97
5.6	Evaluation des modèles appliqués au CCP	97
5.7	Evaluation des modèles appliqués au Trésor	97
5.8	Evaluation des modèles appliqués aux scellés BA	98

# Liste des sigles et acronymes

<b>SGA</b>	<i>Société Générale Algérie</i>
<b>BA</b>	<i>Banque d'Algérie</i>
<b>ALM</b>	<i>Asset and Liability Management</i>
<b>BO</b>	<i>Business Object</i>
<b>DCO</b>	<i>Date Comptable</i>
<b>BAU</b>	<i>Business As Usual</i>
<b>IA</b>	<i>Intelligence Artificielle</i>
<b>RNN</b>	<i>Recurrent Neural Network</i>
<b>LSTM</b>	<i>Long Short Term Memory</i>
<b>GRU</b>	<i>Gated Recurrent Unit</i>
<b>GAN</b>	<i>Generative Adversarial Networks</i>
<b>CGAN</b>	<i>Conditional Generative Adversarial Networks</i>
<b>ARIMA</b>	<i>AutoRegressive Integrated Moving Average</i>
<b>RMSE</b>	<i>Root Mean Squared Error</i>
<b>MSE</b>	<i>Mean Squared Error</i>
<b>MAE</b>	<i>Mean Absolute Error</i>
<b>MAPE</b>	<i>Mean Absolute Percentage Error</i>



# Introduction générale

Le secteur bancaire est considéré comme la cheville ouvrière d'un pays. Il joue un rôle crucial dans l'économie du pays grâce à la mobilisation des ressources financières, l'intermédiation entre les acteurs économiques, et la stimulation de l'activité économique. Il est en constante évolution pour répondre aux besoins des différents agents économiques et s'adapter aux exigences réglementaires internationales. En Algérie, la Banque d'Algérie est chargée de la régulation et de la supervision du système bancaire algérien. L'un des plus gros défis des banques est la gestion des risques liés à leurs activités, risques qui peuvent avoir un impact significatif sur leur stabilité financière. Le risque de liquidité est au cœur de leur préoccupation, car une insuffisance de liquidités peut compromettre leur capacité à répondre à leurs exigences.

Un moyen d'anticiper ce risque est l'exercice de stress-test qui a été mis en place après la crise financière de 2008, plus connue sous le nom de la crise des 'subprimes'. La crise a été déclenchée par l'effondrement du marché immobilier américain, en particulier celui des prêts hypothécaires à risque (subprimes). Les prêts hypothécaires subprimes étaient accordés à des emprunteurs ayant une faible solvabilité, souvent avec des conditions de remboursement peu favorables. A cet effet, la Société Générale Algérie filiale du groupe de renommée mondiale Société Générale et le leader des banques privées sur le marché algérien souhaite mettre en place un outil de stress-test de sa trésorerie afin de tester sa résilience face à des scénarios de stress. Dans ces conditions, aura-t-elle assez de liquidités pour honorer ses engagements ? Si non, doit-t-elle se refinancer ? Le département ALM de la SGA nous a donné pour mission de concevoir et d'implémenter l'outil de stress-test dans le cadre de notre projet de fin d'études. En notre qualité de data scientists, notre objectif dans le cadre de ce projet :

- Concevoir et mettre en œuvre des modèles d'apprentissage automatique pour effectuer des prédictions précises sur l'évolution future de la trésorerie.
- Établir et intégrer des scénarios de stress à ces prévisions afin d'obtenir des projections de trésorerie adaptées aux situations de tension financière.

Afin d'atteindre notre objectif, nous avons organisé notre travail en 3 parties. Nous commencerons par une étude de l'existant où nous examinons en détail l'organisme d'accueil ensuite nous passerons à un diagnostic complet. Dans la partie état de l'art, nous aborderons les notions liées à notre domaine et les technologies avancées de l'IA appliquées à la finance pour ensuite passer à la partie pratique, à savoir l'élaboration du prévisionnel, les scénarios de stress et la conception de l'outil.

Première partie

État des lieux

# Introduction

Dans cette partie, nous consacrons le premier chapitre à l'organisme d'accueil, la SGA, en se concentrant sur ses structures organisationnelles clés, notamment la Direction financière et le Département ALM. Nous mettons en lumière les activités clés de l'ALM, établissant ainsi le contexte de notre étude.

Le deuxième chapitre aborde une évaluation détaillée de l'existant à la SGA, avec une attention particulière portée à la prévision de la trésorerie et au stress-test de cette dernière. Nous effectuons également un diagnostic externe à l'aide des analyses SWOT et PESTEL. À partir de là, nous définissons la problématique principale de notre étude et expliquons la méthodologie que nous avons choisie pour y répondre. Cette partie se termine par une conclusion qui fait le lien avec les parties suivantes de notre mémoire.

# Chapitre 1

## Organisme d'accueil

### Introduction

Le secteur bancaire est un secteur économique important qui offre des services financiers tels que la gestion de dépôts, les prêts, les investissements et les services de paiement. Les banques jouent un rôle crucial dans l'économie en fournissant des services financiers qui permettent aux particuliers et aux entreprises de gérer leurs finances et de réaliser des investissements. Dans cette partie de notre mémoire, nous allons nous concentrer sur l'état des lieux en nous penchant sur la Société Générale Algérie, banque privée leader en Algérie. Nous analyserons ses activités globales ainsi que son département d'Asset and Liability Management (ALM), également connu sous le nom de gestion de l'actif et du passif. L'objectif de cette section est de dresser un état des lieux détaillé de la Société Générale, en mettant l'accent sur ses activités, sa structure organisationnelle. Nous nous intéresserons également plus spécifiquement à son département d'ALM, qui joue un rôle clé dans la gestion des risques et de la performance de l'entreprise.

Les banques sont des entités dont l'activité consiste à recevoir des fonds remboursables du public et à octroyer des crédits.

### 1.1 Groupe Société Générale

Société Générale est une entreprise française multinationale de banque et de services financiers qui a été fondée en 1864. Son siège social est situé à Paris, en France. Elle est l'une des plus anciennes et des plus grandes banques en France et dans le monde, avec des opérations dans de nombreux pays à travers le monde.

La société opère dans trois principaux segments d'activité : la banque de détail et les services financiers spécialisés, la banque d'investissement et la gestion d'actifs, et les marchés mondiaux. Dans le segment de la banque de détail, elle offre des services bancaires de base aux particuliers et aux petites entreprises, tels que les comptes d'épargne, les comptes courants, les prêts personnels, les cartes de crédit, les services bancaires en ligne, et les produits d'assurance. Elle propose également des services financiers spécialisés tels que la gestion de patrimoine, la banque privée et la location de voitures.

En 2021, Société Générale comptait environ 31 millions de clients dans le monde entier, dont 8,4 millions en France. Ces clients incluent des particuliers, des petites et moyennes entreprises, des grandes entreprises et des institutions financières. Elle employait environ

133 000 collaborateurs dans le monde entier, dont environ 52 000 en France. Ces collaborateurs travaillent dans différents domaines tels que la banque de détail, la banque d'investissement, la gestion d'actifs, les opérations de back-office, les services informatiques et autres.

## 1.2 Société Générale Algérie

Société Générale Algérie est une filiale de la Société Générale. Elle est présente en Algérie depuis maintenant 24 ans.

Société Générale Algérie est le premier acteur bancaire à capitaux privés du pays en nombre d'agences et en clients offrant une large gamme de produits et services financiers aux particuliers, aux petites et moyennes entreprises et aux grandes entreprises. Elle propose notamment des services de banque de détail tels que les comptes courants, les cartes de crédit, les prêts personnels, les prêts immobiliers et les services bancaires en ligne. Elle propose également des services de banque d'investissement, tels que les opérations de fusion et d'acquisition, les introductions en bourse, le financement de projets et la gestion d'actifs.

Société Générale Algérie dispose d'un réseau de 103 agences réparties sur 31 Wilayas, ainsi que d'un site web et d'une application mobile pour permettre à ses clients de gérer leurs comptes et leurs transactions bancaires en ligne. La filiale emploie environ 1 500 collaborateurs locaux et compte 450 000 clients.

Voici l'organigramme de la SGA :



FIGURE 1.1 – Organigramme de la SGA

## 1.3 Direction financière

La direction financière de la SGA est chargée de gérer toutes les activités financières de la banque, y compris la gestion des risques, la planification et le contrôle budgétaire, la comptabilité, la fiscalité et le financement. Elle a un rôle crucial dans l'élaboration de stratégies pour assurer la croissance et la pérennité de la banque.

Nous vous présentons l'organigramme de la direction financière :

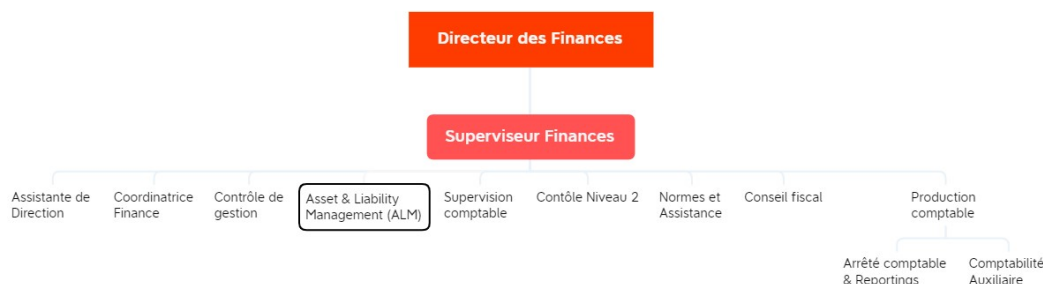


FIGURE 1.2 – Organigramme de la direction financière de la SGA

## 1.4 Département ALM

Le département Asset and Liability Management est une structure rattachée à la direction financière de la SGA dont la fonction principale est de gérer les risques de liquidité, de taux d'intérêt et de change. L'ALM est responsable du contrôle de premier niveau. Elle est chargée d'effectuer la mesure périodique des risques structurels, du reporting, de l'élaboration de proposition de couverture et la mise en œuvre de décisions prises par le comité ALM.

Voici l'organigramme de l'ALM :



FIGURE 1.3 – Organigramme de l'ALM

## 1.5 Domaine de l'ALM

### 1.5.1 Passifs et Actifs

Le bilan bancaire est un état financier qui présente la situation patrimoniale d'une banque à un moment donné. Il se compose de deux principales parties : les actifs et les passifs. Les actifs d'une banque sont les mobilisations investies pour générer des revenus.

Ils comprennent Les crédits, les biens immobiliers, les investissements ( bon du trésor, obligations). Ils sont essentiels pour garantir la stabilité financière de l’institution et sa capacité à fournir des services bancaires aux clients. Les passifs d’une banque sont une obligation financière que doit rembourser la banque à une autre partie. Il représente une dette ou une responsabilité envers des tiers, qui doit être payée ou satisfaite à une date ultérieure comme les dépôts clients, les fonds propres et les emprunts. Nous expliquerons en **Annexe** les principaux postes d’un bilan bancaire. La figure 1.4 présente de manière simplifiée le bilan d’une banque :

<b>SCHÉMA DE BILAN COMPTABLE</b>	
<b>ACTIF</b>	<b>PASSIF</b>
<p><b>Actifs immobilisés</b></p> <ul style="list-style-type: none"> <li>• Immobilisations corporelles</li> <li>• immobilisations incorporelles</li> <li>• immobilisations financières</li> </ul>	<p><b>Capitaux propres</b></p> <ul style="list-style-type: none"> <li>• capital social</li> <li>• réserves</li> <li>• résultats compte courant d'associé</li> </ul>
<p><b>Actifs circulants</b></p> <ul style="list-style-type: none"> <li>• stocks</li> <li>• créances clients</li> <li>• valeurs mobilières</li> <li>• disponibilités</li> </ul>	<p><b>Dettes</b></p> <ul style="list-style-type: none"> <li>• financières</li> <li>• d'exploitation - fournisseurs</li> <li>• diverses</li> </ul>
<p><b>Trésorerie active</b></p>	<p><b>Trésorerie passive</b></p>
<p><b>Total bilan à l'actif</b></p>	<p><b>Total bilan au passif</b></p>

FIGURE 1.4 – Illustration du bilan de gestion

## 1.5.2 Activités de l’ALM

Nous allons maintenant parler des nombreuses activités de l’ALM au sein de la banque. Voici les principales missions de l’ALM :

- Produire les reportings ALM des entités de son périmètre, tout en s’assurant du respect des normes Groupe sur les écoulements.
- Mesurer, analyser les indicateurs sur les risques structurels et identifier les déséquilibres.
- Préparer les comités ALM (ALCO Régional et Entités) et Comités de Trésorerie.
- Comités des Prix : grilles tarifaires, présentation des Taux de Cession Interne (TCI), mesures de rentabilité.
- Proposer des modalités de couverture de risques structurels de taux, change et liquidité (Préconisations en Comité ALCO).
- Anticiper sur les évolutions des positions ALM à travers des simulations (études d’impact suite à un changement de modèle, stress test, suivi du deal pipe), et participer à l’amélioration des modèles d’écoulement. Modélisation des écoulements des actifs et passifs non échéancés.

### Comité ALM :

Le comité ALM est un comité de gouvernance traitant des problématiques de gestion des risques structurels de liquidité, de taux et de change. Ce comité se réunit de manière périodique (généralement trimestriellement) et décide des actions de couverture ou de placement à mettre en œuvre, conformément aux orientations stratégiques sur la gestion

de liquidité.

## **Conclusion**

Dans ce chapitre, nous avons présenté l'organisme d'accueil SGA, le domaine de l'ALM et les notions clés liées à ce dernier dont le risque bancaire, concept indispensable pour le business understanding.



# Chapitre 2

## Étude de l'existant

### Introduction

Dans cette section, nous allons mettre en évidence les principaux facteurs internes et externes qui peuvent influencer la mise en place d'un outil de stress-test de la trésorerie de la Société Générale. À la fin de cette partie, nous allons formuler la problématique qui constituera le fil conducteur de notre mémoire puis présenter la méthodologie de travail adoptée. Cette méthodologie nous permettra de répondre de manière rigoureuse et structurée à la première partie de notre problématique, à savoir, le prévisionnel de la trésorerie.

### 2.1 Le prévisionnel de la trésorerie au niveau de la SGA :

Le prévisionnel de la trésorerie au sein de la SGA est présenté dans une section d'un tableau d'un fichier Excel comportant les soldes prédits des différents postes significatifs.

Un postes significatifs représente les avoirs de la banque qui sont en lien avec le solde BA. Le solde BA représente le solde de la SGA au sein de la banque régulatrice en Algérie "La Banque d'Algérie".

Le premier objectif de ce prévisionnel est de calculer le solde BA à partir des prévisions effectuées sur les postes les plus significatifs qui le constituent. Le second est de prédire le coefficient de liquidité.

*Diagnostic 1 : Or que les soldes prédits sont peu fiables car ils n'ont pas été calculés convenablement : Les valeurs du solde de chacun des postes significatifs sont simplement remplacées par les mêmes valeurs passées ou par la moyenne des données historiques. Les résultats sont donc biaisés car les variations potentielles qui pourraient survenir dans l'avenir sont souvent négligées.*

### 2.2 Stress-test de la trésorerie :

#### 2.2.1 Définition :

Un stress-test dit aussi test de résistance bancaire est un exercice visant à simuler des conditions financières "stressées" afin d'anticiper et d'évaluer la résilience d'une institution bancaire face à des scénarios économiques adverses.

## 2.2.2 Stress-test de la trésorerie au sein de la SGA :

Le stress-test de la trésorerie de la SGA est construit à partir de divers scénarios. À titre d'exemple la non possibilité d'avoir recours au refinancement auprès de la BA, la fuite de 75 % des dépôts à terme Corporate Retail ou encore la limite de l'emprunt interbancaire. Ces scénarios sont ensuite appliqués aux **données du prévisionnel de la trésorerie**.

*Diagnostic 2 : Vu que le stress-test de la trésorerie est appliqué à des données biaisées ( Diagnostic 1), le résultat de ce dernier sera donc affecté.*

## 2.3 Diagnostic externe :

Dans le cadre du diagnostic externe de notre étude, nous procéderons à une analyse approfondie de la SGA en utilisant deux outils d'analyse essentiels : l'analyse SWOT (Forces, Faiblesses, Opportunités, Menaces) et l'analyse PESTEL (Politique, Économique, Social, Technologique, Environnemental, Légal). Ces méthodes nous permettront d'évaluer de manière holistique les aspects internes et externes qui peuvent influencer la mise en œuvre de l'outil de stress-test de la trésorerie de la SGA.

### 2.3.1 Analyse SWOT

#### Forces : (interne)

- La Société Générale est une banque de renommée internationale, avec une solide expérience dans le domaine de la gestion financière.
- La direction financière de la SGA dispose d'une structure de gestion actif-passif ALM, ce qui facilite l'implémentation d'un outil de stress test de la trésorerie.

#### Faiblesses : (interne)

- La refonte d'un nouvel outil peut prendre du temps significatifs et nécessiter des investissements importants.
- Le personnel de la SGA peut avoir besoin de formations pour apporter des modifications à l'outil de stress test.

#### Opportunités : (externe)

- L'implémentation d'un outil de stress test de la trésorerie peut améliorer la compétitivité de la SGA sur le marché de la gestion financière.
- L'outil de stress test de la trésorerie permet une aux exigences réglementaires en matière de gestion des risques financiers.
- La mise en place de l'outil de stress test de la trésorerie peut améliorer la qualité de l'analyse de la trésorerie et de la gestion des risques financiers.

#### Menaces : (externe)

- Les concurrents peuvent disposer d'outils similaires ou plus avancés pour la gestion de la trésorerie et des risques financiers.

- Les changements réglementaires peuvent rendre l’outil de stress test de la trésorerie obsolète et il peut nécessiter des mises à jour coûteuses.

### **2.3.2 Analyse PESTEL**

#### **Politique**

- La réglementation financière algérienne exige que les institutions financières mettent en place des systèmes de gestion des risques financiers.
- L’instabilité politique peut affecter la stabilité économique de l’Algérie et avoir un impact sur la performance financière de la Société Générale Algérie.

#### **Économique**

- La croissance économique de l’Algérie peut offrir des opportunités pour la Société Générale Algérie de se développer.
- Les fluctuations des taux de change peuvent avoir un impact sur les opérations de la Société Générale Algérie.

#### **Socioculturel**

- La culture algérienne peut influencer les préférences des clients en matière de services financiers.
- Les différences culturelles peuvent rendre la formation du personnel plus complexe.

#### **Technologique**

- L’outil de Stress Test nécessite des compétences technologiques spécialisées pour son développement et son maintien.
- L’adoption de nouvelles technologies peut créer des opportunités pour la SGA de se différencier de ses concurrents.

#### **Environnemental**

- La Société Générale Algérie, en tant que filiale à 100% du Groupe Société Générale, est soumise non seulement à la réglementation locale, mais aussi à la taxonomie européenne en matière environnementale. Cela n’entraîne pas nécessairement de modifications majeures des processus ni d’augmentation des coûts, mais cela reste une considération importante pour l’image et l’éthique de l’entreprise. En respectant ces réglementations et normes, la SGA démontre son engagement envers les principes de la Responsabilité Sociétale des Entreprises (RSE) et des critères Environnementaux, Sociaux et de Gouvernance (ESG), ce qui peut renforcer sa réputation et sa position sur le marché.
- La prise de conscience croissante du public vis-à-vis du changement climatique pourrait entraîner une pression accrue sur les banques pour qu’elles dirigent davantage de financements vers des projets de Finance à Impact Positif Durable (Sustainable Positive Impact Finance - SPIF), ce qui pourrait influencer la composition de leur portefeuille de prêts.

## Légal

- La Banque d’Algérie, comme toutes les banques centrales, est soumise à des réglementations légales strictes. Par conséquent, toute modification de ces réglementations peut avoir un impact sur l’outil de stress test.
- Les lois sur la protection des données et la confidentialité peuvent également avoir un impact sur la manière dont les données sont collectées.

## 2.4 Enoncé de la problématique

Après avoir réalisé notre diagnostic, nous avons clairement identifié l’importance des prévisions pour les activités de l’ALM ainsi que le besoin de concevoir un outil permettant d’effectuer l’exercice du stress-test. Cet exercice, essentiel pour la Société Générale Algérie, comme pour toute banque souhaitant respecter ses engagements vis-à-vis de ses clients, implique la réalisation de divers tests pour mesurer sa résilience face à des scénarios de crise.

Notre projet vise, dans une première phase, à moderniser les prévisions initialement établies par la structure gestion actif-passif (ALM), en partenariat avec la Direction du Contrôle de Gestion (CDG). Nous envisageons d’utiliser des modèles d’apprentissage automatique avancés pour améliorer la précision des prévisions. Comment l’apprentissage automatique peut-il aider à moderniser les prévisions de trésorerie ?

La deuxième phase consistera à déterminer les scénarios de stress et à les appliquer aux prévisions. Quels seront nos scénarios de stress ? Comment seront-ils appliqués aux prévisions BAU ? Quelle sera la forme finale de notre outil de stress-test ?

En résumé, le projet s’articulera autour de deux phases principales. La première phase consistera à élaborer des prévisions à l’aide de techniques d’apprentissage automatique, fournissant ainsi des estimations précises et fiables. La deuxième phase mettra l’accent sur l’élaboration et l’intégration de scénarios de stress dans ces prévisions. L’objectif final étant d’évaluer l’impact du stress sur l’ensemble des actifs de trésorerie contribuant au solde BA.

## 2.5 Méthodologie de travail

Afin de mieux cerner la problématique, nous avons décidé d’adopter une méthodologie CRISP-DM( Cross-Industry Standard Process for Data Mining). La CRISP-DM est un cadre standard utilisé pour guider les projets data en exploration de data ou Data Mining.

On peut remarquer depuis la figure [2.1](#) que cette méthode est agile et itérative. Nous présentons maintenant les 6 étapes clés d’une méthodologie CRISP-DM :

1. **Compréhension du problème** : Cette étape consiste à définir clairement le problème à résoudre, les objectifs à atteindre, les données disponibles et les contraintes éventuelles.
2. **Collecte de données** : Il s’agit de collecter les données nécessaires pour résoudre le problème. Cette étape implique souvent la collecte de données brutes à partir de

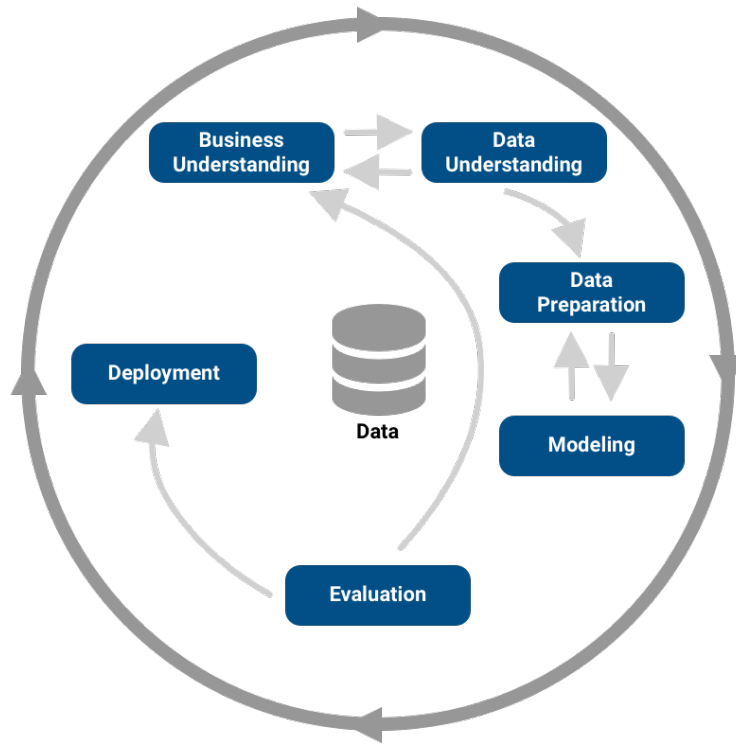


FIGURE 2.1 – Cycle de vie de la méthodologie CRISP-DM

différentes sources.

3. **Préparation des données** : Les données collectées doivent être nettoyées, transformées et structurées pour être utilisées pour l'analyse. Cette étape peut également inclure l'élimination des données redondantes, la création de nouvelles variables et la normalisation des données.
4. **Modélisation** : À cette étape, différents modèles sont développés pour répondre aux objectifs de l'analyse de données. Les modèles peuvent être des modèles prédictifs, des modèles descriptifs ou des modèles exploratoires.
5. **Évaluation** : Les modèles développés doivent être évalués pour déterminer leur qualité et leur pertinence par rapport aux objectifs du projet.
6. **Déploiement** : Cette dernière étape consiste à mettre en œuvre les résultats de l'analyse de données pour résoudre le problème initial. Cela peut impliquer la mise en place d'un système de prise de décision automatisé ou la présentation des résultats aux parties prenantes.

## Conclusion

Nous avons, dans ce chapitre, abordé la problématique, sa naissance et en quoi notre solution contribuerait à l'amélioration de la gestion de la liquidité et surtout de la trésorerie. Enfin, nous avons présenté la méthodologie de travail qu'on allait suivre dans notre PFE.

# Deuxième partie

## Etat de l'art

# Introduction

La deuxième partie de notre PFE est dédiée à l'état de l'art et à la revue bibliographique. Nous allons, en premier lieu, analyser puis synthétiser les connaissances liées au risque bancaire et à la liquidité. En deuxième lieu, nous allons fournir une revue approfondie sur les stress-tests et les modèles de machine learning les plus adaptés à ces derniers.

# Chapitre 3

## Liquidité et risque bancaire

### Introduction

La liquidité et les risques bancaires sont deux concepts étroitement liés qui revêtent une importance cruciale pour le bon fonctionnement du système financier. Nous allons, dans cette partie, découvrir les concepts liés à ces derniers afin de mieux cerner le domaine :

- Qu'est-ce-qu'un risque bancaire ? Quels sont ses types ?
- Qu'en est t-il de la liquidité ? Pourquoi ce concept est important pour les banques ?

### 3.1 Définition du risque bancaire

Le cadre de référence de l'Autorité des Marchés Financiers a défini le risque comme étant : «La possibilité qu'un événement survienne et dont les conséquences seraient susceptibles d'affecter les personnes, les actifs, l'environnement, les objectifs de la société ou sa réputation.»

Le risque bancaire est donc la possibilité qu'une banque subisse des pertes financières en raison d'événements imprévus ou de circonstances exceptionnelles. Il existe différents types de risques bancaires : le risque de taux, le risque de change, le risque de liquidité dit risques structurels et les autres risques.

### 3.2 Les différents risques bancaires

Il existe différents risques dans le secteur bancaire :

#### 3.2.1 Risques structurels

Le risque est dit structurel s'il a trait à la «structure » de la banque, qu'on observe notamment en analysant le contenu de son bilan comptable

Les stress tests est appliqué aux risques structurels, il vise à identifier les vulnérabilités potentielles et les points faibles de la banque. Voici les différents types de risques structurels :

1. **Risque de taux** : Jacqueline Ducret (2002) a défini le risque de taux d'intérêt d'un établissement financier comme le risque encouru en cas de variation des taux d'intérêt du fait de l'ensemble des opérations de bilan et hors-bilan, à l'exception, le cas échéant, des opérations soumises aux risques de marché.



2. **Risque de change** : Le risque de taux de change est associé aux variations des taux de change entre différentes devises.
3. **Risque de liquidité** : Dans le cadre de notre projet, on s'intéresse à ce risque. Une banque est en crise de liquidité si le comportement d'un certain nombre de déposants, qui pour une raison ou une autre, décident de retirer leurs dépôts de manière non anticipée. Dans le cas où la banque ne peut pas répondre à cette demande imprévisible, elle peut être obligée de liquider une partie de ses actifs. (Pierre Henry Chiappori, 1997)

“C’est la possibilité que, sur une période donnée, une banque soit incapable d’honorer ses engagements immédiats. Ce risque présente deux composantes mesurant respectivement un effet « quantité » et un effet « prix » : le montant aléatoire des flux sortants de monnaie ou de trésorerie (décaissements) et le coût aléatoire d’obtention de la liquidité de financement de différentes sources (liquidité de marché, liquidité banque centrale, dépôts à vue).”(Drehmann Nikolaou, 2009).

### 3.2.2 Les autres risques

Les risques non structurels sont liés aux facteurs externes et aux événements spécifiques qui peuvent affecter les performances d’une banque. Ces risques ne sont pas directement liés aux activités de base de la banque. Les principaux risques non structurels sont les suivants :

1. **Risque de marché** : “Le risque de marché est le risque de perte sur les positions du bilan relatif à l’ensemble des avoirs et les dettes d’une institutions financières et du hors bilan notamment les sommes engagées non payées ou reçus à la suite des variations des prix sur le marché.” Christian Gouriéroux, André Tiomo” (2007).
2. **Risque de contrepartie** :Le risque de contrepartie est la possibilité que l’une des parties à une transaction financière ne remplisse pas ses obligations contractuelles.
3. **Risque de réputation** : Le risque de réputation est le risque que l’image et la réputation de la banque soient endommagées par des événements tels que des scandales, des violations de la réglementation, des fautes professionnelles ou des comportements inappropriés.
4. **Risque opérationnel** : Le Comité de Bâle II définit le risque opérationnel comme « le risque de pertes provenant de processus internes inadéquats ou défectueux, de personnes et systèmes ou d’événements externes ».

### 3.3 Les différents types de liquidités

Il existe plusieurs types de liquidités, nous citons les principales :

1. **Liquidité de financement** : Plusieurs autres définitions liées à la liquidité de financement ont vu le jour : Selon la commission de banque française, la liquidité de financement est définie comme “ sa capacité à rembourser à bonne date et à coût raisonnable ses exigibilités, en pouvant mobiliser notamment très rapidement ses avoirs lorsque le besoin s’en présente. Les sources de liquidité les plus communément reconnues sont la trésorerie (le cash) ou les titres qui peuvent être immédiatement

convertis en cash– et à un coût raisonnable –, soit auprès d’une banque centrale (sous forme de «monnaie banque centrale », liquide par définition) soit sur un marché.” Quant au comité de Bâle, il a défini la liquidité bancaire comme “La capacité d’une institution de financer l’expansion de ses avoirs et d’honorer ses obligations aux échéances prévues, sans enregistrer de pertes inacceptables.” Notre étude portera sur la liquidité de financement.

2. **Liquidité banque centrale** : “C’est la capacité de la banque centrale à fournir la liquidité dont le système financier a besoin. Le risque de liquidité inhérent dans ce cas est presque inexistant parce que la banque centrale peut toujours fournir de la monnaie banque centrale dont elle a le monopole d’émission.” (Azzouzi Idrissi Madies, 2012)
3. **Liquidité de marché** : Le risque de marché est le risque de ne plus pouvoir trouver la liquidité nécessaire sur les marchés. Il existe deux types de liquidité de marché : (1) la liquidité obtenue sur le marché interbancaire, où la liquidité se prête seulement entre banques, et (2) la liquidité obtenue sur les marchés d’actifs, où les vendeurs d’actifs reçoivent de la liquidité, en contrepartie des actifs cédés, de la part des acheteurs.

### 3.4 Coefficient de liquidité :

Le ratio de liquidité d’une entreprise permet de mesurer la capacité de celle-ci à faire face à ses obligations à court terme en comparant les actifs à court terme inscrits au bilan avec les passifs à court terme. Le coefficient de liquidité doit être supérieur à 100%. Dans le cas contraire, l’entreprise est en situation d’illiquidité.

Il existe deux méthodes de calcul :

$$\begin{aligned} & (\text{actifs circulants} - \text{stocks} / \text{dettes à court terme de moins de 1 an}) \times 100 \\ & (\text{créances clients} + \text{disponibilités} / \text{dettes à court terme de moins de 1 an}) \times 100 \end{aligned}$$

### 3.5 Supervision bancaire et comité de Bâle

#### 3.5.1 Définition de la supervision bancaire :

La supervision bancaire est l’activité permanente des autorités de contrôles et de régulation qui vise à protéger les déposants et les opérateurs économiques ainsi qu’à prévenir les risques bancaires découlant d’une mauvaise gestion ou des engagements trop importants de la part des banques et établissements financiers. (Banque d’Algérie, 2010).

#### 3.5.2 Présentations du comité :

Le comité de Bâle est un organisme de proposition et de réflexion sur la supervision bancaire qui fut créé en 1974 par le groupe des dix. Il est hébergé par la Banque des règlements internationaux à Bâle. Il rassemble aujourd’hui 28 pays.

#### 3.5.3 Missions de Bâle :

Les objectifs principaux du comité sont :

- Établir des exigences minimales de contrôle prudentiel que les banques se doivent de respecter.
- Promouvoir les meilleures pratiques bancaires et de surveillance.
- Renforcer la stabilité et la solidité du système bancaire.

### 3.6 Besoin en liquidité :

Les besoins de liquidité sont déterminés par la construction d'une échelle d'échéances comprenant les flux prévisionnels de trésorerie entrants et sortants sur une série d'intervalles de temps spécifiés. La différence entre les flux entrants et les flux sortants sur chaque exercice (c'est-à-dire l'excédent ou le déficit de fonds) donne un point de départ à partir duquel on peut mesurer l'excédent ou le déficit de liquidité futur de la banque à un moment donné.

**Trésorerie bancaire** La trésorerie d'une banque représente les liquidités disponibles à un moment donné  $t$ . Elle inclut les entrées et les sorties de fonds de la banque, telles que les dépôts et les retraits des clients, les remboursements de prêts et les investissements réalisés. Dans le contexte algérien, elle se résume à son solde au niveau de la Banque d'Algérie. La trésorerie permet donc à la banque de faire face à ses obligations financières immédiates.

### 3.7 Situation d'illiquidité et méthodes de refinancement des banques

Si la banque est en situation d'illiquidité, c'est-à-dire qu'elle ne peut pas honorer ses engagements si il y'a par exemple un nombre conséquent de retrait ou une baisse de dépôts, elle peut se refinancer de la façon suivante afin de se couvrir du :

#### Marché interbancaire :

Le marché interbancaire est un marché où les banques et les institutions financières peuvent se refinancer pour leurs besoins de liquidités à court terme (horizon de moins d'un an) en empruntant et prêtant des fonds entre elles. Il est important de rester actif sur ce marché car il permet aux banques de répondre à leurs besoins en liquidités immédiates tout en maintenant leurs réserves mais aussi il leur donne la possibilité de diversifier leur portefeuille de prêts. Ce marché est régulé par les banques centrales et régulateurs financiers pour assurer son intégrité.

#### Banque Centrale :

Les banques peuvent emprunter des fonds auprès de la banque centrale pour se refinancer à court terme. Elle peut imposer les termes de l'emprunt ou des garanties. La banque centrale achète des titres tels que les Bons du trésor depuis l'open market afin de réguler la politique monétaire. En Algérie, les opérations d'open market sur le marché monétaire sont effectuées à l'initiative de la BA. Voici les différents types d'opérations qui se font sur l'open market :

- Les opérations principales de refinancement.
- Les opérations de refinancement à plus long terme.

- Les opérations de réglage fin.
- Les opérations structurelles.

La banque peut également se refinancer auprès de la BA grâce aux facilités permanentes. L'un des types de facilités permanentes sont les facilités de prêt marginal.

les facilités de prêt marginal est une opération par laquelle une banque peut obtenir de la BA, contre la présentation d'effets éligibles, des liquidités à 24 heures à un taux prédéfini.

### **Opérations de REPO :**

Les opérations de repo (sales and repurchase agreement) dit pension de titres en français sont des opérations exigeant des garanties pour se couvrir dans le cas où l'emprunteur ne peut pas rembourser son prêt comme l'achat de titres de créances ou d'obligations temporairement contre de la liquidité, l'emprunteur s'engage à les racheter à un prix convenu à l'avance.

### **Titrisation :**

La titrisation est une technique financière permettant de se financer en transformant des actifs tels que des prêts hypothécaires ou des créances commerciales en dettes négociables. Son mécanisme consiste à :

Regrouper un portefeuille de titres similaires → Le portefeuille est cédé à une structure ad-hoc → Les titres sont placés auprès d'investisseurs dans des marchés financiers → Réception de paiements de créances de ces actifs en faveur des investisseurs. Cependant, la titrisation n'est pas utilisée au niveau de la SGA.

## **3.8 Réglementation BA sur le contrôle de risque de liquidité, le ratio des fonds propres et des ressources permanentes :**

La BA est l'organe de régulation du secteur bancaire. Elle est chargée d'établir les règles sur la supervision des banques et des établissements financiers algériens.

Elle adopte un dispositif réglementaire inspiré des recommandations du comité de Bâle pour faire face au risque de liquidité. En voici les plus pertinentes :

- Les banques doivent avoir à un moment  $t$  assez de liquidités pour répondre à leur engagement. (Article 2, Règlement N°2011-04 du 24 Mai 2011).
- Les banques doivent fournir un rapport trimestriel contenant la somme des actifs disponibles et à court terme et des engagements donnés qui n'est rien d'autre que le coefficient de liquidité minimum du mois à venir et des deux derniers mois en plus d'un coefficient d'observation pour les 3 mois suivant la date du dernier arrêté (Ratio de liquidité générale). (Article 3 et 4, Règlement N°2011-04 du 24 Mai 2011)
- Chaque élément du numérateur (actifs) et du dénominateur (passifs) à une quotité déterminée par la BA. (Article 4, instruction N°07-2011 du 21 décembre 2011).
- Les banques et les établissements financiers mettent en place des plans d'urgence formalisés qui leur permettent de se préparer et de faire face à des situations de crise. (Article 24, Règlement N°2011-04 du 24 Mai 2011)
- Les banques se doivent de respecter les réserves obligatoires imposées par la Banque d'Algérie pour réguler l'offre de monnaie dans l'économie.

<b>Pondération</b>	<b>Actif</b>
100%	Comptes créditeurs à vue en dinars, emprunts sur le marché interbancaire, emprunts obligatoires et subordonnés.
70%	Les dépôts à terme (en dinars) et les bons de caisse.
30%	Les livrets d'épargne banque.
25%	Les dépôts à vue (en dinars) des entreprises.
20%	Les autres dépôts à vue (en dinars).

TABLE 3.2 – Quotités des actifs à court terme

<b>Pondération</b>	<b>Passif</b>
100%	Avoirs en caisse, avoirs en comptes devises fonds propres, avoirs à vue en dinars, dépôts auprès du Trésor Public et Algérie Poste, comptes débiteurs à vue, prêt sur le marché, bon du trésor, titres participatifs du Trésor.
75%	Encours à couvrir consentis à la clientèle sous forme de crédit d'exploitation, de crédits d'investissement, d'opérations de crédit-bail et de location simple.
60%	Obligations et autres valeurs mobilières à revenu fixe émises par l'État et faisant l'objet d'une cotation sur un marché officiel.
50%	Les engagements de financement reçus des banques et établissements financiers agréés en Algérie.
10%	Les actions et autres titres à revenu variable faisant l'objet d'une cotation sur un marché officiel algérien.

TABLE 3.1 – Quotités des passifs à court terme

## Conclusion

La gestion des risques bancaires et la garantie d'une liquidité suffisantes sont des éléments cruciaux pour la stabilité et la solidité du système financier. Les réglementations et la supervision bancaire jouent un rôle essentiel dans la prévention des risques et la protection des intérêts des parties prenantes. Les banques doivent continuellement évaluer et gérer leurs risques, ainsi que maintenir des niveaux appropriés de liquidité pour faire face aux situations imprévues et assurer leur viabilité à long terme.

# Chapitre 4

## L'IA pour le Stress Test

### 4.1 Stress-test

#### 4.1.1 Introduction

Un stress test est une technique qui consiste à soumettre un système, une entreprise, un produit, ou un individu à des conditions de stress extrêmes ou inhabituelles, dans le but d'évaluer sa capacité à résister à ces conditions et à maintenir un niveau de performance acceptable.

Les stress tests peuvent être utilisés dans de nombreux domaines, tels que la finance, la médecine, l'ingénierie, la technologie, l'aviation, etc. Ils peuvent également être utilisés pour évaluer les risques et les vulnérabilités d'un système ou d'une entreprise face à des situations imprévues ou des changements soudains dans l'environnement.

Ils peuvent être effectués de différentes manières, selon le domaine et les objectifs visés. Par exemple, dans le domaine financier, les stress tests peuvent être effectués en utilisant des modèles économiques et des scénarios de crise pour évaluer les risques et les vulnérabilités des banques et des institutions financières. Dans le domaine de la médecine, les stress tests peuvent être effectués en mesurant la capacité d'un cœur à fonctionner dans des conditions de stress, ou en évaluant la résistance d'un patient à des substances ou à des situations de stress.

Les stress tests peuvent être effectués en utilisant différentes méthodes, telles que la simulation informatique, les tests sur terrain, les tests en laboratoire, etc. Les résultats des stress tests peuvent aider à identifier les points faibles d'un système ou d'un individu, ce qui peut conduire à des améliorations et à des mesures préventives pour éviter les échecs ou les futurs accidents.

#### 4.1.2 Historique

Les stress tests ne sont pas un concept nouveau et ont été utilisés dans diverses formes depuis des siècles. Par exemple, dans l'Antiquité, les soldats romains étaient soumis à des marches forcées pour évaluer leur condition physique et leur capacité à supporter le stress de la guerre.

Dans le domaine de l'ingénierie, les stress tests ont été utilisés depuis longtemps pour évaluer la résistance des matériaux et des structures. Par exemple, les ponts, les avions et les navires sont soumis à des tests de stress pour s'assurer qu'ils peuvent supporter des conditions de charge extrêmes sans se briser.

En médecine, les stress tests ont également été utilisés depuis longtemps pour évaluer

la fonction cardiaque et la capacité d'un patient à tolérer l'exercice. Les tests de stress cardiaque impliquent généralement de faire faire de l'exercice à un patient sur un tapis roulant tout en surveillant son rythme cardiaque pour évaluer la capacité de son cœur à fonctionner sous stress.

## **En finance**

Les stress tests en banque ont été introduits pour la première fois aux États-Unis à la suite de la crise financière de 2008. En 2009, le gouvernement américain a mandaté la Federal Reserve pour effectuer des stress tests sur les 19 plus grandes banques du pays afin de déterminer leur capacité à résister à une crise économique.

Le premier stress test réalisé en 2009 a révélé que certaines banques avaient besoin de fonds supplémentaires pour faire face à des scénarios de crise. Cela a conduit à une injection de capital dans le système bancaire pour renforcer la solidité financière des banques.

Depuis lors, les stress tests sont devenus une partie intégrante de la réglementation bancaire dans de nombreux pays. Les banques doivent soumettre régulièrement des rapports de stress tests aux autorités de réglementation financière, qui évaluent la capacité des banques à résister aux chocs économiques et à maintenir la stabilité financière.

Les stress tests ont évolué au fil des ans pour inclure une gamme plus large de scénarios économiques et pour être plus sophistiqués dans leur approche. Ils sont considérés comme un outil essentiel pour renforcer la résilience des banques et pour prévenir les crises financières.

### **4.1.3 Stress tests au sein des banques**

Les stress tests en banque sont des évaluations régulières qui visent à tester la solidité financière et la capacité de résilience des banques et des institutions financières en cas de scénarios économiques défavorables. Les régulateurs financiers exigent des banques qu'elles effectuent des stress tests pour s'assurer qu'elles disposent d'un coussin de capital suffisant pour faire face aux risques potentiels.

Les stress tests en banque simulent des scénarios économiques défavorables, tels qu'une crise financière, une récession économique, une baisse de la valeur des actifs, une hausse des taux d'intérêt, etc. Les banques sont évaluées sur la base de leur capacité à maintenir leur solvabilité et leur liquidité en cas de scénarios de stress.

Les résultats des stress tests sont utilisés par les régulateurs financiers pour évaluer la solidité financière des banques et déterminer si elles doivent prendre des mesures supplémentaires pour renforcer leur bilan ou leur capital. Les résultats des stress tests sont également utilisés par les investisseurs pour évaluer la santé financière des banques et décider des investissements réalisés.

### **4.1.4 Catégories**

**Selon le meneur du test :**

- **Bottom-Up** : le test est mené par la banque.

- **Top-Down** : le test est mené par l'autorité monétaire, la banque ne fait que fournir ses données.

#### **Selon la provenance du choc :**

- **Idiosyncratique** : le choc est de provenance interne.
- **Systémique** : le choc est de provenance externe.

### **4.1.5 Typologie**

#### **Test de sensibilité :**

Ce type de stress test évalue l'impact de certains facteurs sur la valeur des actifs et les résultats financiers d'une banque. Les paramètres peuvent inclure les taux d'intérêt, les taux de change, les prix des matières premières ou les niveaux de volatilité des marchés financiers. Les tests de sensibilité sont souvent utilisés pour évaluer les risques de marché ou les risques de taux d'intérêt.

#### **Test inversé :**

Contrairement aux tests de sensibilité, le test inversé évalue les changements nécessaires pour que la banque atteigne un certain niveau de résilience financière. Par exemple, si une banque souhaite maintenir un ratio de fonds propres minimum de 10%, un test inversé pourrait être effectué pour déterminer les scénarios économiques qui nécessiteraient une augmentation de capital pour atteindre ce niveau. Selon le Comité de Bâle : "Les stress tests inversés sont menés à partir d'un résultat connu de stress test et se demande ensuite quels événements pourraient mener à un tel impact sur la banque."

#### **Analyse de scénarios :**

Ce type de stress test implique la création de scénarios économiques hypothétiques et l'évaluation de leur impact sur les résultats financiers d'une banque. Les scénarios peuvent inclure des événements économiques tels qu'une récession, une hausse des taux d'intérêt ou une crise financière. L'analyse de scénarios peut être utilisée pour évaluer la capacité d'une banque à résister à des événements économiques extrêmes et à maintenir sa solvabilité.

Dans tous les cas, les stress tests nécessitent une modélisation et une simulation rigoureuses pour évaluer l'impact de différents scénarios économiques sur la banque et pour déterminer les mesures à prendre pour renforcer sa résilience financière.

### **4.1.6 Approches**

Il existe deux approches principales pour effectuer ces tests de stress : l'approche historique et l'approche hypothétique.



### **Approche historique :**

Cette approche consiste à utiliser des données historiques pour simuler des scénarios de stress. Les données historiques sont utilisées pour identifier les périodes de stress passées, telles que les crises financières, et pour modéliser la manière dont l'institution financière aurait réagi dans ces conditions. Cette approche est basée sur l'hypothèse que les conditions économiques futures seront similaires aux périodes de stress passées. Les scénarios sont généralement axés sur des événements spécifiques tels qu'une récession économique, une chute des prix de l'immobilier ou des taux d'intérêt élevés.

### **Approche hypothétique :**

Cette approche consiste à créer des scénarios de stress hypothétiques qui ne se sont jamais produits dans le passé. Cette approche est basée sur l'hypothèse que les conditions économiques futures peuvent être différentes de celles du passé. Les scénarios peuvent inclure des événements tels qu'une récession économique, une crise financière, une baisse importante des prix des matières premières, une augmentation des taux d'intérêt, etc. Les scénarios sont construits à partir d'informations économiques actuelles ou passées, ainsi que de prévisions économiques.

#### **4.1.7 Stress test de la trésorerie**

Le stress test de la trésorerie est un type de stress test qui évalue la capacité d'une entreprise à faire face à une crise de liquidité. Il s'agit d'un test qui mesure la capacité d'une entreprise à répondre à ses besoins en liquidités pendant une période de stress financier prolongée, souvent d'une durée d'un mois à un an.

Il implique la création de scénarios hypothétiques pour évaluer les risques de liquidité auxquels l'entreprise pourrait être confrontée. Ces scénarios peuvent inclure des événements tels qu'une panique sur les marchés financiers, une augmentation soudaine de la demande de liquidités ou une réduction significative de l'accès aux sources de financement.

Le stress test de la trésorerie peut également inclure des mesures telles que la vérification des niveaux de liquidités, la modélisation des flux de trésorerie, l'évaluation de la capacité de l'entreprise à se refinancer sur les marchés financiers et l'analyse des sources de financement disponibles. Les résultats du stress test peuvent ensuite être utilisés pour identifier les domaines de vulnérabilité et pour élaborer des plans d'urgence en cas de crise de liquidité.

Les régulateurs financiers exigent souvent que les banques et les autres institutions financières effectuent des stress tests de trésorerie pour s'assurer qu'elles disposent de suffisamment de liquidités pour faire face à des périodes de stress financier. Les résultats de ces tests sont souvent communiqués aux régulateurs et au public afin de renforcer la transparence et la confiance dans le secteur financier.

## 4.2 Machine Learning et Séries Temporelles

### 4.2.1 Intelligence Artificielle en Finance

#### Introduction à l'IA

L'IA se réfère à la capacité d'un système informatique à exécuter des tâches qui nécessitent normalement l'intelligence humaine. L'IA vise à simuler et à reproduire des comportements intelligents, tels que la perception, la compréhension du langage, l'apprentissage, le raisonnement, la prise de décision et l'adaptation à de nouvelles situations.

L'IA repose sur le développement et l'application de techniques et d'algorithmes avancés qui permettent aux machines d'apprendre à partir des données et d'effectuer des tâches de manière autonome ou semi-autonome. Les principaux domaines de l'IA comprennent l'apprentissage automatique (machine learning), le traitement du langage naturel (natural language processing), la vision par ordinateur (computer vision), les réseaux neuronaux (neural networks) et la logique symbolique.

**L'apprentissage automatique (Machine Learning)** est l'un des aspects clés de l'IA, où les machines sont entraînées à reconnaître des schémas et à effectuer des prédictions en utilisant des données d'entraînement. Les algorithmes d'apprentissage automatique permettent aux machines d'apprendre à partir de l'expérience et d'ajuster leur comportement en fonction des informations reçues.

**Le traitement du langage naturel** concerne la capacité des machines à comprendre, interpréter et générer le langage humain de manière naturelle. Cela inclut la reconnaissance vocale, la traduction automatique, la génération de texte et la compréhension des intentions et des émotions dans le langage.

**La vision par ordinateur** concerne la capacité des machines à comprendre et à interpréter les images et les vidéos. Les algorithmes de vision par ordinateur permettent aux machines de reconnaître des objets, des visages, des mouvements et de comprendre le contenu visuel.

**Les réseaux neuronaux** sont des structures informatiques inspirées du fonctionnement du cerveau humain. Ils sont utilisés pour modéliser et simuler des processus cognitifs et des comportements intelligents. Les réseaux neuronaux profonds, également appelés réseaux de neurones profonds, sont des architectures de réseaux neuronaux complexes qui permettent d'obtenir des performances élevées dans des tâches telles que la classification d'images et la reconnaissance vocale.

**La logique symbolique** concerne la représentation et le raisonnement sur des connaissances et des informations sous forme de symboles et de règles formelles. Elle est utilisée dans des domaines tels que l'expertise, le raisonnement déductif et la planification.

En combinant ces différents aspects, l'intelligence artificielle vise à créer des systèmes informatiques capables de résoudre des problèmes complexes, d'interagir avec les humains de manière naturelle et de s'adapter à des situations nouvelles et changeantes.

## Appliquée à la finance

L'intégration de l'intelligence artificielle (IA) dans les stress tests financiers offre un potentiel considérable pour améliorer la modélisation des scénarios de stress et évaluer l'impact sur les institutions financières. L'IA permet de combiner des techniques avancées de machine learning avec l'analyse des séries temporelles, ouvrant ainsi de nouvelles perspectives pour capturer des modèles de comportement complexes et anticiper la réaction des institutions face à des conditions économiques et financières défavorables.

Les techniques de machine learning appliquées aux séries temporelles permettent de détecter des modèles, des saisons et des comportements non linéaires, ce qui est essentiel pour prédire les variations des indicateurs financiers dans le cadre des stress tests. Grâce à l'apprentissage automatique, les modèles peuvent apprendre à partir des données historiques, en prenant en compte les tendances passées, les cycles saisonniers, les effets de long terme et les relations complexes entre les différentes variables. L'intelligence artificielle, en combinant le machine learning avec l'analyse des séries temporelles, offre des avantages significatifs dans les stress tests financiers. Les modèles de machine learning peuvent être entraînés sur des ensembles de données historiques de séries temporelles afin de prédire les scénarios de stress et d'évaluer les risques associés. Grâce à cette approche, il est possible d'obtenir une meilleure estimation des pertes potentielles, une évaluation plus précise de la résilience financière et une meilleure prise de décision pour les institutions soumises aux stress tests. En utilisant l'IA pour les stress tests, les institutions financières peuvent bénéficier d'une meilleure compréhension des vulnérabilités potentielles, des impacts systémiques et des risques spécifiques qui pourraient résulter de scénarios de stress. Les modèles de machine learning peuvent intégrer des données provenant de différentes sources, telles que les données macroéconomiques, les données financières, les données de marché et les données opérationnelles, afin de prendre en compte de multiples facteurs qui influencent la résilience financière. Cela permet aux institutions financières d'obtenir des informations plus précises et détaillées sur leur profil de risque, d'identifier les domaines qui nécessitent une attention particulière et de développer des stratégies d'atténuation des risques plus efficaces.

En résumé, l'utilisation du machine learning et de l'analyse des séries temporelles dans le cadre des stress tests financiers offre des avantages significatifs pour améliorer la modélisation, la prévision et l'évaluation des risques. Ces approches permettent une modélisation plus précise des scénarios de stress, une meilleure prévision des tendances financières et une évaluation plus fine des risques spécifiques auxquels les institutions financières sont exposées. Grâce à l'intégration de l'IA, les stress tests peuvent fournir des informations plus robustes et précieuses pour soutenir la prise de décision et renforcer la résilience du secteur financier dans un environnement économique complexe et en constante évolution.

### 4.2.2 Introduction au Machine Learning

#### Définition et Principes Fondamentaux

Le Machine Learning, une branche de l'intelligence artificielle, est une méthodologie pour créer des systèmes informatiques qui apprennent à partir des données pour améliorer leurs performances ou effectuer certaines tâches. Au lieu d'écrire explicitement des instructions pour exécuter une tâche, les machines sont formées à l'aide de grands ensembles de données, tirant des conclusions ou faisant des prédictions à partir de ces données.

La notion de “Machine Learning” est ancrée dans le concept de modèles - des représentations abstraites ou simplifiées basées sur les observations des données. Ces modèles sont construits avec une phase d'apprentissage, où un algorithme ajuste ses paramètres pour minimiser l'erreur de ses prédictions. Une fois formés, ces modèles peuvent être utilisés pour faire des inférences sur des données inconnues.

Trois principaux types d'apprentissage automatique sont généralement reconnus :

1. **Apprentissage supervisé** : Dans ce type d'apprentissage, nous fournissons à l'algorithme des paires d'exemples d'entrée-sortie et l'entraînons à associer les entrées aux sorties correctes. Les tâches typiques sont la régression (prédire une sortie continue à partir d'une entrée) et la classification (prédire une étiquette de classe à partir d'une entrée).
2. **Apprentissage non supervisé** : Dans ce type d'apprentissage, nous fournissons à l'algorithme des exemples d'entrée sans aucune étiquette de sortie. L'objectif ici est de découvrir la structure et les relations intrinsèques au sein des données d'entrée. Les tâches courantes sont le clustering (regrouper des entrées similaires) et la réduction de la dimensionnalité (simplifier les entrées sans perdre leur structure essentielle).
3. **Apprentissage par renforcement** : Dans ce type d'apprentissage, un agent interagit avec un environnement, reçoit des récompenses ou des pénalités pour ses actions et apprend à choisir les actions qui maximisent la récompense cumulée. C'est un cadre de choix pour des problèmes comme le jeu, la navigation et le contrôle robotique.

Certains principes fondamentaux guident la conception et l'application des méthodes de Machine Learning :

- **La capacité et le surapprentissage** : Il y a un équilibre délicat entre la capacité d'un modèle à s'adapter aux données d'entraînement et à généraliser aux données de test. Un modèle trop complexe peut s'adapter aux données d'entraînement mais échouer à généraliser (surapprentissage), tandis qu'un modèle trop simple peut échouer à capturer des tendances significatives dans les données d'entraînement (sous-apprentissage).
- **Le compromis biais-variance** : Le biais se réfère à l'erreur systématique introduite par le choix d'un modèle spécifique, tandis que la variance se réfère à l'erreur due à la sensibilité du modèle à la fluctuation dans les données d'entraînement. Un équilibre approprié entre le biais et la variance est essentiel pour obtenir un modèle performant.
- **La validation et l'évaluation** : Les modèles de Machine Learning doivent être validés sur des ensembles de données indépendants pour évaluer leur performance et éviter le surapprentissage. De plus, différentes métriques d'évaluation peuvent être utilisées selon le problème à résoudre.
- **Les implications éthiques** : Le Machine Learning, tout en étant extrêmement puissant, a également des implications éthiques profondes. Les questions de biais, d'équité, de transparence, de respect de la vie privée et d'explicabilité sont des préoccupations majeures qui doivent être prises en compte lors de la conception et de l'application des méthodes de Machine Learning.
- **L'importance des données** : Les modèles de Machine Learning sont aussi bons que les données sur lesquelles ils sont formés. Les données de haute qualité, représentatives et bien préparées sont essentielles pour former des modèles efficaces.

De plus, le soin avec lequel les données sont recueillies et préparées peut avoir un impact significatif sur les performances du modèle. Cela comprend des aspects tels que le traitement des valeurs manquantes, la détection et la gestion des valeurs aberrantes, et la transformation appropriée des variables.

- **Interprétabilité et explicabilité** : L'interprétabilité fait référence à la mesure dans laquelle un humain peut comprendre le processus de prise de décision d'un modèle, tandis que l'explicabilité fait référence à la mesure dans laquelle les prédictions d'un modèle peuvent être expliquées en termes humains. Avec l'essor des modèles de Machine Learning de plus en plus complexes, tels que les réseaux de neurones profonds, ces deux aspects sont devenus des domaines de recherche actifs.
- **Les défis computationnels** : Le Machine Learning implique généralement de travailler avec de grands ensembles de données et des modèles complexes, ce qui peut être très gourmand en ressources computationnelles. De nouvelles méthodes et technologies sont constamment développées pour rendre le calcul plus efficace, depuis l'amélioration des algorithmes jusqu'à l'utilisation de hardware spécialisé.

En résumé, le Machine Learning est une discipline riche et complexe qui exige une compréhension profonde des principes fondamentaux et des considérations pratiques. Il a une vaste gamme d'applications et a le potentiel de transformer de nombreux aspects de notre société. Cependant, son utilisation doit être guidée par une compréhension éclairée de ses forces et de ses limites, ainsi que par une considération éthique et judicieuse.

## Applications du Machine Learning

Le Machine Learning trouve des applications dans une multitude de domaines, transformant et améliorant notre façon d'interagir avec le monde. Voici quelques-unes des applications les plus importantes :

1. **Reconnaissance d'image et de vidéo** : Les algorithmes d'apprentissage automatique sont utilisés pour la reconnaissance d'objets, la détection de visages, l'analyse de scènes et bien plus encore. Ces applications sont utilisées dans la sécurité, la surveillance, la réalité augmentée, l'automatisation industrielle et la vision par ordinateur en général.
2. **Traitement du langage naturel (NLP)** : Le Machine Learning permet aux machines de comprendre, analyser et générer du langage humain. Les applications incluent la traduction automatique, la reconnaissance vocale, la génération de texte, l'analyse de sentiment, la classification de documents, les chatbots, etc.
3. **Systèmes de recommandation** : Les systèmes de recommandation utilisent le Machine Learning pour fournir des recommandations personnalisées aux utilisateurs. Ils sont utilisés dans le streaming musical, la vidéo à la demande, le commerce électronique, etc.
4. **Médical et Santé** : Le Machine Learning est utilisé dans la détection précoce des maladies, la prédiction de l'évolution des maladies, la personnalisation des traitements, l'analyse de données médicales massives, la radiologie diagnostique, la recherche génomique, etc.
5. **Autonomie des véhicules** : Les véhicules autonomes utilisent le Machine Learning pour la navigation, l'évitement d'obstacles, la reconnaissance de la signalisation routière, etc.

6. **Finance** : Le Machine Learning est utilisé pour l'analyse des données financières, la détection de fraude, la prédiction des prix des actions, l'optimisation de portefeuille, etc.
7. **Cybersécurité** : Le Machine Learning aide à détecter les activités malveillantes, les attaques de logiciels malveillants et à renforcer la sécurité des systèmes informatiques.
8. **Robotique** : Le Machine Learning permet aux robots d'apprendre à interagir avec leur environnement, à accomplir des tâches complexes, à reconnaître des objets, etc.

Ces applications ne représentent qu'une partie des nombreuses utilisations du Machine Learning. Son potentiel est vaste et continue de croître, offrant des opportunités d'innovation et d'amélioration dans divers domaines de notre société.

### 4.2.3 Séries Temporelles

Qui dit Stress test dit forcément, en premier lieu, une prévision de séries temporelles dans le cas Business as Usual. C'est la raison pour laquelle il serait important d'en parler. Une série temporelle est une séquence de données numériques qui sont collectées à des intervalles réguliers dans le temps. Elle peut représenter une mesure de tout phénomène qui évolue dans le temps, comme les ventes mensuelles d'un produit, le nombre de visiteurs d'un site web, la température moyenne d'une ville chaque jour, les flux entrants et sortants de la trésorerie, etc.

Une série temporelle peut être représentée sous forme de tableau, de graphique ou de diagramme linéaire, où l'axe horizontal représente le temps et l'axe vertical représente la valeur de la mesure.

L'analyse des séries temporelles est importante dans de nombreux domaines, notamment la finance, l'économie, la météorologie, l'écologie, la biologie et la physique, car elle permet de détecter les tendances, les cycles, les saisons, les effets de court et long terme, les événements extrêmes, etc. Elle peut également être utilisée pour prévoir ou simuler des événements futurs, en utilisant des modèles mathématiques et statistiques.

#### Les composantes d'une série chronologique :

Le but de la décomposition d'une série chronologique est de distinguer dans l'évolution de la série, une tendance « générale », des variations saisonnières qui se répètent périodiquement, et des variations accidentelles imprévisibles. Une brève description de ces composantes est donnée ici :

1. **Saisonnalité** : Cette composante représente des fluctuations périodiques à l'intérieur d'une certaine période (par exemple, chaque mois / trimestre / année). La saisonnalité est toujours d'une période fixe et connue.
2. **Tendance** : La tendance générale d'une série chronologique à augmenter, diminuer ou stagner sur une longue période est appelée tendance séculaire ou simplement tendance. Ainsi, on peut dire que la tendance est un mouvement à long terme dans une série chronologique. Par exemple, les séries relatives à la croissance démographique, au nombre de maisons dans une ville, etc. montrent une tendance à la hausse, alors qu'une tendance à la baisse peut être observée dans les séries relatives aux taux de mortalité, aux épidémies, etc.

3. **Cyclique** : La variation cyclique dans une série chronologique décrit les changements à moyen terme de la série, causés par des circonstances, qui se répètent par cycles. La durée d'un cycle s'étend sur une période plus longue de temps, généralement deux ans ou plus.
4. **Irrégulière** : Les variations irrégulières ou aléatoires d'une série chronologique sont causées par des influences imprévisibles, qui ne sont pas régulières et qui ne se répètent pas non plus dans un modèle particulier. Ces variations sont causées par incidences telles que guerre, grève, tremblement de terre, inondation, révolution, etc. Il n'existe pas de technique statistique définie pour mesurer les fluctuations aléatoires d'une série chronologique.

### La transformation de série chronologique :

Les séries chronologiques nécessitent souvent une transformation avant la prévision ou l'analyse. La transformation des séries chronologiques est souvent utile car de nombreuses méthodes de prévision conviennent mieux aux séries chronologiques conformes à certaines hypothèses. Elle peut également simplifier le problème de prévision en supprimant les composantes de tendance et saisonnières. Si une opération de transformation est appliquée aux données de série temporelle, une opération inverse similaire doit être appliquée à la prédiction. Identifier la meilleure méthode de transformation peut s'avérer très difficile pour un problème de prédiction donné, car il n'existe pas de règles standard concernant le prétraitement des données chronologiques. Dans ce qui suit, nous présentons les opérations de transformation les plus courantes pour l'analyse et la prévision de séries chronologiques.

1. **Transformation log** : La transformation log est un type de transformation de puissance dans lequel une fonction logarithmique est appliquée aux valeurs de série temporelle. Cette transformation est souvent utilisée dans les prévisions et analyses de séries temporelles pour stabiliser la variance d'une série. En d'autres termes, les prévisions basées sur  $X_t = \log Y_t$  peuvent être bien meilleures que les prédictions directes de  $Y_t$  si  $X_t$  a une variance plus stable que  $Y_t$ . Il est également utile pour prévoir des séries chronologiques positives à croissance exponentielle.
2. **La différenciation** : La transformation Différence est une transformation de série chronologique très célèbre qui vise à éliminer les tendances et la saisonnalité. Alors que les transformations de puissance visent à stabiliser la variance d'une série chronologique, la différenciation vise à stabiliser la moyenne d'une série chronologique. La série différenciée est le changement entre chaque observation de la série originale selon l'équation suivantes :

$$Y'_t = Y_t - Y_{t-1} \dots (1)$$

L'équation (1) est appelée différenciation au premier ordre et, si elle est appliquée à une série, elle supprimera la tendance linéaire.

Si la série différenciée obtenue n'est pas stationnaire, une différenciation de second ordre peut être nécessaire et est donnée par :

$$Y''_t = Y'_t + Y''_{t-1}$$

Avec  $Y'_t$  la différenciation au premier ordre et  $Y''_t$  la différenciation au deuxième ordre. Afin de supprimer la structure saisonnière d'une série chronologique, une différence saisonnière est souvent nécessaire et est obtenue en calculant la différence entre une observation et l'observation précédente de la même saison.

$$Y'_t = Y_t - Y_{t-m}$$

Avec  $m$  le nombre de saisons.

## Techniques de Prévision des Séries Temporelles

Les séries temporelles sont une suite d'observations recueillies à des intervalles de temps réguliers. La prévision des séries temporelles fait référence à l'utilisation de modèles statistiques pour prédire les points de données futurs basés sur les données passées. Plusieurs techniques sont couramment utilisées :

1. **Méthode de lissage exponentiel simple** : Cette technique utilise un facteur de lissage pour pondérer l'impact des observations passées sur les prévisions futures. Elle est utile lorsque les données sont relativement stables sans tendance ou saisonnalité discernable. Le principal inconvénient de cette méthode est qu'elle peut sous-estimer les prévisions lorsque les données présentent une tendance.
2. **Méthode de lissage exponentiel double (ou méthode de Holt)** : Cette méthode étend la méthode de lissage exponentiel simple pour inclure une composante de tendance. Elle est plus adaptée lorsque les données présentent une tendance, qu'elle soit croissante ou décroissante.
3. **Méthode de lissage exponentiel triple (ou méthode de Holt-Winters)** : Cette méthode ajoute une composante de saisonnalité à la méthode de Holt. Elle est utilisée lorsque les données montrent à la fois une tendance et une saisonnalité.
4. **Modèles autorégressifs (AR)** : Ces modèles prédisent les valeurs futures en fonction des valeurs passées. Ils utilisent la dépendance entre une observation et un certain nombre d'observations passées (observations retardées). Ces modèles sont utiles lorsque les données montrent des corrélations à court terme.
5. **Modèles de moyenne mobile (MA)** : Ces modèles prédisent les valeurs futures en se basant sur les erreurs des prévisions passées. Ils utilisent la dépendance entre une observation et une erreur résiduelle à partir d'une moyenne mobile appliquée à des observations passées.
6. **Modèles ARIMA (AutoRegressive Integrated Moving Average)** : Ces modèles combinent les modèles AR et MA et tentent également de supprimer les tendances dans les données (c'est-à-dire qu'ils sont 'intégrés'). Les modèles ARIMA sont utiles lorsque les données présentent une tendance et/ou une saisonnalité.
7. **Modèles de régression** : Ils sont utilisés lorsque la variable que nous voulons prédire est corrélée à une ou plusieurs variables indépendantes. Les modèles de régression peuvent être linéaires ou non linéaires.
8. **Modèles de réseaux de neurones** : Ces modèles sont basés sur des systèmes d'apprentissage qui tentent de simuler la manière dont le cerveau humain apprend. Ils sont particulièrement utiles lorsque les relations entre les variables sont non-linéaires ou lorsque les données contiennent de nombreuses variables d'entrée.

Chacune de ces techniques a ses propres avantages, limites et hypothèses. L'efficacité de ces méthodes peut varier en fonction des caractéristiques spécifiques de la série temporelle.



## 4.3 Etude bibliographique

### 4.3.1 Tableau comparatif

Voici un tableau comparatif des différents articles étudiés, illustrant l'application des techniques de Machine Learning pour le stress test financier, ainsi que pour la prévision financière, qui est un élément intégral de ce processus :

Titre	Année de publication	Auteurs	Mots-clés	Problématique
Deep Learning for financial banking stress-test analytics	2020	Farid Razzak	Stress-test, Tier 1 ratio, Generative Adversarial Networks, Variational Auto-encoder, Conditional Generative Adversarial Networks, Bank Capital, Transfer Learning, Attention mechanism	Prévision des capitaux et des pertes des banques
"Can banks survive the next financial Crisis? An adversarial deep learning model for bank stress testing"	2018	Nikhil Malik, Param Vir Singh, Urooj Khan	Generative Deep Learning, Generative Adversarial Networks, DataScope AutoEncoders, LSTM, FinTech, Algorithms in Business, Stress Testing, Financial Regulations, Plausible and Adverse	L'utilisation du Deep Learning lors des crises financières.

Causal Data Science for Financial Stress Testing	2018	Gelin Gao Bud Mishra Daniele Ramazzotti	Stress Testing, Graphical Models, Causality, Suppes-Bayes Causal Networks, Classification, Decision Trees	l'utilisation des méthodes et techniques de la Data Science causales pour évaluer et quantifier les risques financiers
Historical Analysis and Time Series Forecasting of Stock Market using FB Prophet	2022	Ciram Saikishna, Nalam Sai Venkat Sumanth, Muppavarapu Madhu Sudhan Rao, Thangakumar J	FB Prophet, Time series forecasting, Streamlit, Yfinance, Plotly, Matplotlib, Null Values Handling	La prédiction du comportement du marché boursier
A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India	2022	Shobhit Chaturvedi, Elangovan Rajasekar, Sukumar Natarajan, Nick McCullen	LSTM, RNN, SARIMA, FB Prophet, Central Energy Authority	Prévision précise de la demande énergétique.

Titre	Data	Résultats / Métrique	Algorithmes	Synthèse
-------	------	----------------------	-------------	----------

Deep Learning for financial banking stress-test analytics	Pour la génération de scénario : Wharton Research Data Services (WRDS, 2019), Global Financial Data (GFD), the U.S. Federal Reserve Board of Governors (Reserve, 2015) Prévisionnel : Comptes annuels (financial statements)	Générateur de scénario : <i>RMSE</i> » <i>MCVAE</i> » <i>VAE</i> » <i>CGAN</i> » <i>GAN</i> Prédicteur de performance : <i>RMSE</i> » <i>DA-RNN</i> » <i>LSTM</i> » <i>GRU</i>	<i>GRU</i> , <i>LSTM</i> , <i>DA-RNN</i>	L'utilisation des <i>MCVAE</i> pour générer des scénarios plus plausibles Le <i>DA-RNN</i> sont utilisés pour prédire les performances dans le scénario généré par les <i>MCVAE</i>
"Can banks survive the next financial Crisis? An adversarial deep learning model for bank stress testing"	16 macro economic variables such as GDP, Unemployment Rate, CPI reported by Federal Reserve Microeconomic variables Bank Data (caractéristiques et performances passées)	Génération de scénario : LogLikelihood <i>CGAN</i> » <i>MVN</i> Prédiction de performance : <i>RMSE</i> » <i>LSTM</i> » <i>VAR</i> <i>CGAN</i> + <i>LSTM</i> est plus plausible et négatif	<i>CGAN</i> , <i>MVN</i> (Multi Variate Normal) <i>LSTM</i> , <i>VAR</i> , <i>FE-QAR</i>	L'utilisation des <i>CGAN</i> pour générer des scénarios plus plausibles Les <i>LSTM</i> sont utilisés pour prédire les performances dans le scénario généré par les <i>CGAN</i>
Causal Data Science for Financial Stress Testing	data développée par des relations causales composée de 10 Portfolio	Critère d'information d'Akaike ( <i>AIC</i> ), Critère d'Information de Bayes ( <i>BIC</i> ) appliqués à différents échantillons	<i>SBCN</i> , Stand Bayesian Network	Il est important de comprendre les relations de causalité entre les différentes variables financières et les facteurs de stress.

<p>Historical Analysis and Time Series Forecasting of Stock Market using FB Prophet</p>	<p>Les données utilisées pour alimenter les algorithmes proviennent de “Yfinance”, qui fournit des données de séries temporelles rafraîchies quotidiennement</p>	<p>/</p>	<p>FB-prophet, LSTM, Regression techniques</p>	<p>l'article cherche à déterminer quel algorithme de prévision fournit la meilleure précision, le taux d'erreur le plus faible, gère efficacement les données désordonnées, et n'est pas dérangé par les valeurs nulles. L'article conclut que l'algorithme Prophet de Facebook répond le mieux à ces critères.</p>
---	--	----------	--	---

<p>A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India</p>	<p>108 mois de données historiques de 2008 à 2017 pour prédire 24 mois de données de 2017 à 2019</p>	<p>Pour la prédiction de la demande totale mensuelle : <i>RMSE</i> et <i>MAPE</i> Facebook Prophet &gt; CEA &gt; SARIMA &gt; LSTM RNN  Pour la prédiction de la demande de pointe mensuelle : <i>RMSE</i> et <i>MAPE</i> Facebook Prophet &gt; CEA &gt; SARIMA &gt; LSTM RNN</p>	<p>SARIMA, LSTM, RNN, FB-prophet, Central Energy Authority (CEA) Trend-based Model</p>	<p>Le modèle de la CEA prévoyait bien la demande énergétique totale mensuelle, mais sous-estimait significativement la demande de pointe mensuelle. Le modèle Prophet de Facebook s'est bien comporté pour les deux types de prévisions. Les modèles SARIMA et LSTM RNN ont montré des erreurs de prévision plus importantes.. Par conséquent, l'étude recommande d'utiliser le modèle Prophet de Facebook pour les prévisions énergétiques futures en Inde.</p>
--	--	--	--	--

Les articles étudiés illustrent l'application croissante et innovante du Machine Learning dans le domaine des stress tests financiers et de la prévision financière. Plusieurs méthodes, y compris la régression, les réseaux de neurones et les arbres de décision, ont été appliquées pour analyser une variété de données financières, y compris les données de marché et les informations macroéconomiques. Les résultats de ces études indiquent que le Machine Learning peut améliorer la précision des prévisions financières et des stress tests, contribuant ainsi à une meilleure gestion des risques. Cependant, les auteurs soulignent également les défis, tels que le risque d'overfitting et la nécessité d'interprétabilité des modèles.

Dans la sous-section suivante, nous allons non seulement énumérer, mais également four-

nir des définitions concises des différents algorithmes de Machine Learning identifiés au cours de notre étude bibliographique exhaustive.

### 4.3.2 Algorithmes identifiés :

#### ARIMA

L'ARIMA est une méthode de modélisation de séries temporelles qui a été introduite par George Box et Gwilym Jenkins en 1976. Elle est conçue pour gérer les séries temporelles non stationnaires en utilisant des différences pour rendre la série stationnaire avant de l'appliquer à un modèle ARMA (AutoRegressive Moving Average). Les modèles ARIMA sont souvent utilisés dans le domaine de la prévision économique, les prévisions météorologiques, la modélisation de la bourse et bien d'autres domaines où l'analyse de séries temporelles est nécessaire.

Un modèle ARIMA est défini par trois paramètres :  $p$ ,  $d$  et  $q$ .

- $p$  est l'ordre du composant autoregressif (AR)
- $d$  est l'ordre de la différence nécessaire pour rendre la série temporelle stationnaire (I)
- $q$  est l'ordre du composant de moyenne mobile (MA)

En combinant ces trois composants, le modèle ARIMA est capable de modéliser une variété de séries temporelles complexes.

Voici les équations principales du modèle ARIMA :

Equation du composant AutoRegressive (AR) :

$$AR(p) : Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (4.1)$$

Equation du composant Integrated (I) :

$$I(d) : \nabla^d Y_t \quad (4.2)$$

Equation du composant Moving Average (MA) :

$$MA(q) : Y_t = C + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (4.3)$$

Dans ces équations,  $Y_t$  représente la série temporelle à l'instant  $t$ ,  $C$  est une constante,  $\epsilon_t$  est l'erreur à l'instant  $t$ , et  $\nabla^d$  est l'opérateur de différenciation. Les coefficients  $\phi$  et  $\theta$  sont appris pendant l'entraînement.

Le composant autoregressif (AR) capture la dépendance entre une observation et un certain nombre de ses observations passées. Le composant intégré (I) rend la série temporelle stationnaire en supprimant la tendance et la saisonnalité. Le composant de moyenne mobile (MA) capture la dépendance entre l'observation et l'erreur résiduelle de l'observation passée.

En combinant ces trois composants, le modèle ARIMA peut modéliser efficacement les séries temporelles non stationnaires en prenant en compte à la fois les tendances et la saisonnalité.

## RNN

Un réseau de neurones récurrents (RNN) est un type d'algorithme d'apprentissage profond utilisé pour le traitement de séquences de données, tels que des séquences de texte, des signaux sonores et des séries temporelles. Contrairement aux réseaux de neurones classiques, les RNN peuvent prendre en compte l'historique des données en utilisant une boucle de rétroaction qui leur permet de stocker des informations et de les utiliser pour traiter les données futures.

La structure de base d'un RNN consiste en une couche cachée de neurones qui est connectée à elle-même. À chaque étape de temps, le RNN prend en entrée les données d'entrée et l'état caché de l'étape de temps précédente. Il utilise ensuite ces informations pour calculer la sortie et l'état caché actuel, qui sera transmis à la prochaine étape de temps.

La boucle de rétroaction dans un RNN lui permet de conserver une mémoire à court terme des données d'entrée précédentes, ce qui lui permet de prendre en compte le contexte et de traiter les données de manière séquentielle. Cette mémoire à court terme est appelée "état caché", qui est mis à jour à chaque étape de temps à partir des données d'entrée et de l'état caché précédent.

Les RNN ont été utilisés dans de nombreuses applications, notamment la traduction automatique, la reconnaissance de la parole, la prédiction de séries temporelles et la génération de texte. Cependant, les RNN ont tendance à rencontrer des problèmes de vanishing gradients, où la rétropropagation de l'erreur à travers de nombreuses étapes de temps peut entraîner une diminution rapide du gradient et une lenteur de l'apprentissage.

Pour surmonter ce problème, des variantes des RNN ont été proposées, notamment les LSTM (Long Short-Term Memory) et les GRU (Gated Recurrent Units), qui utilisent des mécanismes de portes pour réguler le flux d'informations et éviter la disparition des gradients.

Les équations principales du modèle RNN sont les suivantes :

Equation de la cellule RNN :

$$h_t = f(x_t, h_{t-1}) \quad (4.4)$$

Equation de la sortie de la cellule RNN :

$$y_t = g(h_t) \quad (4.5)$$

Equation de la perte de l'ensemble des cellules RNN :

$$J = -\frac{1}{N} \sum_{t=1}^N \log(P(y_t|x_1, \dots, x_N)) \quad (4.6)$$

Dans ces équations,  $x_t$  représente le vecteur d'entrée à l'instant  $t$ ,  $h_t$  est l'état caché du réseau à l'instant  $t$ , et  $y_t$  est la sortie du réseau à l'instant  $t$ . Les fonctions  $f$  et  $g$  sont des fonctions non-linéaires qui sont appliquées aux entrées et à l'état caché pour calculer l'état caché et la sortie à chaque instant. La fonction de perte  $J$  mesure l'écart entre la sortie du réseau et la sortie attendue, et est utilisée pour entraîner le réseau à ajuster ses paramètres pour minimiser cette perte.

## LSTM

Les LSTM (Long Short-Term Memory) sont une variante de réseau de neurones récurrents (RNN) qui ont été développés pour résoudre les problèmes de vanishing gradients, qui se produisent lorsque les gradients diminuent exponentiellement à mesure que l'algorithme de rétropropagation se propage dans le temps. Les LSTM utilisent des mécanismes de portes pour réguler le flux d'informations et contrôler la quantité d'information stockée dans l'état caché.

La structure de base d'un LSTM consiste en une cellule LSTM qui est connectée à une boucle de rétroaction. La cellule LSTM est composée de trois portes : la porte d'entrée, la porte de sortie et la porte d'oubli. Chacune de ces portes contrôle le flux d'informations entrantes et sortantes de la cellule LSTM.

La porte d'entrée contrôle la quantité d'informations nouvelles à ajouter à l'état caché, en fonction des données d'entrée et de l'état caché précédent. La porte d'oubli détermine la quantité d'informations précédentes à oublier, en fonction des données d'entrée et de l'état caché précédent. La porte de sortie détermine la quantité d'informations à sortir de la cellule LSTM, en fonction des données d'entrée et de l'état caché actuel.

L'état caché dans un LSTM est composé de deux parties : l'état caché à court terme et l'état caché à long terme. L'état caché à court terme stocke les informations actuelles de la séquence, tandis que l'état caché à long terme stocke les informations pertinentes de la séquence jusqu'à présent. Les portes de la cellule LSTM contrôlent la quantité d'informations stockées dans l'état caché à long terme et l'état caché à court terme.

Les LSTM ont été utilisés dans de nombreuses applications, notamment la reconnaissance de la parole, la traduction automatique, la génération de texte et la prédiction de séries temporelles. Les LSTM ont également été utilisés pour améliorer les performances des modèles de traitement du langage naturel, tels que les réseaux de neurones convolutifs (CNN) et les Transformers.

Voici les équations principales du modèle LSTM :

Equation d'entrée (input gate) :

$$i(t) = \sigma(W_i[h(t-1), x(t)] + b_i) \quad (4.7)$$

Equation de sortie (output gate) :

$$o(t) = \sigma(W_o[h(t-1), x(t)] + b_o) \quad (4.8)$$

Equation de mise à jour de la cellule (cell update gate) :

$$f(t) = \sigma(W_f[h(t-1), x(t)] + b_f) \quad (4.9)$$

Equation de sortie de la cellule (cell state output) :

$$C(t) = f(t) * C(t-1) + i(t) * \tanh(W_c[h(t-1), x(t)] + b_c) \quad (4.10)$$

Equation de sortie de l'état caché (hidden state output) :

$$h(t) = o(t) * \tanh(C(t)) \quad (4.11)$$



Dans ces équations,  $x(t)$  est l'entrée à l'étape de temps  $t$ ,  $h(t)$  est l'état caché à l'étape de temps  $t$ ,  $C(t)$  est l'état de cellule à l'étape de temps  $t$ , et les variables  $W$  et  $b$  sont les poids et les biais appris par le modèle pendant l'entraînement. La fonction  $\sigma$  représente la fonction d'activation sigmoïde, qui transforme les valeurs en une plage de 0 à 1, tandis que la fonction  $\tanh$  représente la fonction d'activation tangente hyperbolique, qui transforme les valeurs en une plage de -1 à 1.

## GRU

Les GRU (Gated Recurrent Unit) sont une autre variante de réseau de neurones récurrents (RNN) qui ont été introduits pour surmonter les problèmes de vanishing gradients dans les RNN traditionnels. Les GRU utilisent des portes pour contrôler la quantité d'informations stockées dans l'état caché, tout en permettant un flux continu d'informations dans la boucle de rétroaction.

Contrairement aux LSTM, les GRU ont seulement deux portes : la porte de réinitialisation et la porte de mise à jour. La porte de réinitialisation détermine la quantité d'informations précédentes à oublier, tandis que la porte de mise à jour détermine la quantité d'informations nouvelles à ajouter à l'état caché. La porte de réinitialisation permet à la cellule GRU de s'adapter à des séquences plus longues, tandis que la porte de mise à jour permet de gérer les changements de tendances dans les séquences.

Comme les LSTM, les GRU ont été utilisés dans diverses applications de traitement du langage naturel, telles que la traduction automatique, la reconnaissance de la parole et la génération de texte. Les GRU sont également souvent utilisés en tandem avec d'autres architectures de réseaux de neurones, tels que les réseaux de neurones convolutifs (CNN) et les Transformers pour améliorer les performances des modèles de traitement du langage naturel.

Voici les équations principales du modèle GRU :

Equation de la mise à jour de la porte de réinitialisation :

$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \quad (4.12)$$

Equation de la mise à jour de la porte d'oubli :

$$z_t = \sigma(W_z * [h_{t-1}, x_t]) \quad (4.13)$$

Equation de l'état candidat :

$$h'_t = \tanh(W * [r_t * h_{t-1}, x_t]) \quad (4.14)$$

Equation de l'état caché :

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t \quad (4.15)$$

Dans ces équations,  $x_t$  représente le vecteur d'entrée à l'instant  $t$ ,  $h_t$  est l'état caché du réseau à l'instant  $t$ , et  $h'_t$  est l'état candidat à l'instant  $t$ . Les portes de réinitialisation  $r_t$  et d'oubli  $z_t$  contrôlent l'importance relative de l'état caché précédent et de l'état candidat

pour calculer l'état caché actuel  $h_t$ . Les matrices de poids  $W_r$ ,  $W_z$  et  $W$  sont apprises pendant l'entraînement.

La porte de réinitialisation  $r_t$  permet au modèle de décider quelle partie de l'état caché précédent doit être utilisée pour calculer l'état caché actuel. La porte d'oubli  $z_t$  permet au modèle de décider quelle partie de l'état précédent doit être oubliée. L'état candidat  $h'_t$  est une nouvelle hypothèse pour l'état caché, qui est pondérée par la porte de réinitialisation  $r_t$ . L'état caché final  $h_t$  est calculé comme une combinaison linéaire de l'état précédent et de l'état candidat, pondérée par la porte d'oubli  $z_t$ .

## Generative DL (Deep Learning Génératif)

Generative Deep Learning (GDL) est une branche de l'apprentissage profond qui se concentre sur la génération de nouvelles données à partir d'un modèle appris. Les modèles de GDL sont capables de générer des données réalistes, tels que des images, du son, du texte, et même des vidéos. GDL est souvent utilisé dans des applications telles que la synthèse d'images, la création de musique, la génération de texte, la traduction de langue et la création de vidéos.

Les modèles de GDL sont souvent des réseaux de neurones profonds, tels que les réseaux de neurones convolutifs (CNN), les réseaux de neurones récurrents (RNN) et les auto-encodeurs. L'apprentissage de ces modèles se fait souvent à l'aide de l'apprentissage supervisé, semi-supervisé ou non supervisé. L'un des modèles de GDL les plus couramment utilisés est le Generative Adversarial Network (GAN), qui consiste en deux réseaux de neurones, un générateur et un discriminant. Le générateur prend une distribution de bruit en entrée et produit une sortie qui est considérée comme une image synthétique. Le discriminant prend en entrée l'image synthétique produite par le générateur, ainsi qu'une image réelle, et détermine laquelle des deux est la plus probable d'être réelle. Les deux réseaux de neurones sont entraînés simultanément dans un processus appelé "adversarial training", où le générateur essaie de tromper le discriminant en produisant des images synthétiques réalistes, tandis que le discriminant essaie de distinguer les images réelles des images synthétiques. Une autre méthode de GDL est l'utilisation de modèles de Markov cachés (HMM) pour générer des séquences de données. Les HMM sont des modèles probabilistes qui modélisent les transitions entre les états d'une séquence de données, ce qui en fait une méthode utile pour la modélisation de séquences, telles que la musique et la parole. Enfin, les réseaux de neurones récurrents, tels que les LSTM, sont également couramment utilisés pour la génération de séquences. Les RNN peuvent être entraînés à apprendre les motifs d'une séquence de données, puis à générer de nouvelles séquences à partir de ces motifs.

En résumé, Generative Deep Learning est une méthode d'apprentissage profond qui permet de générer de nouvelles données à partir d'un modèle appris. Les modèles de GDL sont souvent des réseaux de neurones profonds, tels que les GAN, les HMM et les LSTM, et sont utilisés pour la génération de divers types de données, tels que des images, du son, du texte et des vidéos.

## Generative Adversarial Networks (GAN)

Les Generative Adversarial Networks (GAN) sont une classe de systèmes d'apprentissage automatique conçus par Ian Goodfellow et ses collègues en 2014. Deux réseaux de neurones sont mis en concurrence dans un jeu de type "minimax". Dans ce jeu, l'un des

réseaux, appelé le "générateur", produit des données, tandis que l'autre réseau, appelé le "discriminateur", évalue ces données. L'objectif est d'amener le générateur à produire des données de plus en plus réalistes, au point que le discriminateur ne puisse plus distinguer les données réelles des données générées par le générateur.

Voici les équations principales du modèle GAN :

Equation de perte du discriminateur :

$$L_D = -[\log(D(x)) + \log(1 - D(G(z)))] \quad (4.16)$$

Equation de perte du générateur :

$$L_G = -\log(D(G(z))) \quad (4.17)$$

Equation de mise à jour des poids du discriminateur :

$$\theta_D = \theta_D - \alpha * \nabla L_D \quad (4.18)$$

Equation de mise à jour des poids du générateur :

$$\theta_G = \theta_G - \alpha * \nabla L_G \quad (4.19)$$

Dans ces équations,  $x$  est un échantillon de données réel,  $z$  est un vecteur de bruit généré aléatoirement,  $G$  est le générateur,  $D$  est le discriminateur,  $\alpha$  est le taux d'apprentissage,  $\theta_D$  et  $\theta_G$  sont les poids du discriminateur et du générateur respectivement, et  $\nabla$  représente l'opérateur de gradient. La perte du discriminateur  $L_D$  mesure la capacité du discriminateur à distinguer les échantillons réels des échantillons générés. La perte du générateur  $L_G$  mesure la capacité du générateur à tromper le discriminateur en produisant des échantillons générés qui ressemblent à des échantillons réels. Les poids du discriminateur et du générateur sont mis à jour en utilisant la rétropropagation du gradient pour minimiser leur perte respective.

## Transformers

Les Transformers sont une architecture de réseau neuronal conçue pour traiter des données séquentielles, comme du texte ou de l'audio, qui ont des dépendances à long terme. Ils ont été introduits en 2017 par Vaswani et ses collègues chez Google Brain et sont largement utilisés dans des applications de traitement de langage naturel.

La principale caractéristique des Transformers est leur mécanisme d'attention. Au lieu de traiter les données séquentiellement et de tenir compte de l'état caché précédent comme le font les réseaux de neurones récurrents (RNN), les Transformers utilisent un mécanisme d'attention qui permet d'attribuer des poids aux différentes parties de l'entrée en fonction de leur importance relative pour la tâche en cours.

Cette fonctionnalité permet aux Transformers de capturer des dépendances à long terme entre les différentes parties de l'entrée, sans la nécessité d'un état caché. C'est particulièrement utile pour la compréhension du texte, où les dépendances entre les mots peuvent être très éloignées dans la phrase.

Les Transformers sont souvent utilisés pour des tâches de traitement du langage naturel, telles que la traduction automatique, la génération de texte et la classification du

texte. En particulier, le modèle de langage pré-entraîné appelé BERT (Bidirectional Encoder Representations from Transformers) a été largement adopté pour diverses tâches de traitement du langage naturel en raison de ses performances supérieures sur une variété de tâches de NLP.

L'architecture des Transformers repose sur l'empilement de plusieurs couches d'encodage et de décodage, qui permettent d'apprendre des représentations hiérarchiques de l'entrée. L'encodage se fait en convertissant la séquence d'entrée en une représentation de vecteur d'attention, tandis que le décodage utilise cette représentation pour générer une sortie séquentielle.

Voici les équations principales du modèle Transformer :

Equation d'attention :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q * K^T}{\sqrt{d_k}}\right) * V \quad (4.20)$$

Equation de la sortie de la couche d'attention multi-têtes :

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) * W^O \quad (4.21)$$

Equation de la sortie de la couche de feedforward :

$$\text{FFN}(x) = \max(0, x * W_1 + b_1) * W_2 + b_2 \quad (4.22)$$

Equation de sortie de la couche d'encodage :

$$\text{EncoderLayer}(x) = \text{LayerNorm}(x + \text{MultiHead}(\text{SelfAttention}(x), x, x)) + \text{FFN}(x) \quad (4.23)$$

Equation de sortie de la couche de décodage :

$$\begin{aligned} \text{DecoderLayer}(x, \text{enc\_output}) = & \text{LayerNorm}(x + \text{MultiHead}(\text{SelfAttention}(x), x, x) \\ & + \text{MultiHead}(\text{Attention}(\text{enc\_output}, \text{enc\_output}, x), \\ & \text{enc\_output}, \text{enc\_output})) \\ & + \text{FFN}(x) \end{aligned} \quad (4.24)$$

Dans ces équations, Q, K, et V sont les vecteurs de requête, clé et valeur, respectivement, d'une matrice d'entrée x. Les termes  $d_k$  et  $h$  représentent la dimension de la clé et le nombre de têtes d'attention multi-têtes, respectivement. Les termes  $W^O$ ,  $W_1$ ,  $W_2$ ,  $b_1$  et  $b_2$  sont des matrices de poids et des vecteurs de biais appris par le modèle pendant l'entraînement. La fonction softmax calcule les poids d'attention normalisés pour les vecteurs de valeur V, tandis que la fonction max représente la fonction d'activation ReLU, qui est utilisée pour la couche de feedforward. La fonction LayerNorm représente la normalisation de la couche.

## FB Prophet

Prophet est un outil de prévision basé sur un modèle additif où des tendances non linéaires sont ajustées avec la saisonnalité annuelle, hebdomadaire et journalière, ainsi que les effets des vacances. Il a été introduit par Sean J. Taylor et Ben Letham de l'équipe Core Data Science de Facebook en 2017.

La caractéristique principale de Prophet est sa capacité à gérer des données de séries temporelles avec des tendances et des composants saisonniers robustes, ainsi que la présence de valeurs manquantes, d'outliers, ou de changements de tendance. Contrairement à d'autres modèles de séries temporelles, Prophet est conçu pour fonctionner avec des données de séries temporelles qui ont des caractéristiques complexes comme plusieurs saisons, des vacances, et des tendances non-linéaires.

Prophet utilise une approche de modélisation additive, où la tendance, la saisonnalité et les composants des vacances sont modélisés séparément puis additionnés pour produire la prévision finale. Cela permet de capturer une variété de formes de tendance et de saisonnalité, ce qui en fait un outil flexible pour la prévision de séries temporelles.

Prophet est souvent utilisé pour la prévision de séries temporelles dans divers domaines tels que les prévisions économiques, les prévisions météorologiques, la demande en électricité, et les ventes de détail.

Le fonctionnement de Prophet repose sur la décomposition de la série temporelle en ses composants de tendance, saisonnalité et vacances, et l'estimation de ces composants séparément. Une fois que ces composants ont été estimés, ils sont additionnés pour obtenir la prévision finale.

En résumé, Prophet est un outil de prévision conçu pour traiter des données de séries temporelles avec des tendances et des composants saisonniers robustes. Prophet utilise une approche de modélisation additive pour estimer les composants de tendance, saisonnalité et vacances séparément, puis les additionne pour produire la prévision finale.

Voici les équations principales du modèle Prophet :

Équation de la tendance :

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad (4.25)$$

Équation du modèle additif :

$$g(t) = \text{piecewise\_linear}(t) \text{ ou } \text{piecewise\_logistic}(t) \quad (4.26)$$

Équation de la saisonnalité :

$$s(t) = \sum (a_i * \cos(2 * \pi * i * t/P) + b_i * \sin(2 * \pi * i * t/P)) \quad (4.27)$$

Équation des vacances :

$$h(t) = \sum (v(t) * k) \quad (4.28)$$

Équation de l'erreur :

$$e(t) = \text{normal}(0, \sigma) \quad (4.29)$$

Dans ces équations,  $t$  représente le temps,  $y(t)$  est la valeur observée à l'instant  $t$ ,  $g(t)$  est la tendance,  $s(t)$  est le composant saisonnier,  $h(t)$  est l'effet des vacances, et  $e(t)$  est l'erreur. Les paramètres  $a_i$ ,  $b_i$ ,  $v(t)$ ,  $k$ , et  $\sigma$  sont des coefficients à estimer. Le terme  $P$

représente la période de la saisonnalité.

## Troisième partie

### Conception de l'outil de stress-test

# Introduction

Notre solution passera par deux grandes étapes, la première consiste en la prévision des séries temporelles des postes significatifs, ce qui correspondrait à un prévisionnel BAU. Par la suite, il s'agirait de stresser ce dernier afin d'obtenir un prévisionnel stressé.

Voici un schéma synthétique qui représente notre vision de ce projet :

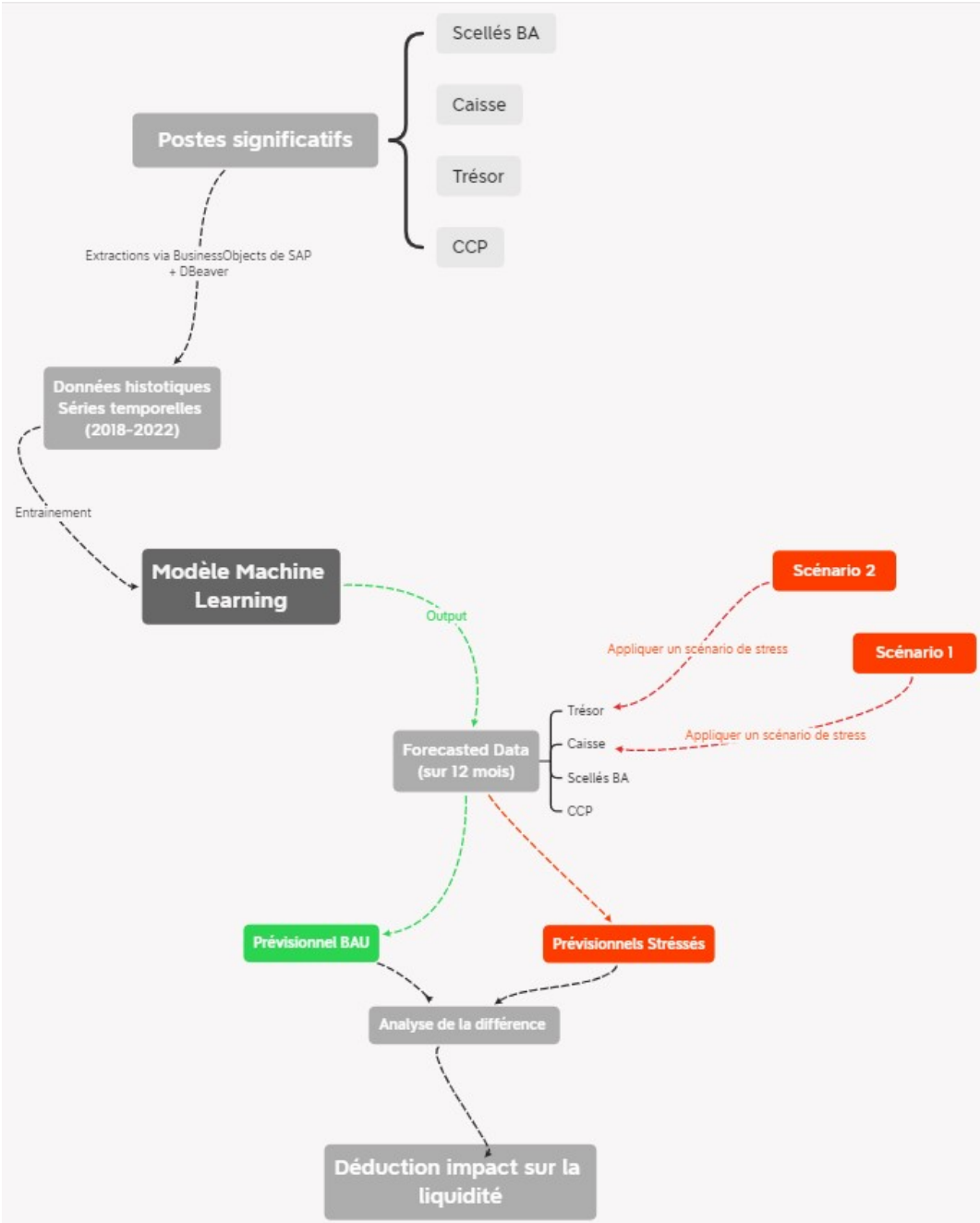


FIGURE 4.1 – Schéma synthétique du projet



# Chapitre 5

## Prévisionnel

### Introduction

Cette partie de notre étude se concentre sur la prédiction et la modélisation des données.

Nous suivrons un processus structuré pour obtenir des prévisions précises et fiables.

Nous commencerons par la collecte, l'analyse et le prétraitement des données. Cette étape portera sur les Mouvements et les Soldes, principalement du poste "Caisse". Les détails concernant les trois (3) autres postes (séries temporelles) se trouvent dans la partie **Annexes**.

Après cela, nous aborderons la phase de modélisation qui se concentrera exclusivement sur les soldes. Cette décision est motivée par la faible volatilité des soldes par rapport aux mouvements, ce qui facilite la prédiction. Nous sélectionnerons les modèles appropriés en fonction des caractéristiques de la série temporelle du poste "Caisse". Nous explorerons différentes architectures de modèles et nous nous concentrerons sur l'optimisation de l'apprentissage automatique pour améliorer les performances des modèles.

Enfin, nous évaluerons les résultats de nos modèles et présenterons les métriques utilisées pour mesurer leur précision et leur efficacité. Nous discuterons des conclusions tirées de nos prévisions et de leur pertinence pour les objectifs de notre étude.

Nous fournirons ici les extraits de code clés pour illustrer notre démarche, tandis que les détails et les notebooks spécifiques à chaque poste seront disponibles en **Annexes**.

## 5.1 Collecte des Données

### 5.1.1 Extraction des données

Afin de décider des données à collecter, on a d'abord commencé par l'exploration du fichier "Prévisionnel de la Trésorerie" dans le but d'identifier les postes représentant l'ensemble des actifs de la trésorerie à strésser.

Les postes identifiés sont listés dans le tableau ci-dessous :

Intitulé	Définition
Caisse	Les caisses correspondent aux mouvements journaliers effectués au niveau des agences et DAB de la SGA en plus de la caisse REGIE qui représente les dépenses courantes.
CCP	Compte CCP de la SGA auprès d'Algérie Poste.
Trésor	Ce compte enregistre les avoirs en dinars à vue auprès du Trésor public et du Trésor de wilaya
Scellés BA	Ce sont des sacs scellés, considérés comme des espèces verrouillés en attente de comptage

TABLE 5.2 – Postes significatifs et leurs définitions

Il a été décidé de recueillir la data retraçant l'historique des mouvements à partir de 2018 jusqu'à 2023 (5 ans et quelques mois), et ce avec une fréquence journalière qui a été choisie afin d'avoir la liberté de la modifier par la suite, si nécessaire, en effectuant un traitement d'agrégation.

Au cours de notre stage, nous avons effectué plusieurs extractions de données à partir du système d'information de la banque. Initialement, nous avons employé le logiciel BusinessObjects de SAP pour ces opérations. Par la suite, pour l'extraction finale des mouvements quotidiens associés aux divers postes, nous avons opté pour l'utilisation de DBeaver. Pour ce faire, une requête SQL relativement complexe a été nécessaire. Cette requête a été conçue par le Responsable du Contrôle de Gestion, qui avait une connaissance approfondie des différentes tables de la base de données ainsi que de leur contenu. Sa familiarité avec la structure de la base de données et la précision des données a été indispensable pour formuler une requête qui fournirait les informations précises requises pour notre projet.

### 5.1.2 Importation des données

Nous avons initié notre processus de travail en important nos données dans un environnement dynamique et interactif, précisément sur la plateforme *Google Colab*. Cette plateforme a été sélectionnée spécifiquement pour sa capacité à faciliter l'exploitation du langage de programmation Python, reconnu pour sa puissance et sa flexibilité en matière d'analyse de données. Les données provenaient de deux fichiers distincts au format *Excel*. L'un des fichiers contenait les Mouvements financiers, tandis que l'autre était dédié aux données de Soldes. Notre première étape a consisté à extraire les données relatives aux Mouvements.

Par la suite, nous avons réalisé une extraction portant sur deux soldes, en particulier les soldes au 31/03/2023 et au 30/04/2023. La première extraction a été utilisée pour calculer les soldes sur l'ensemble des dates, tandis que la seconde a été utilisée pour confirmer l'exactitude des calculs et pour s'assurer que les deux soldes extraits correspondaient correctement aux mouvements en notre possession.

Cela nous a permis de vérifier l'intégrité des données et de nous assurer que les soldes extraits étaient cohérents avec les mouvements associés

Calcul du Solde :

$$\text{Solde}_j = \text{Solde}_{j-1} + \text{Mouvement}_j$$

## Indexation : Solde et Mouvements

```
caisse.set_index('dco', inplace=True)
[6] caisseSOLDE.set_index('dco',inplace=True)
```

FIGURE 5.1 – Indexation des données

Nous définissons la colonne "dco" comme index de la DataFrame "caisse" qui correspond ou Mouvements, et la DataFrame "caisseSOLDE" qui correspond au Solde. Cela signifie que la colonne "dco" sera utilisée comme référence temporelle pour les données.

## Rééchantillonnage des données

```
[ ] caisse_hebdo = caisse.resample('W').sum()
[ ] caisseSLD_hebdo= caisseSLD.resample('W').mean()
```

FIGURE 5.2 – Ré-échantillonnage des données par semaine

Nous créons une nouvelle DataFrame appelée "caisse\_hebdo" en agrégeant les données de la DataFrame "caisse" par semaine. La fonction "resample" est utilisée pour regrouper les données selon la fréquence spécifiée, dans ce cas 'W' pour une agrégation hebdomadaire. La méthode "sum()" est utilisée pour calculer la somme des valeurs pour chaque semaine

Nous effectuons une étape similaire pour la DataFrame "caisseSLD". Nous créons une nouvelle DataFrame appelée "caisseSLD\_hebdo" en agrégeant les données par semaine avec une fréquence hebdomadaire ('W'). Cette fois-ci, nous utilisons la méthode "mean()" pour calculer la moyenne des valeurs pour chaque semaine.

## Raisonnement Derrière le Choix du Rééchantillonnage Hebdomadaire

La résolution temporelle de la série temporelle peut avoir un impact significatif à la fois sur la taille des données d'entraînement et sur le nombre de pas à prédire dans le

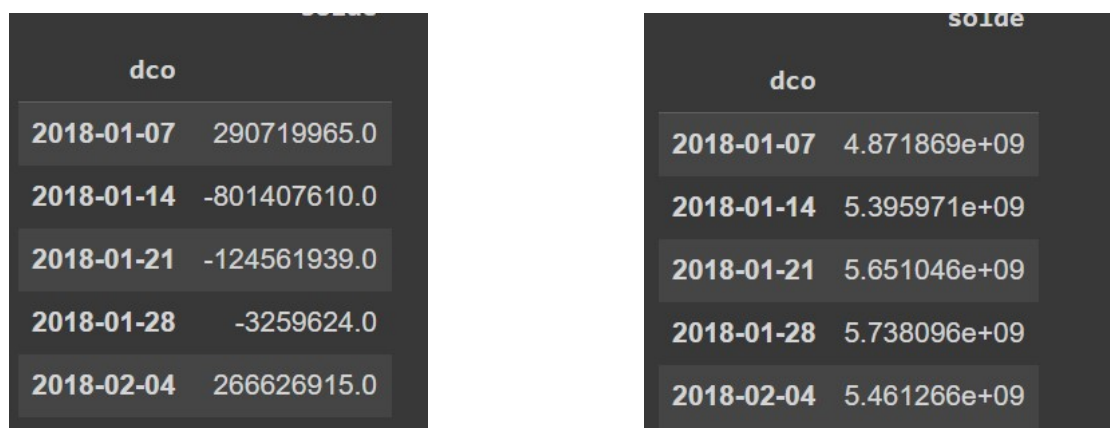
futur. Dans notre cas, il est nécessaire de trouver un équilibre entre ces deux aspects.

Si nous maintenons la résolution à un pas journalier, nous bénéficierions d'une grande quantité de données pour l'entraînement. Cependant, cela entraînerait la nécessité de faire des prévisions pour 365 pas dans le futur, ce qui pourrait entraîner une plus grande incertitude et une diminution potentielle de la précision des prévisions.

Inversement, si nous ré-échantillons la série temporelle à une résolution mensuelle, nous n'aurions que 12 pas à prédire, ce qui simplifierait le problème de prédiction. Cependant, cela réduirait également la quantité de données disponibles pour l'entraînement, ce qui pourrait limiter la capacité du modèle à apprendre des motifs et des tendances dans les données.

En conséquence, le ré-échantillonnage des données à une résolution hebdomadaire semble être un choix judicieux. Il offre un bon compromis en fournissant suffisamment de données pour l'entraînement du modèle, tout en limitant le nombre de pas à prédire à 52. Cela permettrait d'améliorer la robustesse du modèle de prévision, tout en maintenant un niveau de précision acceptable pour les prévisions à un an.

## DataFrames



dco	
2018-01-07	290719965.0
2018-01-14	-801407610.0
2018-01-21	-124561939.0
2018-01-28	-3259624.0
2018-02-04	266626915.0

solde	
2018-01-07	4.871869e+09
2018-01-14	5.395971e+09
2018-01-21	5.651046e+09
2018-01-28	5.738096e+09
2018-02-04	5.461266e+09

(a) Mouvements

(b) Soldes

FIGURE 5.3 – Visualisation du DataFrame de la caisse

Voici le DataFrame dans lequel nous avons calculé les Soldes à partir des Mouvements :

Conformément aux conventions comptables, les actifs sont généralement représentés avec un signe négatif (-). Dans notre approche, nous avons conservé le signe lors du calcul des soldes pour refléter fidèlement cette convention et pour garder la cohérence avec les signes des Mouvements.

Cependant, pour faciliter les calculs et l'interprétation ultérieure, nous avons pris la valeur absolue des soldes, ce qui nous permet d'exprimer tous les soldes en valeurs positives, indépendamment de leur signe d'origine.

```
caisseSOLDE.head()
```

	mvt	solde negatif	solde
dco			
2018-01-02	72363996.0	-5.031733e+09	5031733113
2018-01-03	18621041.0	-5.013112e+09	5013112072
2018-01-04	383860238.0	-4.629252e+09	4629251834
2018-01-07	-184125310.0	-4.813377e+09	4813377144
2018-01-08	-459705367.0	-5.273083e+09	5273082511

FIGURE 5.4 – Données utilisées pour calculer les soldes :

## 5.2 Analyse des Données

### 5.2.1 Statistiques descriptives

#### Mouvements

Le résultat de `df.describe()` fournit des statistiques descriptives pour la série temporelle des mouvements. Voici une interprétation des différentes mesures statistiques données dans le tableau (1.3) :

- **count** : Il y a un total de 275 observations dans la série temporelle des mouvements.
- **mean** : La moyenne des mouvements est d'environ -11,310,070, ce qui indique une tendance négative globale.
- **std** : L'écart-type des mouvements est d'environ 780,429,700, ce qui indique une variabilité significative des valeurs autour de la moyenne.
- **min** : Le mouvement minimum dans la série est d'environ -2,292,276, ce qui représente le mouvement le plus bas enregistré.
- **25%** : Le premier quartile (Q1) est d'environ -463,447,400, ce qui signifie que 25% des mouvements sont inférieurs à cette valeur.
- **50%** : La médiane (deuxième quartile Q2) est d'environ 40,743,520, ce qui représente la valeur centrale de la série des mouvements.
- **75%** : Le troisième quartile (Q3) est d'environ 521,474,500, ce qui signifie que 75% des mouvements sont inférieurs à cette valeur.
- **max** : Le mouvement maximum dans la série est d'environ 2,162,257, ce qui représente le mouvement le plus élevé enregistré.

Ces statistiques descriptives donnent un aperçu global de la distribution et des caractéristiques des mouvements dans la série temporelle. Ils nous aident à comprendre les tendances, la dispersion et la portée des valeurs des mouvements, ce qui est utile pour l'analyse et la modélisation ultérieures.

Statistique	Valeur
Count	275
Mean	-11,310,070
Std	780,429,700
Min	-2,292,276
25%	-463,447,400
50%	40,743,520
75%	521,474,500
Max	2,162,257

TABLE 5.3 – Statistiques descriptives pour la série temporelle des Mouvements de la caisse

## Soldes

Le résultat de `df.describe()` fournit des statistiques descriptives pour la série temporelle du solde. Voici une interprétation des différentes mesures statistiques données dans le tableau (1.4) :

Statistique	Valeur
Count	275
Mean	7,272,456,000
Std	1,310,227,000
Min	3,851,042,000
25%	6,108,871,000
50%	7,348,788,000
75%	8,282,744,000
Max	10,481,760,000

TABLE 5.4 – Statistiques descriptives pour la série temporelle du Solde de la caisse

- **count** : Il y a un total de 275 observations dans la série temporelle du solde.
- **mean** : La moyenne des soldes est d'environ 7,272,456,000, ce qui indique une tendance positive globale.
- **std** : L'écart-type des soldes est d'environ 1,310,227,000, ce qui indique une variabilité significative des valeurs autour de la moyenne.
- **min** : Le solde minimum dans la série est d'environ 3,851,042,000, ce qui représente le solde le plus bas enregistré.
- **25%** : Le premier quartile (Q1) est d'environ 6,108,871,000, ce qui signifie que 25% des soldes sont inférieurs à cette valeur.
- **50%** : La médiane (deuxième quartile Q2) est d'environ 7,348,788,000, ce qui représente la valeur centrale de la série des soldes.
- **75%** : Le troisième quartile (Q3) est d'environ 8,282,744,000, ce qui signifie que 75% des soldes sont inférieurs à cette valeur.
- **max** : Le solde maximum dans la série est d'environ 10,481,760,000, ce qui représente le solde le plus élevé enregistré.

## 5.2.2 Vérification de la complétude et des valeurs manquantes

### Mouvements



```
Vérification de la complétude  
[ ] # Création d'un nouvel index qui couvre toute la plage de dates  
    index_complet = pd.date_range(start=caisseCLEAN.index.min(), end=caisseCLEAN.index.max(), freq='W')  
  
    # Vérification de si l'index de mon DataFrame est identique à cet index complet  
    index_manquant = index_complet.difference(caisseCLEAN.index)  
  
    print(f"Nombre de dates manquantes : {len(index_manquant)}")  
    if len(index_manquant) > 0:  
        print(f"Dates manquantes : {index_manquant}")  
  
Nombre de dates manquantes : 0  
  
Vérification des valeurs manquantes  
[ ] print(caisseCLEAN['solde'].isnull().any())  
  
False
```

FIGURE 5.5 – Vérifications de la complétude et des valeurs manquantes

La série temporelle est complète, ce qui signifie qu'il n'y a pas de dates manquantes. De plus, il n'y a aucune valeur manquante dans les données, ce qui indique que toutes les observations sont présentes et qu'il n'y a pas de données manquantes à traiter.

### Soldes

De même que pour les mouvements, la série temporelle des soldes est intégralement renseignée, signifiant qu'aucune date ne manque. De plus, l'absence de valeurs manquantes dans les données suggère que l'ensemble des observations est complet et ne nécessite aucune procédure de traitement pour combler des données absentes.

### 5.2.3 Visualisation des données

#### Mouvements

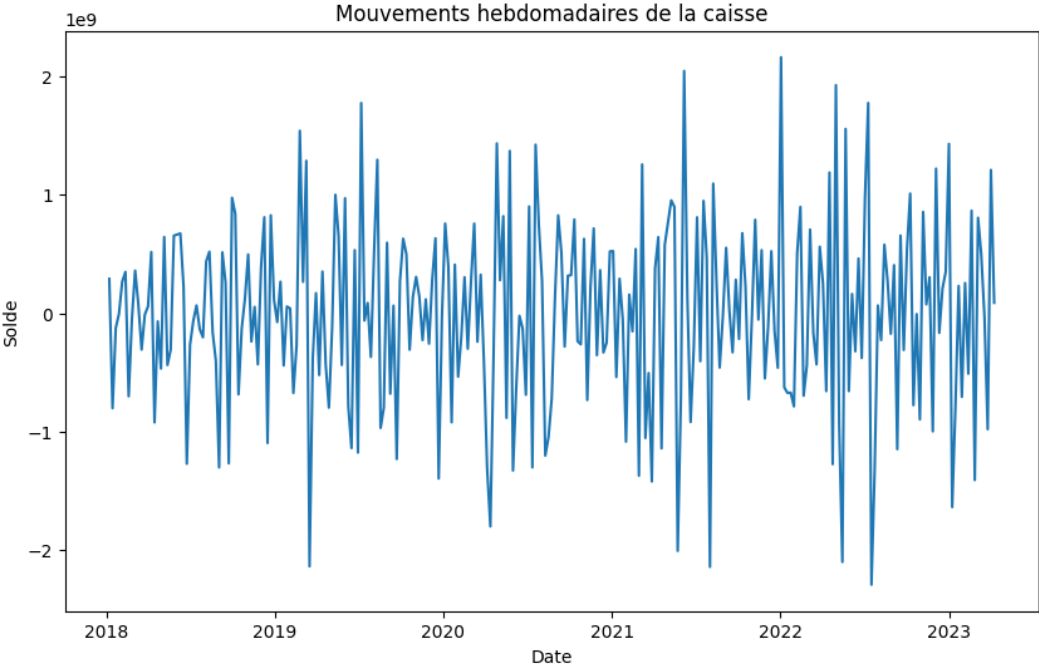


FIGURE 5.6 – Mouvements hebdomadaires de la Caisse

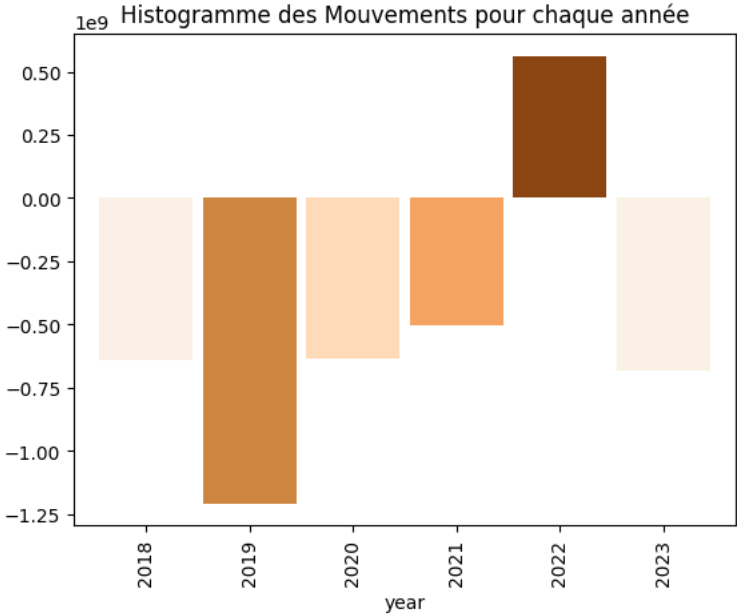


FIGURE 5.7 – Histogramme des Mouvements annuels de la Caisse



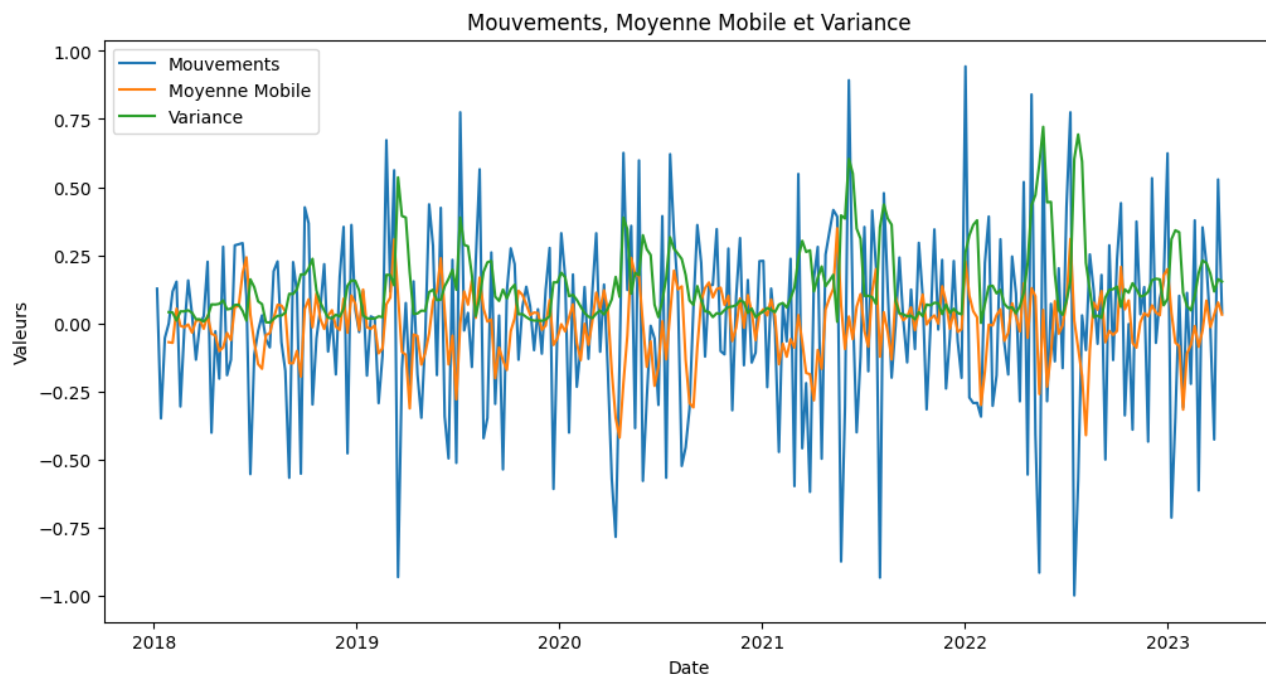


FIGURE 5.8 – Evolution des Mouvements avec Moyenne Mobile et Variance

## Soldes

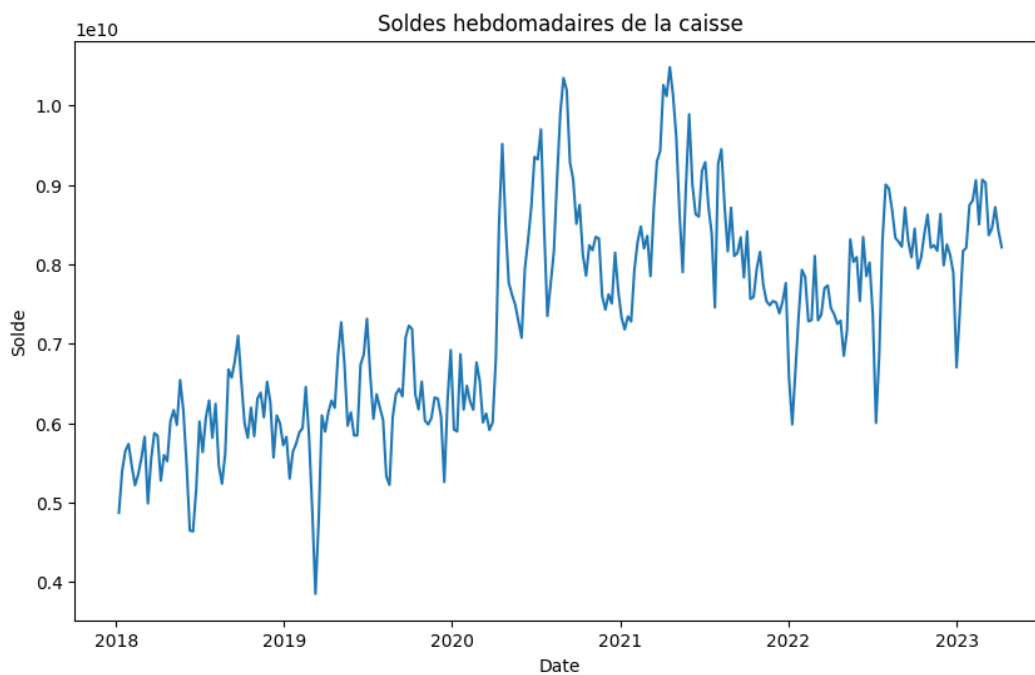


FIGURE 5.9 – Soldes hebdomadaires de la Caisse

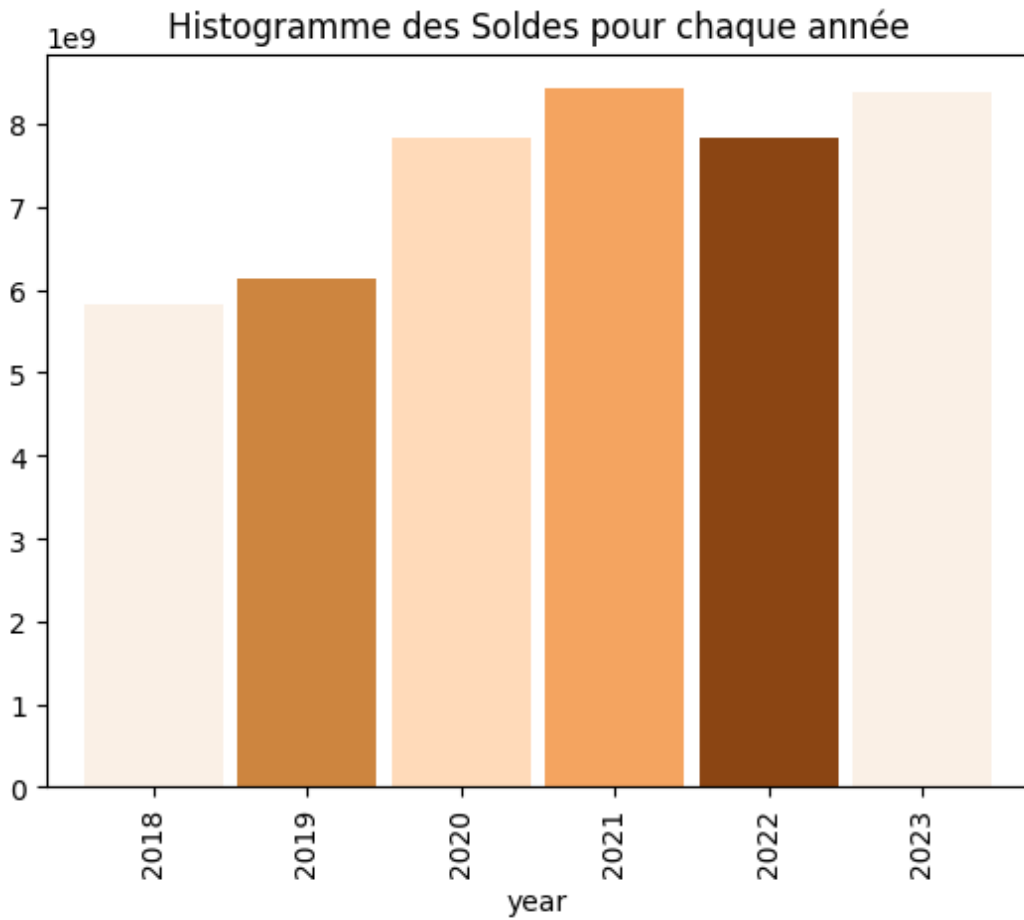


FIGURE 5.10 – Histogramme des Soldes annuels de la Caisse

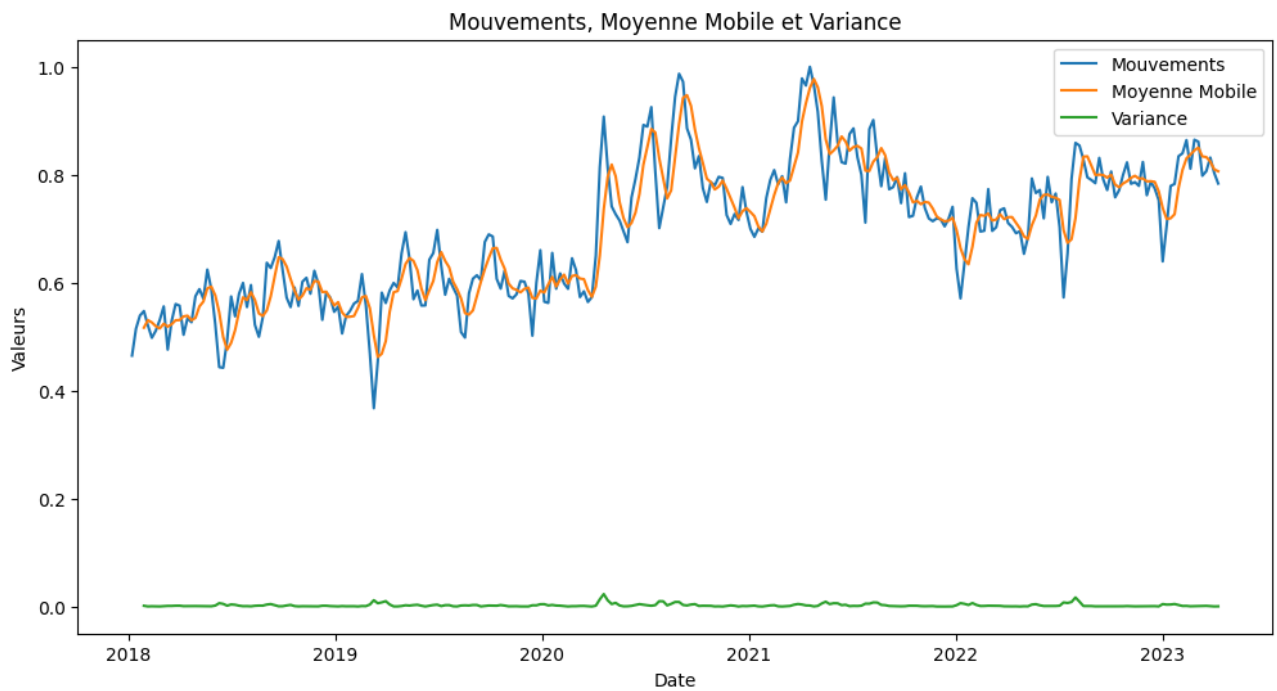


FIGURE 5.11 – Evolution des Soldes avec Moyenne Mobile et Variance

**Interprétation :**

1. On remarque une évolution plus stable pour les soldes.

2. Les soldes ne contiennent que des valeurs positives, contrairement aux mouvements.
3. En comparant les deux histogrammes, On remarque que les mouvements négatifs représentent une augmentation du solde, tandis que les mouvements positifs se traduisent par une baisse du solde (conformément à la convention comptable).
4. En 2020, on enregistre la plus grande augmentation du solde annuel moyen, ceci peut être justifié par le lancement du crédit automobile à 0% ainsi que l'augmentation et améliorations de produits de la SGA.

## 5.2.4 Analyse des valeurs aberrantes

### Mouvements

Le BoxPlot, également connu sous le nom de diagramme en boîte et moustaches, est un outil de visualisation des données statistiques qui permet de représenter graphiquement la distribution et les caractéristiques principales d'un ensemble de données. Il est souvent utilisé pour observer la dispersion, la symétrie, les valeurs extrêmes et les valeurs aberrantes d'une variable.

Le BoxPlot se compose d'une boîte rectangulaire et de deux moustaches (lignes). La boîte représente l'étendue interquartile (IQR), qui englobe le deuxième et le troisième quartile (Q2 et Q3) des données. La médiane est représentée par une ligne à l'intérieur de la boîte.

Les moustaches indiquent la plage des valeurs situées à l'extérieur de l'IQR. Elles sont généralement définies comme étant 1,5 fois l'IQR au-dessus et en dessous de la boîte. Les valeurs extrêmes, qui sont au-delà des moustaches, sont représentées par des points ou des symboles.

Le BoxPlot permet de visualiser rapidement plusieurs informations clés, telles que la médiane, les quartiles, les valeurs extrêmes et les valeurs aberrantes, ce qui aide à comprendre la répartition des données et à détecter les valeurs atypiques. Il est également utile pour comparer plusieurs ensembles de données ou pour observer les variations dans une variable au fil du temps.

**Interprétation :** On peut remarquer qu'il y'a présence de valeurs aberrantes dans cette série temporelle. Ces valeurs, loin d'être le résultat d'erreurs de saisie, sont en réalité le reflet de mouvements significatifs survenus lors de transactions de nivellement (c'est-à-dire, le transfert d'argent vers la Banque d'Algérie) ou lors d'encaissements importants qui ne se produisent que sporadiquement.

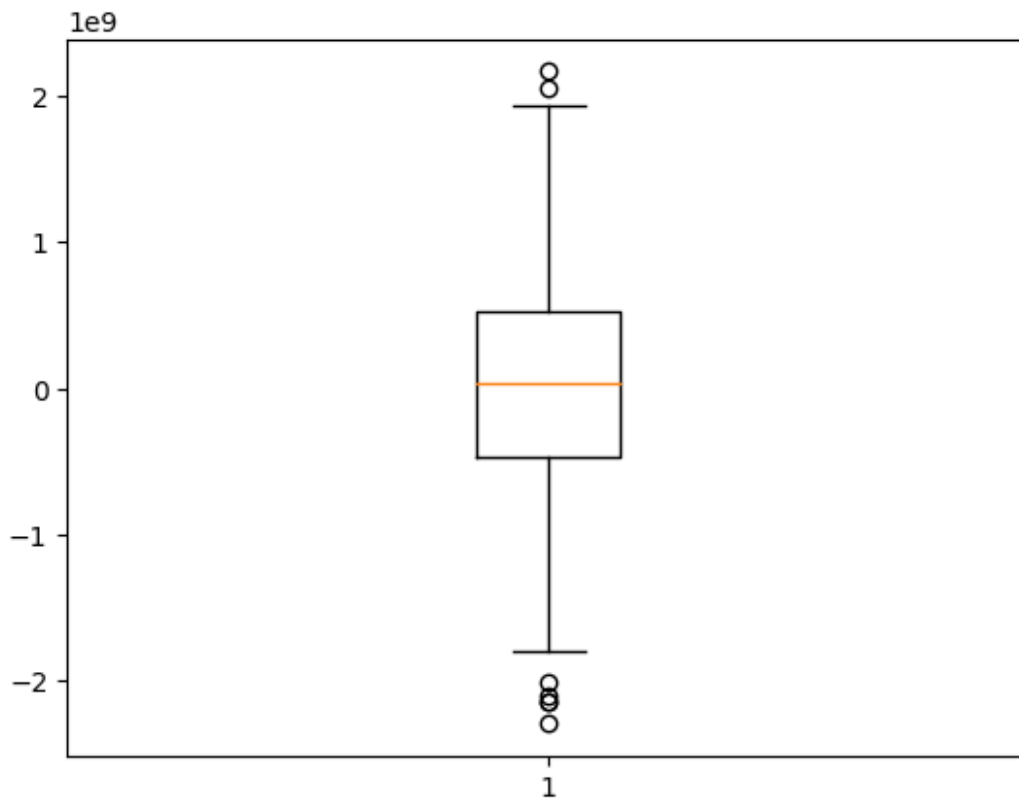


FIGURE 5.12 – Boite à Moustaches (BoxPlot) des mouvements de la caisse

### Soldes

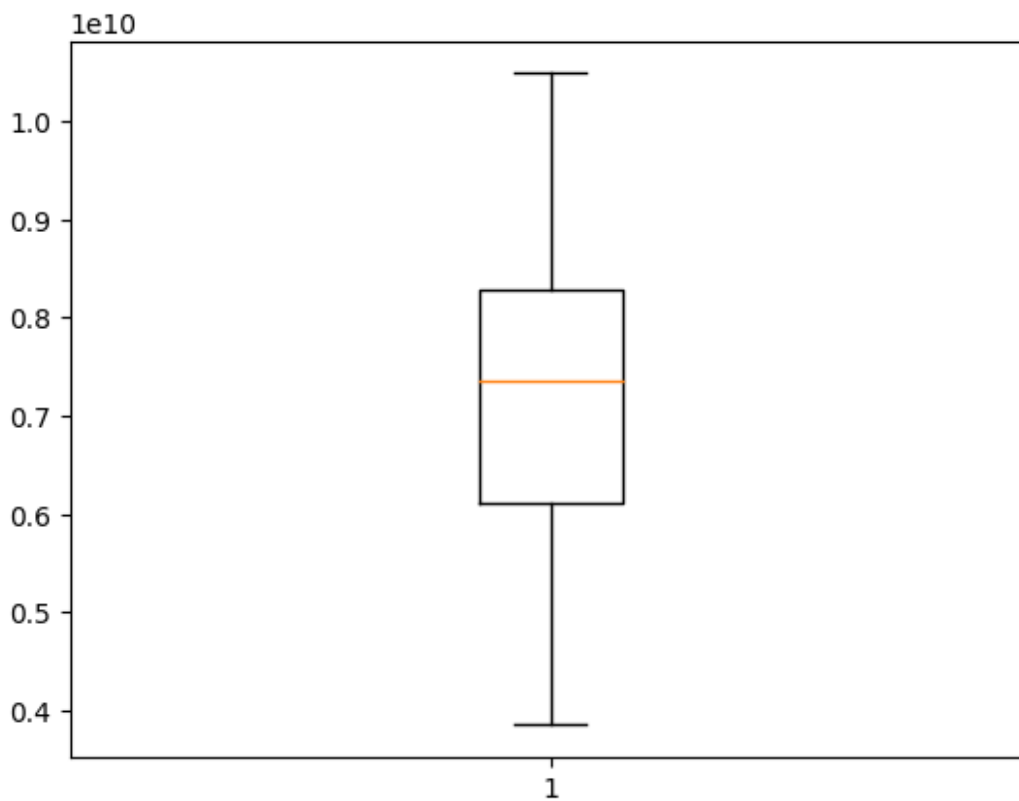


FIGURE 5.13 – Boite à Moustaches (BoxPlot) des soldes de la caisse

**Interprétation :** On peut remarquer que pour la série temporelle du solde, il n'y a pas de valeurs aberrantes. La plage de valeurs des soldes est beaucoup plus petite que celle des mouvements.

### 5.2.5 Décomposition des séries temporelles

La décomposition saisonnière (seasonal decomposition en anglais) est une technique utilisée en analyse des séries temporelles pour décomposer une série en différentes composantes : tendance, saisonnalité et résidus. Cette méthode permet d'identifier les structures sous-jacentes d'une série temporelle et de mieux comprendre les modèles saisonniers et les variations à long terme.

La décomposition saisonnière vise à séparer les différentes composantes d'une série temporelle. Voici une brève description des composantes clés :

**Tendance :** C'est la composante à long terme qui représente l'évolution générale de la série au fil du temps. Elle indique la direction et la tendance générale de la série, qu'elle soit croissante, décroissante ou stable.

**Saisonnalité :** C'est la composante qui représente les variations périodiques régulières dans la série, souvent associées à des effets saisonniers ou cycliques. Elle peut être quotidienne, hebdomadaire, mensuelle ou annuelle, selon la fréquence des variations saisonnières dans les données.

**Résidus :** Ce sont les variations aléatoires ou non expliquées par la tendance et la saisonnalité. Les résidus sont souvent considérés comme le bruit ou l'erreur dans la série temporelle.

La décomposition saisonnière utilise généralement des méthodes mathématiques telles que la méthode de la moyenne mobile ou la décomposition de Fourier pour extraire les différentes composantes d'une série temporelle. Une fois la décomposition effectuée, il devient plus facile d'analyser et de modéliser chaque composante séparément, ce qui peut faciliter les prévisions ou l'identification de modèles spécifiques dans les données.

## Mouvements

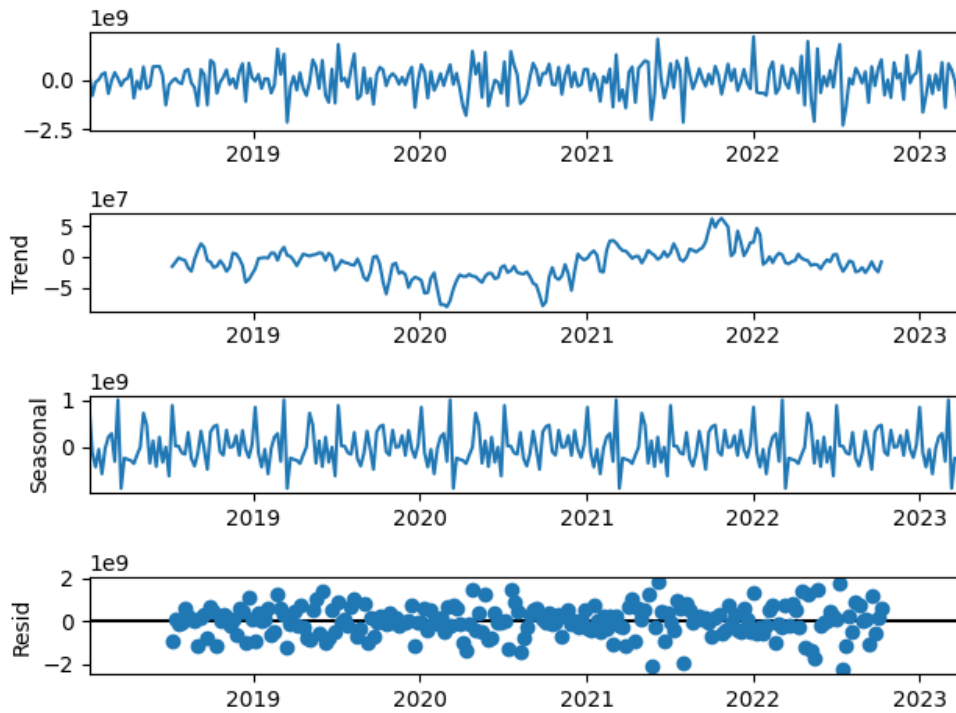


FIGURE 5.14 – Décomposition de la série temporelle des mouvements de la caisse

**Interprétation :** La série temporelle n'a pas de tendance ou de saisonnalité claires.

## Soldes

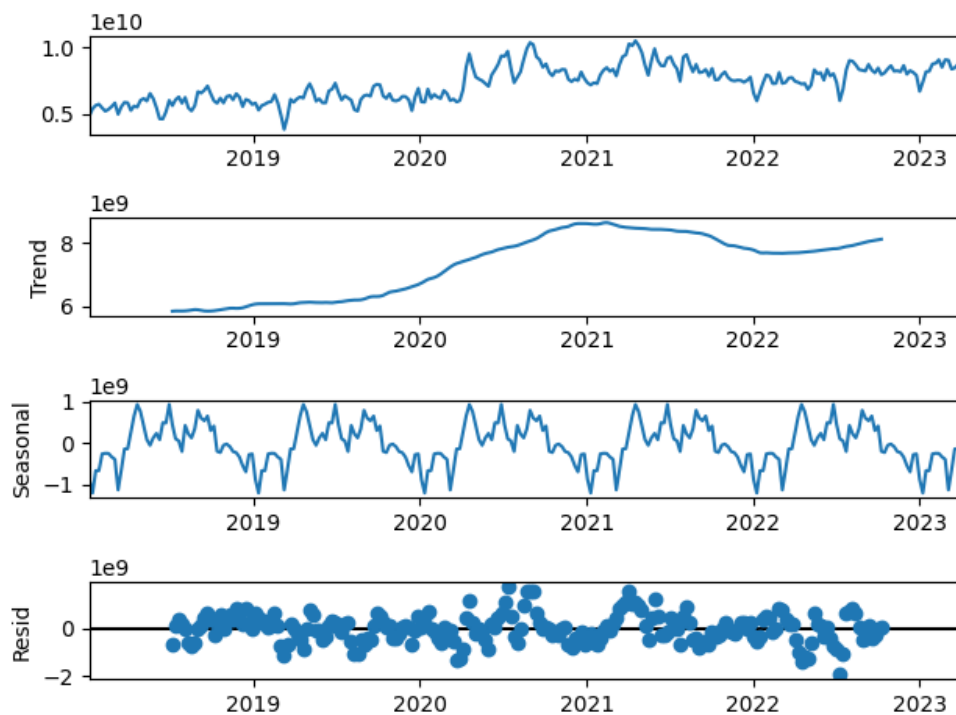


FIGURE 5.15 – Décomposition de la série temporelle du solde de la caisse

**Interprétation :** La série temporelle a une saisonnalité et une tendance à la hausse.

## 5.2.6 Test de stationnarité

L'ADF (Augmented Dickey-Fuller) est un test statistique utilisé en économétrie pour déterminer si une série temporelle est stationnaire ou non. Il est couramment utilisé pour analyser les séries temporelles et détecter la présence de racines unitaires, ce qui indique une non-stationnarité.

La non-stationnarité d'une série temporelle peut être problématique lors de l'analyse et de la modélisation, car elle peut rendre difficile la prévision et la prise de décision basée sur les tendances et les modèles de données. L'ADF permet d'évaluer si une série temporelle suit un processus de marche aléatoire ou un processus stationnaire.

Le test ADF est basé sur le modèle autorégressif (AR) et utilise la régression pour estimer la présence de racines unitaires dans la série temporelle. La version augmentée du test Dickey-Fuller permet de prendre en compte les retards supplémentaires et de mieux traiter les séries temporelles avec des tendances complexes.

L'hypothèse nulle du test ADF est que la série temporelle possède une racine unitaire, ce qui signifie qu'elle est non stationnaire. Si la valeur de test ADF est inférieure à une certaine valeur critique, l'hypothèse nulle est rejetée, ce qui indique que la série temporelle est stationnaire. En revanche, si la valeur de test ADF est supérieure à la valeur critique, l'hypothèse nulle est acceptée, ce qui indique la non-stationnarité de la série temporelle.

Le test ADF est largement utilisé dans l'analyse des séries temporelles, notamment dans la modélisation économique, la finance et d'autres domaines où la stationnarité des données est importante pour les prévisions et les analyses. Il fournit une approche statistique pour évaluer la présence de racines unitaires et la stationnarité des séries temporelles, ce qui aide à mieux comprendre les modèles et les tendances dans les données.

### Mouvements

Les résultats du test ADF (Augmented Dickey-Fuller) sont les suivants :

- ADF Statistic : -5.508154482307871
- p-value : 2.000025210657827e-06
- Num Lags : 15
- Num Observations Used For ADF Regression and Critical Values Calculation : 259
- Critical Values :
  - 1% : -3.4558530692911504
  - 5% : -2.872764881778665
  - 10% : -2.572751643088207

**Interprétation :** La statistique ADF (-5.508154482307871) est inférieure aux valeurs critiques à 1%, 5% et 10%, ce qui suggère que nous pouvons rejeter l'hypothèse nulle de non-stationnarité de la série temporelle. Le p-value (2.000025210657827e-06) est également très faible, ce qui renforce la conclusion selon laquelle la série temporelle est stationnaire. Le nombre de retards utilisés dans la régression est de 15, et 259 observations ont été utilisées pour effectuer le test et calculer les valeurs critiques.

En résumé, les résultats du test ADF indiquent que la série temporelle est stationnaire, ce qui est important pour l'analyse et la modélisation des données.

```
[ ] from statsmodels.tsa.stattools import adfuller

caisseADF= caisse_hebdo.copy()

def adf_test(series):
    result = adfuller(series)
    print(f'ADF Statistic: {result[0]}')
    print(f'p-value: {result[1]}')
    print(f'Num Lags: {result[2]}')
    print(f'Num Observations Used For ADF Regression and Critical Values Calculation: {result[3]}')
    print('Critical Values:')
    for key, value in result[4].items():
        print(f'    {key}, {value}')

adf_test(caisseADF['solde'])

ADF Statistic: -5.508154482307871
p-value: 2.000025210657827e-06
Num Lags: 15
Num Observations Used For ADF Regression and Critical Values Calculation: 259
Critical Values:
1%, -3.4558530692911504
5%, -2.872764881778665
10%, -2.572751643088207
```

FIGURE 5.16 – ADF Test sur les Mouvements de la Caisse

## Soldes

- ADF Statistic : -1.5589364501645024
- p-value : 0.5041813534329218
- Num Lags : 8
- Num Observations Used For ADF Regression and Critical Values Calculation : 266
- Critical Values :
  - 1% : -3.455175292841607
  - 5% : -2.8724677563219485
  - 10% : -2.57259315846006

**Interprétation :** La statistique ADF (-1.5589364501645024) est supérieure aux valeurs critiques à 1%, 5% et 10%, ce qui suggère que nous ne pouvons pas rejeter l'hypothèse nulle de non-stationnarité de la série temporelle. Le p-value (0.5041813534329218) est également relativement élevé, ce qui renforce la conclusion selon laquelle la série temporelle n'est pas nécessairement stationnaire. Le nombre de retards utilisés dans la régression est de 8, et 266 observations ont été utilisées pour effectuer le test et calculer les valeurs critiques.

En résumé, les résultats du test ADF indiquent que la série temporelle n'est pas stationnaire, ce qui peut avoir des implications sur l'analyse et la modélisation des données.

## 5.2.7 Autocorrélation et Autocorrélation Partielle

L'autocorrélation est une mesure statistique utilisée en analyse des séries temporelles pour quantifier la corrélation entre une observation et les observations précédentes dans la même série temporelle. Elle permet de détecter les motifs de corrélation et les dépendances entre les valeurs successives d'une variable.



Plus précisément, l'autocorrélation mesure la similarité entre les valeurs d'une série temporelle à différents décalages (lags). Elle peut prendre des valeurs allant de -1 à 1, où -1 indique une corrélation négative parfaite, 0 indique aucune corrélation et 1 indique une corrélation positive parfaite.

L'autocorrélation partielle, quant à elle, mesure la corrélation entre une observation et ses décalages (lags), tout en éliminant l'influence des décalages intermédiaires. En d'autres termes, elle permet d'examiner la corrélation directe entre deux observations après avoir tenu compte de la corrélation indirecte via les autres observations intermédiaires.

L'autocorrélation et l'autocorrélation partielle sont souvent représentées graphiquement à l'aide de fonctions d'autocorrélation et de fonctions d'autocorrélation partielle, respectivement. Ces graphiques permettent de visualiser les motifs de corrélation et d'identifier les décalages significatifs dans la série temporelle.

Ces mesures d'autocorrélation et d'autocorrélation partielle sont couramment utilisées pour l'analyse des séries temporelles, notamment dans la modélisation et la prévision. Elles fournissent des informations sur la structure de dépendance temporelle des données, ce qui permet de sélectionner des modèles appropriés et d'estimer les paramètres pour une meilleure prédiction des valeurs futures.

## Mouvements

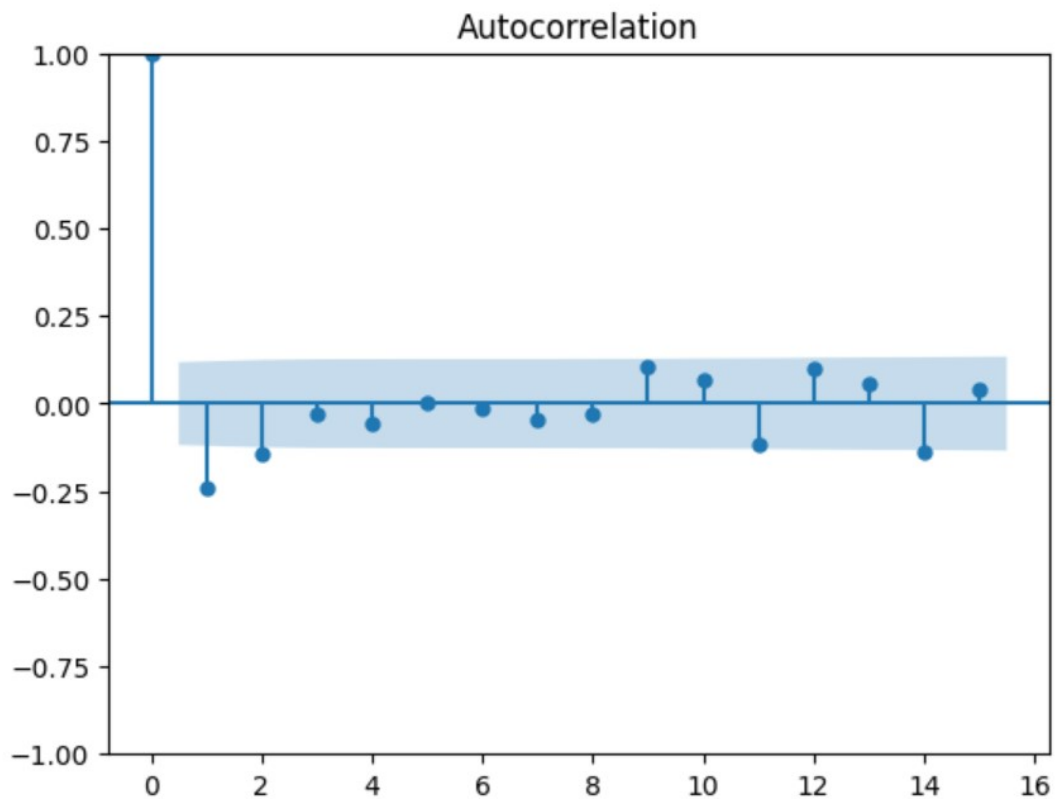


FIGURE 5.17 – Fonction d'autocorrélation des mouvements de la caisse

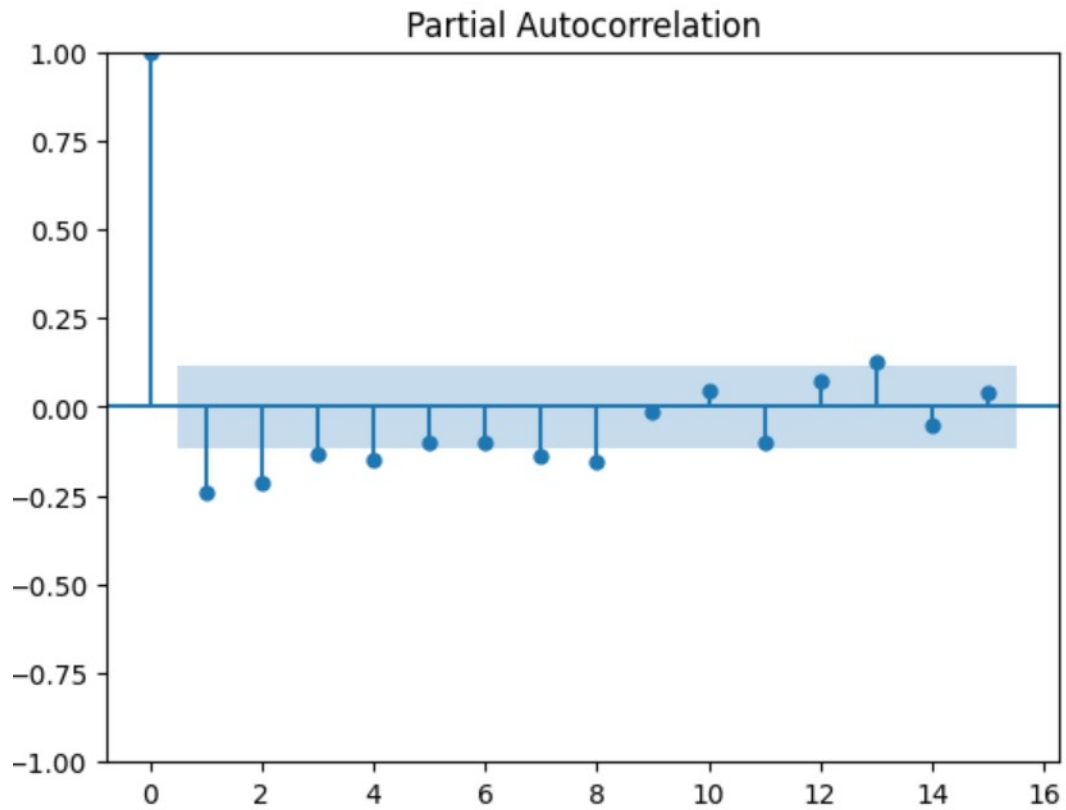


FIGURE 5.18 – Fonction d'autocorrelation partielle des mouvements de la caisse

**Interprétation :** L'observation indique que la série temporelle présente une faible corrélation avec elle-même, c'est-à-dire qu'elle n'exhibe ni autocorrélation significative ni autocorrélation partielle notable.

### Soldes

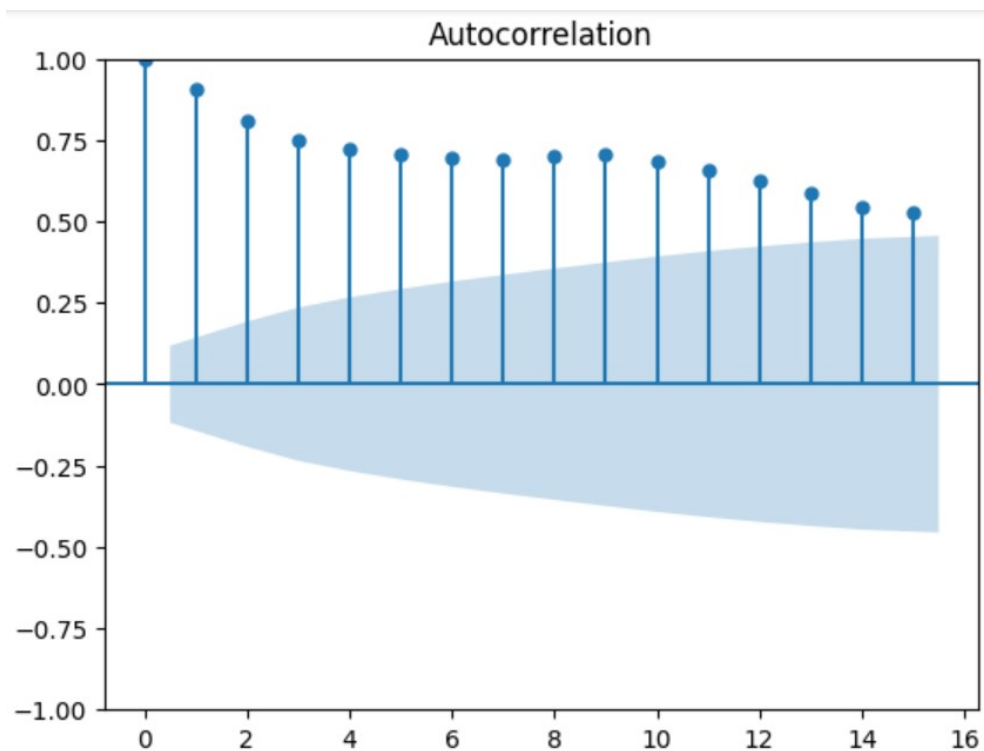


FIGURE 5.19 – Fonction d'autocorrelation du solde de la caisse

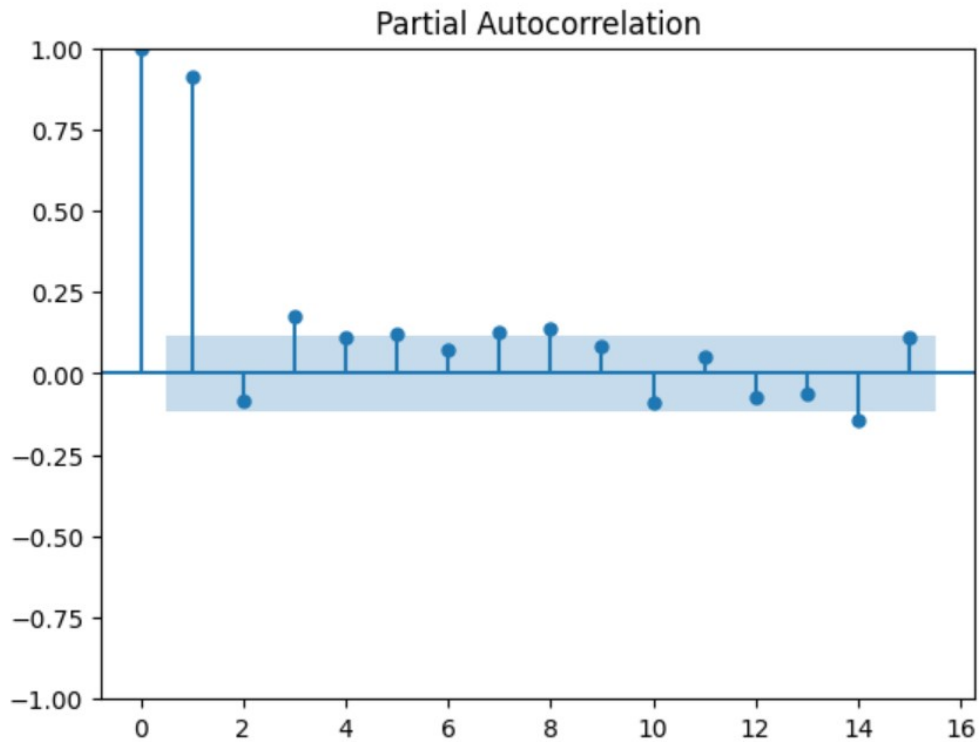


FIGURE 5.20 – Fonction d'autocorrélation partielle du solde de la caisse

**Interprétation :** La série temporelle présente une forte corrélation avec elle-même, indiquant une autocorrélation significative et une autocorrélation partielle notable.

### 5.2.8 Vérification de la normalité

Le QQ plot (Quantile-Quantile plot) est un outil graphique utilisé pour évaluer la normalité d'un échantillon de données en le comparant à une distribution théorique, généralement la distribution normale. Il permet de visualiser si les quantiles observés dans l'échantillon correspondent aux quantiles attendus d'une distribution normale.

Le QQ plot affiche les quantiles observés de l'échantillon sur l'axe des ordonnées (y) et les quantiles attendus d'une distribution normale sur l'axe des abscisses (x). Si les points du QQ plot sont alignés approximativement le long d'une ligne droite, cela suggère une bonne concordance entre les quantiles observés et les quantiles attendus d'une distribution normale. En revanche, si les points s'écartent de la ligne droite, cela indique une déviation par rapport à la normalité.

Plus précisément, le QQ plot compare les valeurs empiriques (quantiles observés) aux valeurs théoriques (quantiles attendus) de la distribution normale. Si les valeurs observées et attendues sont proches les unes des autres, cela indique une bonne adéquation à la distribution normale. Si les valeurs observées sont plus éloignées des valeurs attendues, cela suggère une déviation de la normalité.

Le QQ plot est un outil visuel utile pour évaluer la normalité des données car il permet de repérer facilement les déviations de la distribution normale. Il est couramment utilisé en statistique, en économétrie et dans d'autres domaines où l'évaluation de la normalité est importante pour les analyses et les modélisations statistiques.

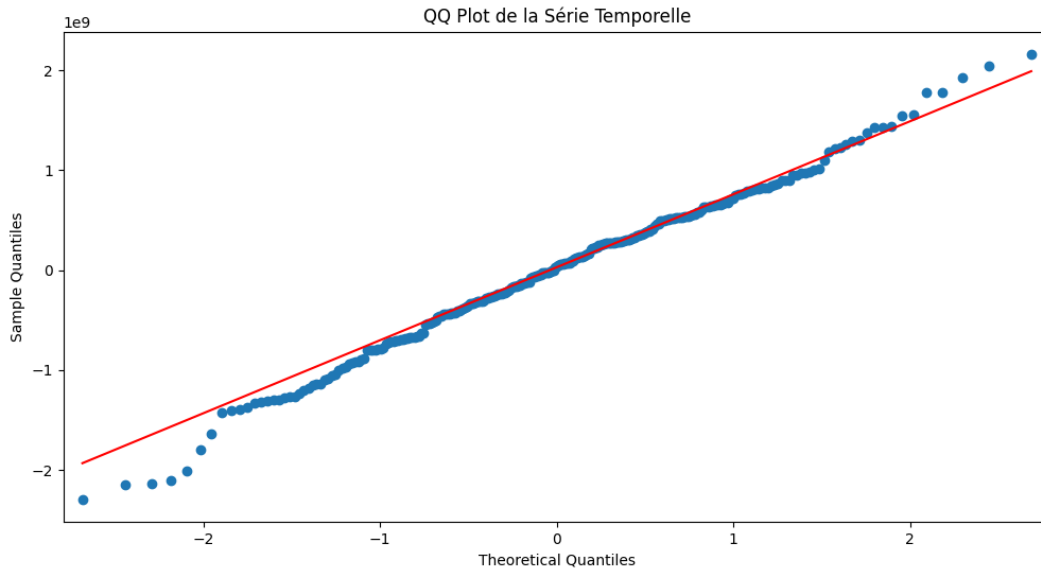


FIGURE 5.21 – QQ Plot des mouvements de la caisse

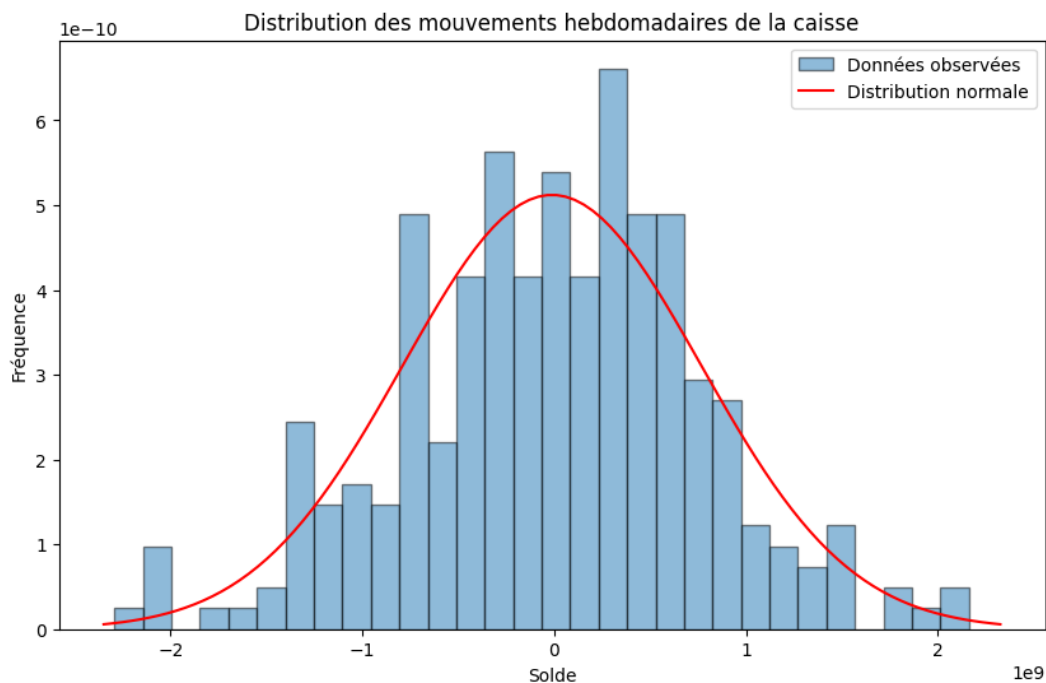


FIGURE 5.22 – Distribution des mouvements hebdomadaires de la caisse

## Mouvements

### Interprétation :

En ce qui concerne le QQ-Plot, la majorité des points de quantiles observés suivent approximativement la ligne diagonale, cela suggère que les données sont bien ajustées à la distribution de référence (dans ce cas, la distribution normale). C'est un indicateur de bonne adéquation. Cependant, plusieurs points s'éloignent de manière significative de la ligne diagonale, cela peut indiquer une déviation des données par rapport à la distribution de référence. Des déviations vers le haut ou vers le bas peuvent suggérer un excès ou une insuffisance de quantiles extrêmes par rapport à la distribution de référence.

L'histogramme de distribution, quant à lui, est un graphique qui représente la ré-

partition des valeurs d'un échantillon dans des intervalles (bins) prédéfinis. Il montre la fréquence ou la densité des observations dans chaque intervalle. Cela permet de visualiser la forme de la distribution des données et de détecter des caractéristiques telles que la symétrie, l'asymétrie, les pics, ou les queues épaisses.

Le lien entre le QQ plot et l'histogramme de distribution réside dans leur objectif commun d'analyser les propriétés de distribution des données. Ensemble, ils fournissent une perspective plus complète de la distribution des données :

Là où le QQ-plot présente une courbure prononcée ou une courbure inversée par rapport à la ligne diagonale de référence, cela suggère une différence entre la distribution des quantiles observés et la distribution de référence. Dans l'histogramme de distribution, on peut remarquer des signes de cette différence, tels que des pics ou des queues épaisses qui ne correspondent pas à la forme attendue de la distribution de référence. Les quantiles observés dans le QQ plot s'écartent de manière significative de la ligne diagonale aux extrémités, cela peut indiquer une queue plus lourde ou une distribution des valeurs extrêmes différente de celle de la distribution de référence.

Pour conclure, On peut remarquer des valeurs aberrantes ou des plages de valeurs dans l'histogramme de distribution qui correspondent à ces déviations.

## Soldes

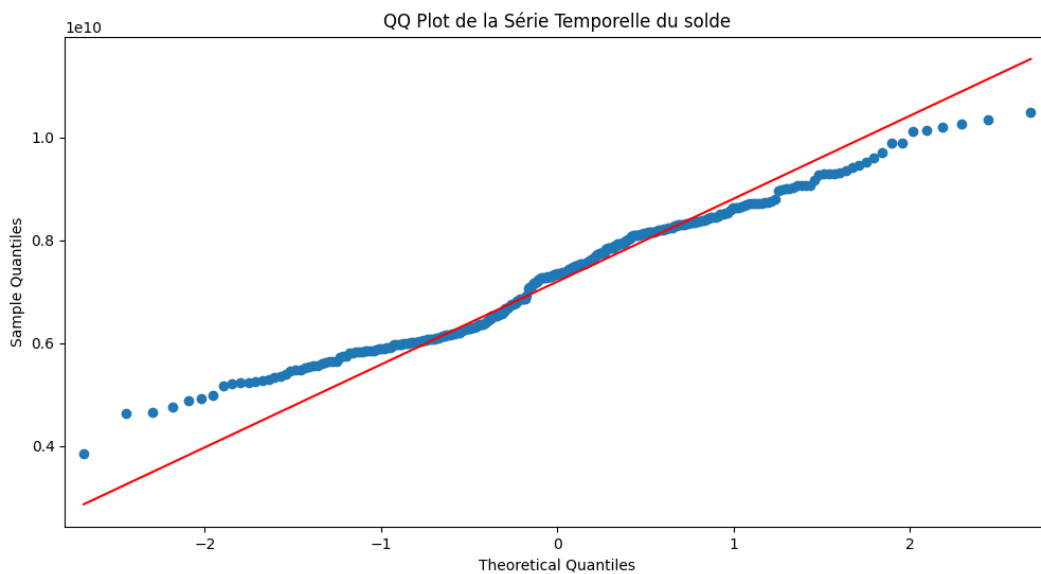


FIGURE 5.23 – QQ Plot du solde de la caisse

**Interprétation :** On remarque ici que le distribution des soldes ne suit pas la distribution normale.

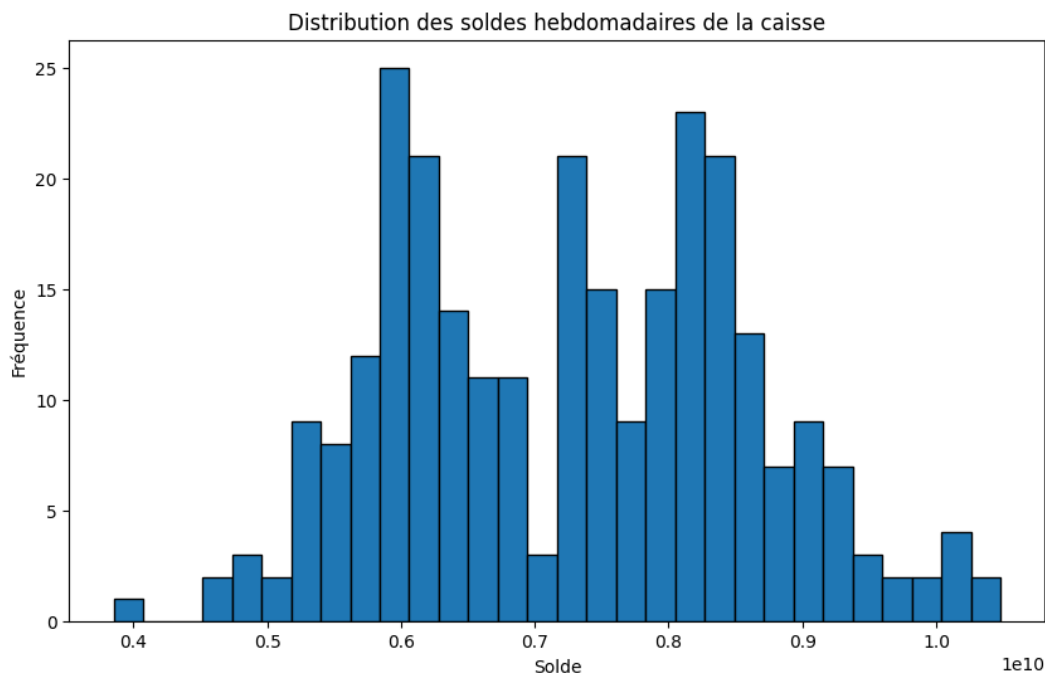


FIGURE 5.24 – Distribution du solde hebdomadaires de la caisse

### 5.2.9 Analyse de la volatilité

L'analyse de la volatilité est une technique d'évaluation statistique qui cherche à déterminer la variation des prix d'un actif financier sur une certaine période. L'actif peut être une action, une obligation, une devise, un indice boursier, etc. La volatilité est un indicateur de risque, montrant l'incertitude ou le risque associé à la taille des changements dans la valeur d'un actif.

Dans l'analyse de la volatilité, deux mesures principales sont généralement utilisées :

1. **Volatilité historique** : Il s'agit de la fluctuation des prix d'un actif au cours d'une période précédente. Elle est calculée à l'aide de techniques statistiques, comme l'écart type ou la variance, pour mesurer l'écart des rendements de l'actif par rapport à sa moyenne.
2. **Volatilité implicite** : Elle est dérivée du prix d'une option ou d'un autre produit dérivé et représente les attentes du marché en matière de volatilité future. Plus la volatilité implicite est élevée, plus les traders s'attendent à de grands mouvements de prix.

Il est important de noter que la volatilité n'est pas une indication de la direction des mouvements de prix, mais simplement de leur amplitude. Une forte volatilité signifie que le prix de l'actif peut changer de manière significative en très peu de temps, ce qui peut offrir des opportunités d'investissement mais aussi augmenter le risque.

Dans le domaine des mathématiques financières, l'analyse de la volatilité implique généralement des concepts tels que la variance, l'écart type et les modèles stochastiques. Voici une formule pour la volatilité (à travers l'écart type) d'une série de valeurs  $x$ , avec une fenêtre de taille "n" :

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2} \quad (5.1)$$

L'écart type est une mesure statistique de la dispersion des valeurs dans un ensemble de données. Plus l'écart type est élevé, plus les valeurs sont éloignées de la moyenne (et vice versa).

## Mouvements

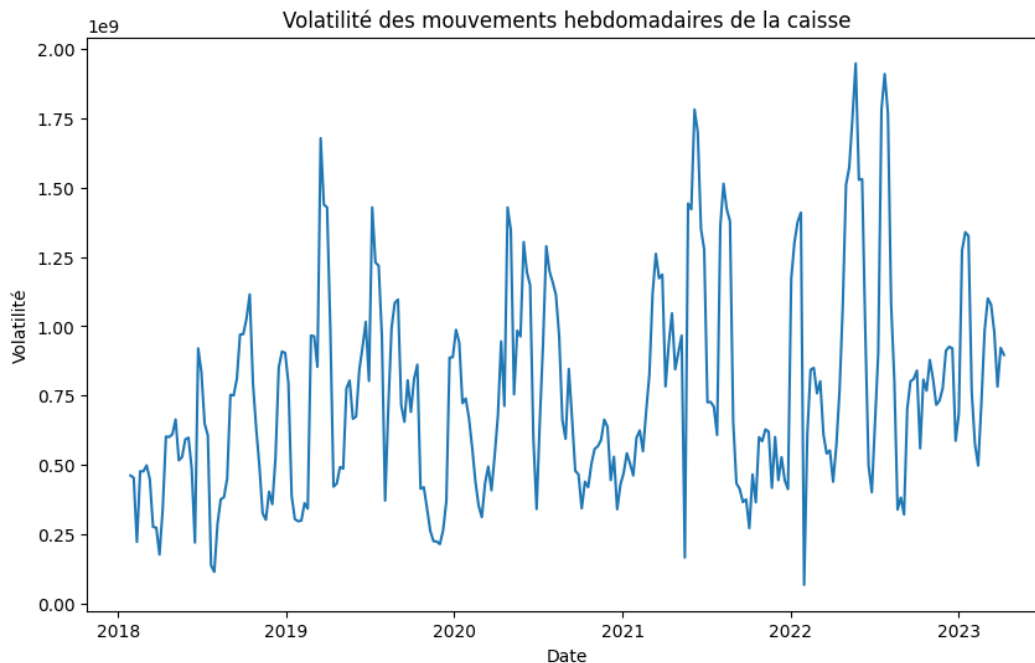


FIGURE 5.25 – Analyse de la volatilité des mouvements hebdomadaires de la caisse

**Interprétation :** Dans notre cas, la volatilité est définie comme l'écart type des soldes hebdomadaires sur une fenêtre mobile de 4 semaines.

Là où la volatilité est plus élevée (mi-2022 par exemple) indique que les mouvements de la caisse fluctuent de manière plus imprévisible, ce qui pourrait indiquer un risque plus élevé.

Le fait d'utiliser une fenêtre mobile pour le calcul de la volatilité signifie que pour chaque point dans le temps, on ne regarde que les 4 semaines précédentes pour le calcul. C'est une méthode courante pour estimer la volatilité dans le temps, car elle donne une idée de la volatilité récente plutôt que de la volatilité sur toute la période de données.

## Solde

**Interprétation :** Comme attendu, on voit que la série temporelle des soldes est moins volatile que celle des mouvements.

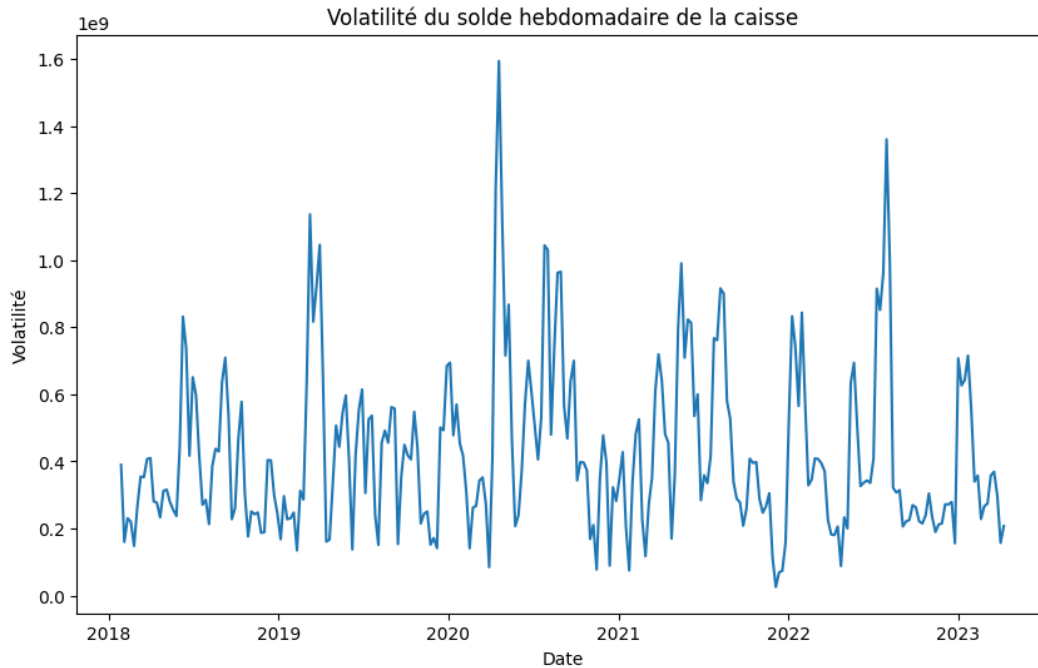


FIGURE 5.26 – Analyse de la volatilité du solde hebdomadaire de la caisse

### 5.3 Préparation des Données et Modélisation

Dans la section précédente, nous avons effectué une analyse approfondie de nos données. Nous avons alors été en mesure de comprendre profondément les caractéristiques intrinsèques de nos séries temporelles et de nous préparer à l'étape cruciale suivante : la modélisation.

La modélisation est une étape essentielle de toute analyse prédictive. Elle consiste à utiliser des algorithmes statistiques ou de machine learning pour prédire une variable d'intérêt. Pour garantir que nos modèles peuvent généraliser efficacement à partir de nos données et produire des prédictions précises sur des données non vues, on se devrait de diviser notre jeu de données en deux sous-ensembles : l'ensemble d'apprentissage (ou "training set") et l'ensemble de test (ou "test set").

Dans cette section, nous allons d'abord examiner une liste de modèles pré-identifiés à partir de l'état de l'art dans le domaine, dans le but de déterminer les techniques de modélisation les plus pertinentes pour notre analyse. Nous passerons en revue l'architecture des modèles sélectionnés, en expliquant de manière approfondie leur structure et leurs principes de fonctionnement.

Par la suite, nous adapterons nos données pour qu'elles correspondent aux exigences structurelles de chaque modèle sélectionné, ce qui nous permettra d'entraîner efficacement ces modèles.

Bien que cette phase puisse s'avérer complexe, elle est indispensable pour générer des prédictions précises et utiles à partir de nos données.

De la même manière que pour la section précédente, les points de discussion dans cette section se concentreront sur la caisse. Il est à noter que le même processus sera appliqué pour les trois autres postes mentionnés.



### 5.3.1 Identification des modèles

Les modèles identifiés dans la section *Étude Bibliographique* de la partie 4 sont les suivants :

- FB Prophet
- Transformers
- Generative Adversarial Networks (GAN)
- GRU (Gated Recurrent Unit)
- LSTM
- RNN (Réseaux de Neurones Récurrents)
- Graphical Models
- Autoencoder

Parmi ces modèles, certains ont été exploités lors de la phase de prévision, tandis que d'autres ont trouvé leur application dans le cadre de la génération de scénarios de stress. Il est crucial de discuter de la pertinence et de l'adaptabilité de ces modèles en ce qui concerne notre première tâche de prévision pour nos séries temporelles comprenant 275 instances. Le choix du modèle dépend fortement de la particularité de nos données :

- **FB Prophet** : Ce modèle, développé par Facebook, est spécialement conçu pour gérer des séries temporelles ayant une forte saisonnalité et tendance. Si nous attendons des variations saisonnières dans nos données, FB Prophet pourrait être notre choix optimal.
- **Transformers** : Ces modèles sont excellents pour les données séquentielles, particulièrement quand il y a des dépendances à long terme. Cependant, ils pourraient nécessiter une quantité plus importante de données pour être entraînés efficacement. Considérant la taille de notre ensemble de données (275 instances), les Transformers pourraient ne pas être le meilleur choix.
- **Generative Adversarial Networks (GAN)** : Les GANs sont principalement utilisés pour générer de nouvelles données qui ressemblent à notre ensemble de données initial. Ils pourraient ne pas être le meilleur choix pour la prévision de séries temporelles, particulièrement pour notre jeu de données de taille relativement petite.
- **GRU (Gated Recurrent Unit), LSTM, RNN (Réseaux de Neurones Récurrents)** : Ces modèles sont très efficaces pour traiter des données séquentielles, ce qui les rend potentiellement appropriés pour notre série temporelle. Les GRU et LSTM, qui sont des variantes des RNN, sont capables de capturer des dépendances à long terme dans les données.
- **Graphical Models** : Ces modèles probabilistes sont utiles pour représenter les dépendances complexes entre variables. Ils pourraient ne pas être le meilleur choix pour notre problème de prévision de séries temporelles.
- **Autoencoder** : Les autoencodeurs sont des réseaux de neurones utilisés pour l'apprentissage non supervisé. Ils pourraient ne pas être idéaux pour une tâche de prévision de série temporelle.

En somme, si nous cherchons un modèle capable de gérer les tendances et la saisonnalité de manière simple et efficace, FB Prophet pourrait être notre premier choix. Si nous soupçonnons que des dépendances à long terme sont présentes dans nos données, nous pourrions envisager d'utiliser un LSTM ou un GRU.

### 5.3.2 Sélection des modèles

Dans notre quête pour développer le meilleur modèle de prévision pour notre ensemble de données de séries temporelles, nous avons examiné une variété de techniques de modélisation modernes. Ces techniques, identifiées à travers une étude bibliographique approfondie de l'état de l'art, comprennent le FB Prophet, les Transformers, les Generative Adversarial Networks (GANs), les Gated Recurrent Units (GRU), les Long Short-Term Memory (LSTM), les Réseaux de Neurons Récurents (RNN), les Graphical Models et les Autoencoders.

Cependant, il est important de noter que toutes les techniques ne sont pas équivalentes dans leur capacité à traiter et à modéliser des séries temporelles. Chaque technique a ses propres forces et faiblesses, et sa pertinence pour notre projet dépend fortement des caractéristiques spécifiques de nos données.

Après un examen soigneux des spécificités de nos données et des capacités des différents modèles, nous avons retenu trois techniques pour une exploration plus approfondie : les Réseaux de Neurons Récurents (RNN), les Long Short-Term Memory (LSTM), et le FB Prophet. Ces modèles ont été choisis pour leur capacité à gérer efficacement les données séquentielles, à capturer les dépendances à long terme, et à modéliser la tendance et la saisonnalité, des caractéristiques qui sont très pertinentes pour notre ensemble de données de séries temporelles.

#### RNN (Réseaux de Neurons Récurents)

Les RNN sont une classe de réseaux de neurones qui sont particulièrement bien adaptés pour traiter des données séquentielles, comme les séries temporelles. Ils ont cette capacité unique de relier des informations passées à l'état présent, ce qui permet de capter les dépendances temporelles dans les données.

L'architecture de base d'un RNN comprend un noeud ou une unité cachée qui reçoit une entrée à chaque pas de temps et donne une sortie. L'état caché à l'étape de temps actuelle est calculé en utilisant l'état caché et l'entrée de l'étape de temps précédente. Ceci est réalisé par une fonction d'activation, généralement la fonction tanh ou ReLU.

La principale raison pour laquelle les RNN ont été choisis pour ce projet est leur capacité à modéliser des dépendances temporelles, ce qui est crucial pour notre problème de prévision de séries temporelles.

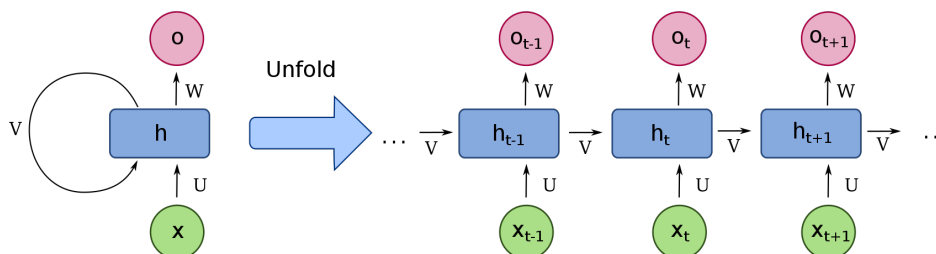


FIGURE 5.27 – Fonctionnement général d'un RNN

L'explication détaillée de ce fonctionnement général se trouve dans la définition des modèles dans la section *Etude Bibliographique de l'Etat de l'Art*

## LSTM (Long Short-Term Memory)

Les LSTM sont une variante spécifique des RNN conçus pour éviter les problèmes de disparition ou d'explosion du gradient qui peuvent survenir lors de l'entraînement des RNN. Ils sont capables de se souvenir et d'apprendre de séquences de données plus longues que les RNN traditionnels, ce qui les rend très adaptés pour notre tâche de prévision de séries temporelles.

L'architecture d'un LSTM comprend trois portes - la porte d'entrée, la porte d'oubli et la porte de sortie, qui ensemble permettent au LSTM de conserver ou d'oublier des informations en fonction de leur importance. Ceci est essentiel pour capturer les dépendances à long terme dans les données.

La sélection des LSTM pour ce projet est motivée par leur capacité à modéliser efficacement des dépendances à long terme dans les données de séries temporelles.

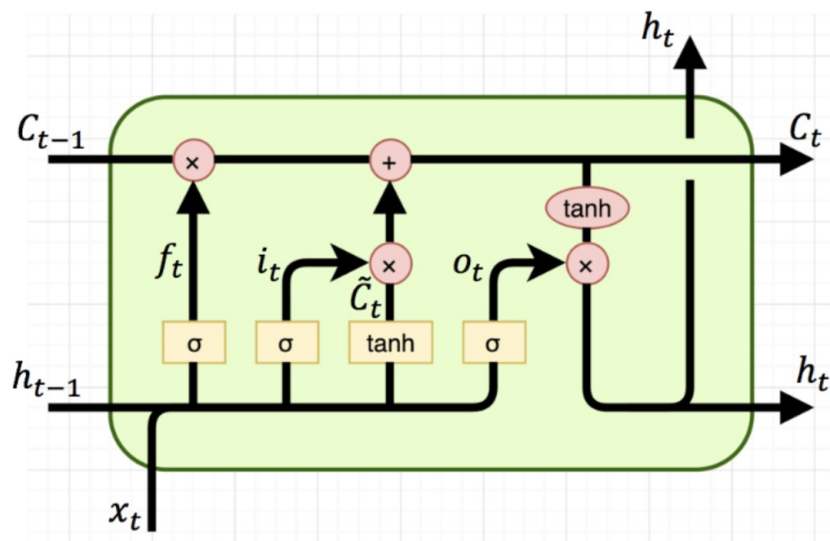


FIGURE 5.28 – Structure générale d'une cellule LSTM

L'explication détaillée cette structure générale se trouve dans la définition des modèles dans la section *Etude Bibliographique de l'Etat de l'Art*

## FB Prophet

FB Prophet est un modèle de prévision de séries temporelles développé par Facebook. Il est conçu pour traiter des séries temporelles ayant une forte tendance et une saisonnalité, ce qui est souvent le cas dans les séries temporelles financières.

L'architecture de FB Prophet repose sur trois composants principaux : la tendance, la saisonnalité et les vacances. Ces composants sont combinés de manière additive ou multiplicative pour obtenir la prévision finale.

FB Prophet a été choisi pour ce projet en raison de sa simplicité et de son efficacité pour traiter les séries temporelles avec une forte saisonnalité et une tendance, comme c'est le cas avec nos données. Il a également l'avantage d'être robuste aux valeurs manquantes et aux points aberrants, ce qui est utile pour les données financières.

## Justification de préférence des LSTM aux GRU :

Les GRU (Gated Recurrent Units) sont une autre variante des RNN qui, tout comme les LSTM, ont été conçus pour résoudre le problème des dépendances à long terme. Ils utilisent deux "portes" (réinitialisation et mise à jour) pour contrôler le flux d'informations, ce qui leur permet de conserver les informations sur de longues périodes.

Cependant, en dépit de ces caractéristiques attrayantes, il y a plusieurs raisons pour lesquelles les GRU n'ont pas été retenus pour ce projet :

- **Performance** : Bien que les GRU aient une architecture plus simple que les LSTM et soient donc plus rapides à entraîner, plusieurs études n'ont pas montré de différence significative en termes de performance entre les GRU et les LSTM. Dans certains cas, les LSTM ont même démontré de meilleures performances, surtout sur des séquences de données plus longues.
- **Adaptabilité aux données** : Les LSTM ont montré une plus grande robustesse face à une variété de types de données et de tâches. Cela signifie qu'ils sont plus susceptibles de bien performer sur notre jeu de données, même s'il présente des caractéristiques inattendues.
- **Prévisibilité** : Les LSTM sont plus couramment utilisés et plus largement étudiés que les GRU. Cela signifie qu'ils sont plus prévisibles en termes de comportement, ce qui peut être un avantage lorsque l'on travaille avec des données financières, où les erreurs peuvent avoir des conséquences importantes.

Pour ces raisons, les LSTM ont été préférés aux GRU pour ce projet.

## Raisons de ne pas retenir les Transformers :

Les Transformers sont des modèles d'apprentissage profond puissants, souvent utilisés dans le traitement du langage naturel. Cependant, il y a plusieurs raisons pour lesquelles les Transformers n'ont pas été retenus pour ce projet de prévision de séries temporelles :

- **Complexité d'entraînement** : Les Transformers sont des modèles assez complexes qui nécessitent une quantité importante de données pour être entraînés efficacement. Dans le contexte de notre projet, avec un ensemble de données de 275 instances, l'entraînement d'un Transformer pourrait conduire à des problèmes comme le surapprentissage ou des performances médiocres.
- **Manque d'interprétabilité** : Les Transformers, en tant que modèles de boîte noire, ne fournissent pas beaucoup d'insights sur comment ils prennent leurs décisions. Cela peut être problématique, surtout dans les domaines où l'explicabilité est importante.
- **Spécificité aux séries temporelles** : Les Transformers ont été initialement conçus pour le traitement du langage naturel, et bien qu'ils aient été adaptés pour les tâches de prévision de séries temporelles, ils ne sont pas spécifiquement conçus pour cela. D'autres modèles comme les RNN, LSTM, et FB Prophet, qui sont plus adaptés pour traiter des séries temporelles, sont susceptibles de fournir de meilleurs résultats.
- **Ressources computationnelles** : L'entraînement de modèles de Transformers peut nécessiter des ressources computationnelles conséquentes, ce qui pourrait ne pas être idéal étant donné les contraintes de ressources potentielles de ce projet.

C'est pour ces raisons que les Transformers n'ont pas été sélectionnés pour ce projet de prévision de séries temporelles.

### 5.3.3 Feature engineering

Le *Feature Engineering*, ou ingénierie des caractéristiques, est le processus qui consiste à utiliser des connaissances spécifiques du domaine pour créer des caractéristiques (descripteurs) qui rendent les algorithmes d'apprentissage automatique plus performants. C'est essentiellement l'art de transformer les données brutes en caractéristiques utiles qui aident les algorithmes à mieux comprendre les schémas sous-jacents des données.

Ce processus peut impliquer plusieurs étapes, dont :

- **Sélection de caractéristiques** : Il s'agit d'identifier les caractéristiques les plus utiles pour la prédiction.
- **Création de caractéristiques** : Il s'agit de construire de nouvelles caractéristiques à partir des caractéristiques existantes.
- **Transformation de caractéristiques** : Les transformations, comme la normalisation ou la mise à l'échelle, peuvent être appliquées pour rendre les caractéristiques plus facilement utilisables par certains algorithmes.
- **Élimination de caractéristiques** : Cela implique de retirer les caractéristiques qui ne contribuent pas à la performance du modèle ou qui pourraient causer des problèmes, comme le surapprentissage.

Le *Feature Engineering* peut avoir un impact significatif sur la qualité des modèles d'apprentissage automatique. En fait, dans de nombreux cas, la qualité des caractéristiques peut être plus importante que le choix de l'algorithme d'apprentissage automatique lui-même.

#### Lag features :

Les "Lag Features" ou caractéristiques retardées sont une méthode couramment utilisée en ingénierie des caractéristiques (feature engineering) pour les séries temporelles.

Dans une série temporelle, une observation à un instant donné est souvent liée à des observations précédentes. Un "lag" est le délai entre deux points de données dans une série temporelle. Par exemple, dans une série temporelle quotidienne, un lag de 1 serait le décalage d'un jour.

Une "lag feature" serait donc une nouvelle caractéristique créée en décalant les valeurs de la série temporelle de quelques périodes. Ils peuvent aider le modèle à comprendre des tendances ou des motifs périodiques dans les données. Il est important de noter que la création de "lag features" peut introduire des valeurs manquantes dans les données pour les premières observations (où il n'y a pas d'observations précédentes pour créer le lag). Ces valeurs manquantes devront être traitées d'une manière ou d'une autre avant de pouvoir entraîner un modèle sur les données. Dans notre cas, on les a supprimé.

Suite au processus d'analyse des données et à la sélection des modèles à utiliser, voici les étapes finales accomplies avant de procéder à l'entraînement des modèles :

1. **Sélection de Données** : En ce qui concerne les données utilisées dans l'entraînement des modèles d'apprentissage automatique, nous avons opté pour les séries temporelles basées sur les soldes, en excluant celles des mouvements. Conformément à ce qui a été mentionné dans l'*Introduction* du chapitre *Prévisionnel*, notre focus lors de la phase de modélisation a été exclusivement sur les soldes. Cette décision est justifiée par la volatilité relativement faible des soldes comparée aux mouvements, facilitant ainsi la tâche de prédiction.

2. **Préparation des données d'entraînement** : Les modèles sélectionnés peuvent nécessiter des architectures différentes pour la phase d'entraînement :
- **Prophet** : Ce modèle nécessite que les colonnes soient nommées 'ds' pour la date et 'y' pour la valeur. Par conséquent, nous avons adapté nos données en conséquence. Il est à noter que pour ce modèle, nous n'avons pas eu à effectuer de division Train-Test. De plus, Prophet n'utilise pas de "lag features".
  - **RNN et LSTM** : Pour ces modèles, nous avons divisé nos séries temporelles en données d'entraînement (80%) et données de test (20%) en utilisant la technique du Train-Test-Split. Avant cette division, nous avons créé six (6) "lag features" pour chacune de nos séries temporelles.

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6
dco						
2022-04-03	7.733842e+09	7.698074e+09	7.365771e+09	7.294598e+09	8.105036e+09	7.297353e+09
2022-04-10	7.447208e+09	7.733842e+09	7.698074e+09	7.365771e+09	7.294598e+09	8.105036e+09
2022-04-17	7.371334e+09	7.447208e+09	7.733842e+09	7.698074e+09	7.365771e+09	7.294598e+09
2022-04-24	7.248765e+09	7.371334e+09	7.447208e+09	7.733842e+09	7.698074e+09	7.365771e+09
2022-05-01	7.290031e+09	7.248765e+09	7.371334e+09	7.447208e+09	7.733842e+09	7.698074e+09

FIGURE 5.29 – Tête du DataFrame des données d'entraînement de la caisse

- **ARIMA** : En ce qui concerne la préparation finale des données pour ARIMA, avant de créer les "lag features" de la série **Trésor**, nous avons effectué une différenciation de la série temporelle tant qu'on a remarqué qu'elle n'était pas stationnaire

La différenciation est une technique couramment utilisée en analyse de séries temporelles, notamment en préparation des données pour des modèles ARIMA. Son but principale est de rendre une série temporelle stationnaire. Une série temporelle est dite stationnaire si ses propriétés statistiques, comme la moyenne et la variance, restent constantes au fil du temps. C'est une hypothèse importante pour de nombreux modèles de séries temporelles, y compris ARIMA.

La différenciation d'une série temporelle est simplement la différence entre chaque valeur et la valeur précédente. Mathématiquement, cela peut être exprimé comme suit :

$$Y_{\text{diff}}[t] = Y[t] - Y[t - 1] \quad (5.2)$$

où  $Y$  est la série temporelle et  $t$  est le temps.

### 5.3.4 Entraînement des modèles et optimisation de l'Apprentissage Automatique

Le tableau ci-dessous est organisé en plusieurs colonnes afin de comparer facilement les différentes approches et de mieux comprendre les caractéristiques de chaque modèle appliqué sur les différentes séries temporelles.

La colonne **Modèle** indique le nom du modèle utilisé, tandis que la colonne **Série temporelle** précise la série temporelle spécifique à laquelle le modèle a été appliqué.

La colonne **Entraînement** est subdivisée en deux sous-colonnes : **Compilation** et **Adaptation**, qui décrivent les étapes spécifiques du processus d'entraînement du modèle.

La colonne **Architecture/Hyperparamètres** est également subdivisée en deux sous-colonnes : **Couches** et **Paramètres**, qui fournissent des détails sur l'architecture du modèle et les hyperparamètres utilisés.

Pour l'optimisation des hyperparamètres, nous avons utilisé la technique de Grid Search. Cette méthode nous a permis de parcourir systématiquement différentes combinaisons de paramètres, d'entraîner notre modèle pour chaque combinaison, et de retenir la configuration qui donnait les meilleurs résultats en termes de performances. Cette pratique a aidé à augmenter l'efficacité de nos modèles en trouvant les paramètres optimaux pour notre jeu de données.

Étant donné la taille considérable de ce tableau, il sera divisé en plusieurs parties afin de faciliter sa consultation. Cette démarche est nécessaire puisque le tableau ne peut pas être accommodé intégralement sur une seule page.

Modèle	Série temporelle	Entraînement	
		Compilation	Adaptation
ARIMA	Caisse	/	/
	Ccp		
	Trésor		
	Scellés BA		
LSTM	Caisse	Optimiseur = Adam Fonction de perte = MSE Métrique = MAE	Taille des lots de données = 1 Epochs = 150
	Ccp	Optimiseur = Adam Fonction de perte = MSE Métrique = MAE	Taille des lots de données = 1 Epochs = 150
	Trésor	Optimiseur = Adam Fonction de perte = MSE Métrique = MAE	Taille des lots de données = 1 Epochs = 100
	Scellés BA	Optimiseur = Adam Fonction de perte = MSE Métrique = MAE	Taille des lots de données = 1 Epochs = 100
RNN	Caisse	Optimiseur = Adam Fonction de perte = MSE	Taille des lots de données = 1 Epochs = 50
	Ccp	Optimiseur = Adam Fonction de perte = MSE	Taille des lots de données = 1 Epochs = 50
	Trésor	Optimiseur = Adam Fonction de perte = MSE	Taille des lots de données = 1 Epochs = 50
	Scellés BA	Optimiseur = Rmsprop Fonction de perte = MSE	Taille des lots de données = 1 Epochs = 100
FB-Prophet	Caisse	Optimiseur : L-BFGS	/
	Ccp		
	Trésor		
	Scellés BA		

FIGURE 5.30 – Modélisation - Compilation et Adaptation

Modèle	Architecture/Hyperparamètres	
	Couches	Paramètres
ARIMA	/	/
		/
		p=1, i=1, q=3 p=5, i=0, q=3
LSTM	LSTM (1ère couche) Dropout	128 unités, return_sequences=True, input_shape=(X_train.shape[1], 1) 0.2
	LSTM (2ème couche) Dense	64 unités, return_sequences=False 1 unité
	EarlyStopping	monitor='loss', patience=20
	LSTM (1ère couche) Dropout	128 unités, return_sequences=True, input_shape=(X_train.shape[1], 1) 0.2
	LSTM (2ème couche) Dense	64 unités, return_sequences=False 1 unité
	EarlyStopping	monitor='loss', patience=20
	Bidirectional LSTM (1ère couche) Dropout	150, activation='relu', return_sequences=True, input_shape=(X_train.shape[1], X_train.shape[2])0.2
	Bidirectional LSTM (2ème couche) Dense	75 unités, return_sequences=False 1 unité
EarlyStopping	monitor='loss', patience=20	
RNN	Bidirectional LSTM (1ère couche) Dropout	128, activation='relu', return_sequences=True, input_shape=(X_train.shape[1], X_train.shape[2])0.2
	Bidirectional LSTM (2ème couche) Dense	64 unités, return_sequences=False 1 unité
	EarlyStopping	monitor='loss', patience=20
	SimpleRNN Dense	100 unités, activation='tanh', input_shape=[None, 1] 1 Densité
	SimpleRNN Dense	100 unités, activation='tanh', input_shape=[None, 1] 1 Densité
	SimpleRNN Dense	150 unités, activation='tanh', input_shape=[None, 1] 1 Densité
	SimpleRNN Dense	100 unités, activation='tanh', input_shape=[None, 1] 1 Densité
	SimpleRNN Dense	100 unités, activation='tanh', input_shape=[None, 1] 1 Densité
FB-Prophet	/	/

FIGURE 5.31 – Modélisation - Optimisation

## 5.4 Résultats et Evaluation

### 5.4.1 Résultats

Compte tenu des données de test qui s'étendent du 3 avril 2022 au 3 avril 2023, les prévisions pour la caisse ont été générées par les modèles RNN, LSTM et Prophet comme suit :



RNN :

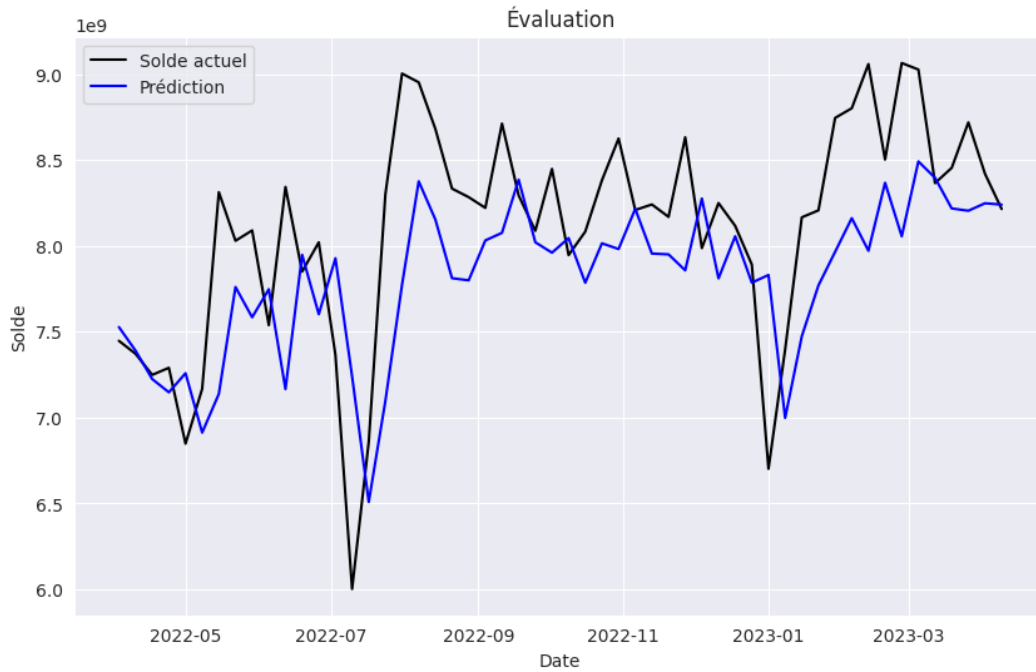


FIGURE 5.32 – Prévisions du modèle RNN pour les soldes hebdomadaires de la caisse

LSTM :

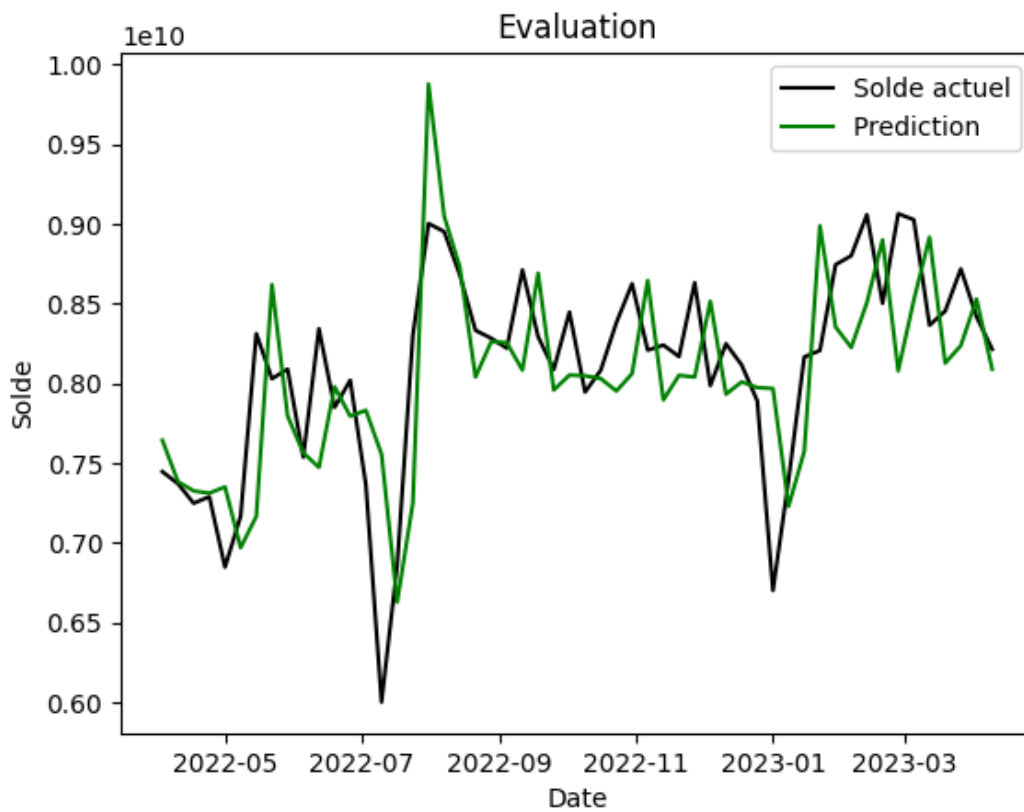


FIGURE 5.33 – Prévisions du modèle LSTM pour les soldes hebdomadaires de la caisse

## Prophet :

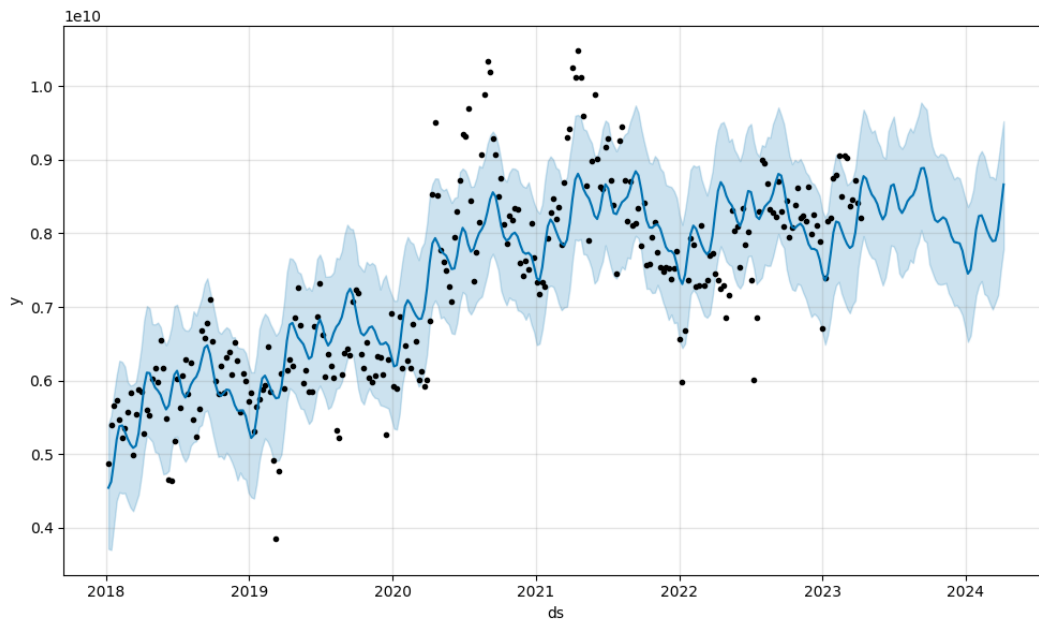


FIGURE 5.34 – Prévisions du modèle Prophet pour les soldes hebdomadaires de la caisse

### 5.4.2 Métriques

- **RMSE (Root Mean Square Error)** : Le RMSE est utilisé pour calculer la racine carrée de la moyenne des erreurs quadratiques. C'est une mesure populaire pour vérifier l'exactitude des prédictions d'un modèle.

La formule du RMSE est :

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

où  $N$  est le nombre total d'observations,  $y_i$  est la valeur observée et  $\hat{y}_i$  est la valeur prédite.

- **MSE (Mean Squared Error)** : Le MSE est la moyenne des carrés des erreurs. C'est une autre mesure populaire pour vérifier l'exactitude des prédictions d'un modèle.

La formule du MSE est :

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

où  $N$  est le nombre total d'observations,  $y_i$  est la valeur observée et  $\hat{y}_i$  est la valeur prédite.

- **MAE (Mean Absolute Error)** : Le MAE est la moyenne des valeurs absolues des erreurs. C'est une mesure robuste aux valeurs aberrantes, car elle ne met pas au carré les erreurs.

La formule du MAE est :

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

où  $N$  est le nombre total d'observations,  $y_i$  est la valeur observée et  $\hat{y}_i$  est la valeur prédite.

- **MAPE (Mean Absolute Percentage Error)** : Le MAPE est la moyenne des valeurs absolues des erreurs en pourcentage. Cette mesure est souvent utilisée lorsque vous voulez exprimer l'erreur en pourcentage plutôt qu'en unités brutes.

La formule du MAPE est :

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

où  $N$  est le nombre total d'observations,  $y_i$  est la valeur observée et  $\hat{y}_i$  est la valeur prédite. Notez que le MAPE peut être problématique lorsque  $y_i$  est égal à zéro pour certaines observations, car cela entraîne une division par zéro.

### 5.4.3 Evaluation

Nous avons employé les métriques définies précédemment pour apprécier la performance de nos modèles. Les résultats recueillis sont présentés ci-après :

	Modèle	RMSE	MSE	MAE	MAPE
<b>Caisse</b>	ARIMA	/	/	/	/
	LSTM	0.07955	0.00632	0.05918	0.09800
	RNN	0.08699	0.00756	0.06802	0.05573
	FB-Prophet	0.17150	0.02941	0.14368	0.23304

TABLE 5.5 – Evaluation des modèles appliqués à la caisse

	Modèle	RMSE	MSE	MAE	MAPE
<b>CCP</b>	ARIMA	/	/	/	/
	LSTM	0.04104	0.00168	0.03158	/
	RNN	0.04778	0.00228	0.03515	/
	FB-Prophet	0.12346	0.01524	0.10771	/

TABLE 5.6 – Evaluation des modèles appliqués au CCP

	Modèle	RMSE	MSE	MAE	MAPE
<b>Trésor</b>	ARIMA	0.52544	0.23409	0.48703	0.36931
	Bi-LSTM	0.22593	0.04910	0.17843	0.17239
	RNN	0.21727	0.04720	0.16628	0.14601
	FB-Prophet	0.34098	0.09978	0.29054	0.21555

TABLE 5.7 – Evaluation des modèles appliqués au Trésor

	<b>Modèle</b>	<b>RMSE</b>	<b>MSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Scellés BA</b>	ARIMA	0.45098	0.21345	0.42205	0.36791
	Bi-LSTM	0.23487	0.04910	0.18782	0.05692
	RNN	0.23183	0.07911	0.22007	0.17546
	FB-Prophet	0.43565	0.09890	0.38769	0.19246

TABLE 5.8 – Evaluation des modèles appliqués aux scellés BA

#### 5.4.4 Conclusion

Suivant les résultats de l'évaluation avec les différentes métriques et afin de passer à la prochaine étape de test de résistance, il a été décidé de retenir pour chacun des postes les modèles suivants :

- **Caisse** : LSTM
- **CCP** : LSTM
- **Trésor** : RNN
- **SBA** : Bi-LSTM

# Chapitre 6

## Stress-test

### 6.1 Introduction

Après avoir réalisé les modèles de prévisions des différents postes, nous allons maintenant passer à la partie stress. Nous allons stresser les comptes Caisse et Trésor.

En ce qui concerne :

- Les scellés BA, leurs mouvements dépendent majoritairement de l'efficacité des agents responsables de leur comptage à la BA. Nous ne sommes pas en mesure de formuler des scénarios de stress pour ces derniers.
- Le compte CCP, la seule part d'incertitude des cashflows de ce poste réside dans les défauts de paiements des chances de crédits de particuliers CCP car ce compte est alimenté par les créances des clients fonctionnaires publics. À cet effet, ce poste est donc déterministe et fonction du portefeuille de crédits consommations et travaux octroyés aux clients non domiciliés.

Cependant, la partie prévisionnel des scellés BA et du CCP restera utile lorsqu'on établit le prévisionnel de la Trésorerie.

Nous allons présenter les différents scénarios de stress pour chaque poste :

#### 6.1.1 Scénario de stress pour la Caisse

La SGA dispose de 103 agences à travers le territoire national. Chacune d'elles dispose d'une caisse principale et d'une caisse secondaire. Nos données représentent les soldes de toutes les caisses de la SGA y compris les Business Center (Agences pour les Grandes Entreprises) et les agences pour les particuliers. Le solde de ces caisses dépend des dépôts et des retraits effectués en cash au niveau de ces agences. L'hypothèse de stress la plus probable et défavorable est : retraits massifs et simultanés.

Voici un exemple de scénario de stress qu'on pourrait faire subir à la caisse :

**Crise de confiance dans le système financier de la SGA**

⇓

**Retraits massifs et simultanés des dépôts**

⇓

**Baisse du solde de la caisse de 60%**

1. **Crise de confiance dans le système financier de la SGA** : Cette première étape pourrait être déclenchée par une série d'événements tels qu'une mauvaise gestion financière, des rumeurs de difficultés financières, des fraudes ou des scandales

au sein de la Société Générale Algérie. Cela provoquerait une perte de confiance de la part des clients et des investisseurs envers la banque.

2. **Retraits massifs et simultanés des dépôts :** En raison de la perte de confiance dans la banque, les clients commenceraient à retirer massivement leurs dépôts. Cela pourrait créer un effet domino car la peur se propagerait parmi les clients, provoquant une panique bancaire.
3. **Baisse du solde de la caisse de 60% :** À la suite des retraits massifs, la banque verrait son solde de caisse chuter de manière significative, par exemple de 60%. Cela signifierait que la Société Générale Algérie aurait maintenant moins de liquidités à sa disposition pour répondre aux demandes des clients, pour ses opérations quotidiennes ou pour investir. Si la situation persistait, cela pourrait potentiellement mettre en péril la solvabilité de la banque, exacerbant encore plus la crise de confiance.

**Formulation mathématique du scénario :** Dans cette partie, le but est de définir une fonction qui définit la série de prévisionnel stressé à partir de la série de prévisionnel BAU.

Imaginons une fonction de stress qui augmente graduellement puis diminue, simulant un choc qui atteint un pic puis s'atténue. Cette dernière prend donc en compte le temps et l'ampleur du stress.

Une telle fonction pourrait être modélisée en utilisant une fonction Gaussienne inversée, qui atteint un maximum à un certain point et diminue de chaque côté de ce point.

La fonction de stress  $S(t)$  est définie comme suit :

$$S(t) = A \cdot \exp\left(-\frac{(t - B)^2}{2C^2}\right) \quad (6.1)$$

où :

- $A$  est l'amplitude du stress (dans notre cas, 0.6 si on veut modéliser une baisse de 60%),
- $B$  est le moment où le stress atteint son maximum (dans notre cas, le moment où les retraits massifs atteignent leur maximum),
- $C$  contrôle la largeur de la cloche de la Gaussienne (c'est-à-dire en combien de temps le stress s'accumule et se dissipe).

Ensuite, la fonction  $g_1(t)$  qui applique le stress à la série temporelle BAU ( $f_1(t)$ ) serait :

$$g_1(t) = f_1(t) \cdot (1 - S(t)) \quad (6.2)$$

Tel que :

- $f_1(t)$  : Fonction de la série temporelle prédite du solde de la caisse en BAU.
- $g_1(t)$  : Fonction de la série temporelle prédite et stressée du solde de la caisse

Dans ce modèle, le stress commence à augmenter à partir de  $t = 0$ , atteint un maximum à  $t = b$ , puis diminue graduellement. L'effet de  $S(t)$  sur  $f_1(t)$  est de réduire la valeur de  $f_1(t)$  à mesure que le stress augmente.

## Cohérence de la fonction de stress avec le scénario :

1. **Augmentation graduelle** : Au début de la crise de confiance, la panique parmi les déposants n'est pas encore généralisée. Seuls quelques-uns, informés ou particulièrement sensibles aux signaux de crise, commencent à retirer leurs fonds. Cela correspond à la phase initiale d'augmentation graduelle de la fonction de stress.
2. **Atteinte d'un pic** : Au fur et à mesure que la crise de confiance s'intensifie, de plus en plus de clients retirent leurs dépôts, augmentant ainsi le stress sur la banque. Cela culmine en un pic lorsque la panique est à son comble, et le maximum de clients retirent leurs fonds. Ce pic correspond au maximum de la fonction de stress.
3. **Diminution graduelle** : Après le pic de la crise, des mesures sont prises pour restaurer la confiance, ou les clients qui voulaient retirer leurs fonds l'ont déjà fait. Le nombre de retraits commence alors à diminuer, tout comme la fonction de stress.

Dans ce contexte, une fonction Gaussienne inversée est appropriée car elle modélise une situation où un choc (la crise de confiance) provoque une réponse (les retraits de dépôts) qui augmente rapidement, atteint un sommet, puis diminue progressivement. C'est une bonne représentation de la façon dont une panique bancaire se déroule typiquement.

Dans ce qui suit, nous ferons référence à ce scénario en tant que *Scénario 1*.

### 6.1.2 Scénario de stress pour le Trésor public

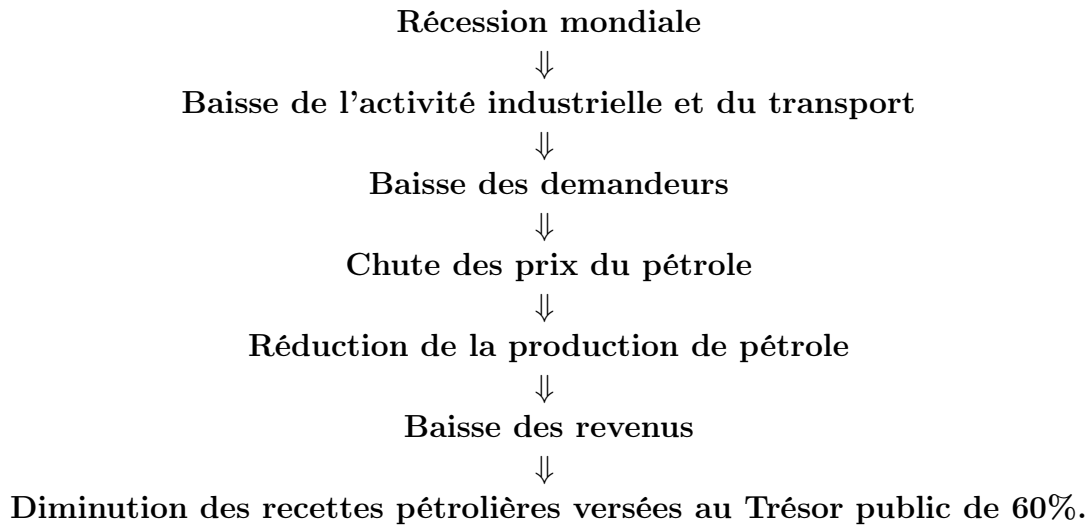
La SGA dispose de comptes au sein du Trésor public qui est alimenté par les recettes générées des activités commerciales des clients SGA opérant dans le marché public.

Lorsqu'un entrepreneur conclut un contrat avec une entité publique, il s'engage à fournir des biens ou des services en échange d'une rémunération. Cette rémunération est versée par l'entité publique au Trésor public, qui agit en tant que dépositaire des fonds publics. Le Trésor public, quant à lui, est alimenté par différentes sources de revenus et recettes publiques. Principalement, les impôts et taxes, les cotisations sociales et revenus des entreprises publiques.

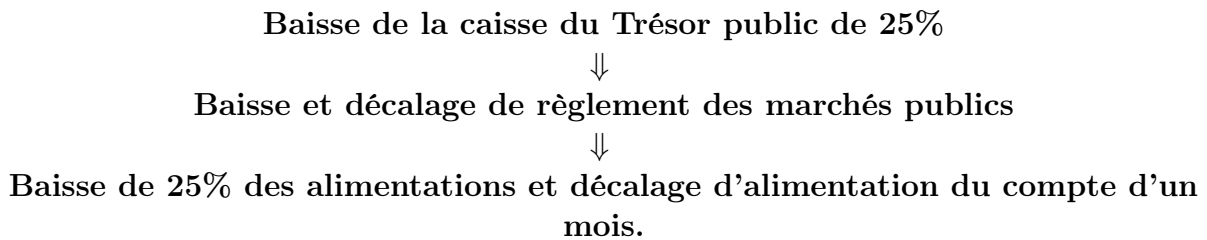
En Algérie, le pétrole est la pierre angulaire de l'économie du pays. Les revenus générés par la fiscalité pétrolière, provenant des impôts, des droits et des redevances liés aux activités pétrolières, sont versés au Trésor public. Le Trésor public utilise une partie de ces revenus pour le financement du budget d'investissement de l'état.

En 2016, le Trésor public en Algérie a connu une baisse d'alimentation en raison de la chute des prix du pétrole. Cette baisse significative des revenus pétroliers a impacté les recettes fiscales de l'État et a eu des répercussions sur le financement des dépenses publiques, l'Algérie étant dépendante du pétrole.

Le scénario de stress choisi est donc :



Étant donné qu'en 2022, la fiscalité pétrolière représente 41% des recettes budgétaires totales :



**Formulation mathématique du scénario :**

Nous pouvons résumer le scénario comme suit :

$$g_2(t) = f_2(t + t') * C \tag{6.3}$$

Tel que :

- $f_2(t)$  : Fonction de la série nominale.
- $g_2(t)$  : Fonction de la série stressée.

Avec :

- $C$  : Coefficient de baisse du compte Trésor.
- $t'$  : Décalage en semaine de la série stressée.

Dans notre cas, nous prenons :

- Baisse de 25% :  $C=0.75$
- Décalage de 4 semaines :  $t=4$

Dans ce qui suit, nous ferons référence à ce scénario en tant que *Scénario 2*.



## 6.2 Résultats du stress

### 6.2.1 Résultat du *scénario 1* sur la caisse

En appliquant le *scénario 1* de stress relatif à la caisse, nous obtenons le graphique ci-dessous :

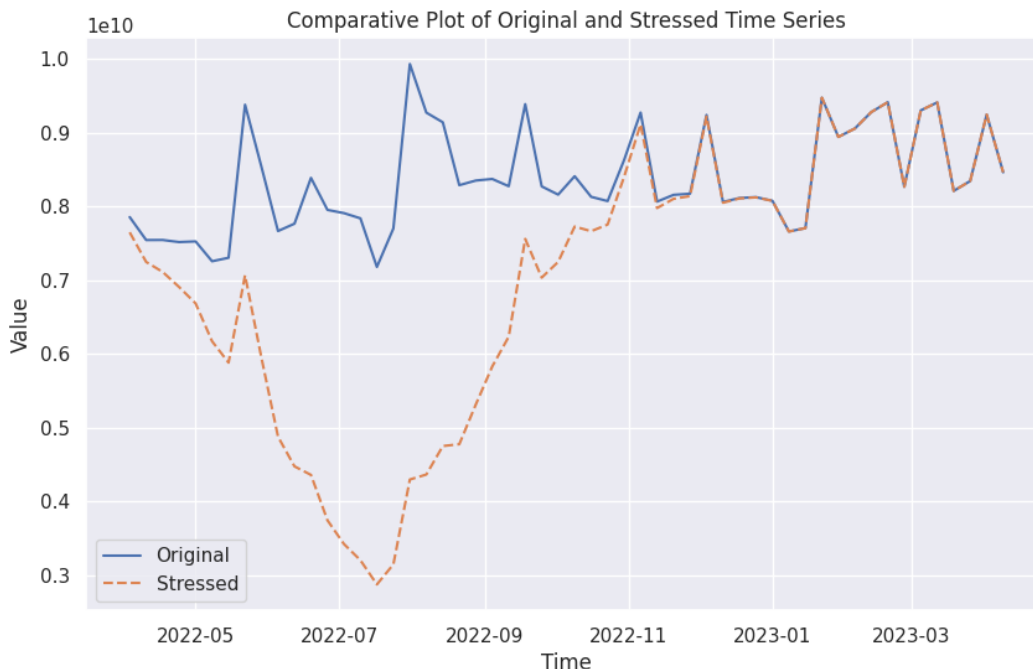


FIGURE 6.1 – Le prévisionnel BAU et le prévisionnel stressé de la caisse

Nous pouvons voir dans la figure [6.1](#) :

Dans notre scénario, le solde de caisse diminuerait de 60%, donc  $A = 0.6$ .

Si on suppose que le stress atteint son maximum 15 semaines après le début de la crise,  $B$  pourrait être défini comme 15.

Si on suppose que le stress s'accumule sur une période de 6 semaines avant d'atteindre son maximum, puis se dissipe sur une période de 6 semaines après avoir atteint son maximum,  $C$  pourrait être défini comme 6.

### 6.2.2 Résultat du *scénario 2* sur le compte Trésor

En appliquant le *scénario 2* de stress relatif au Trésor, nous obtenons le graphique ci-dessous :

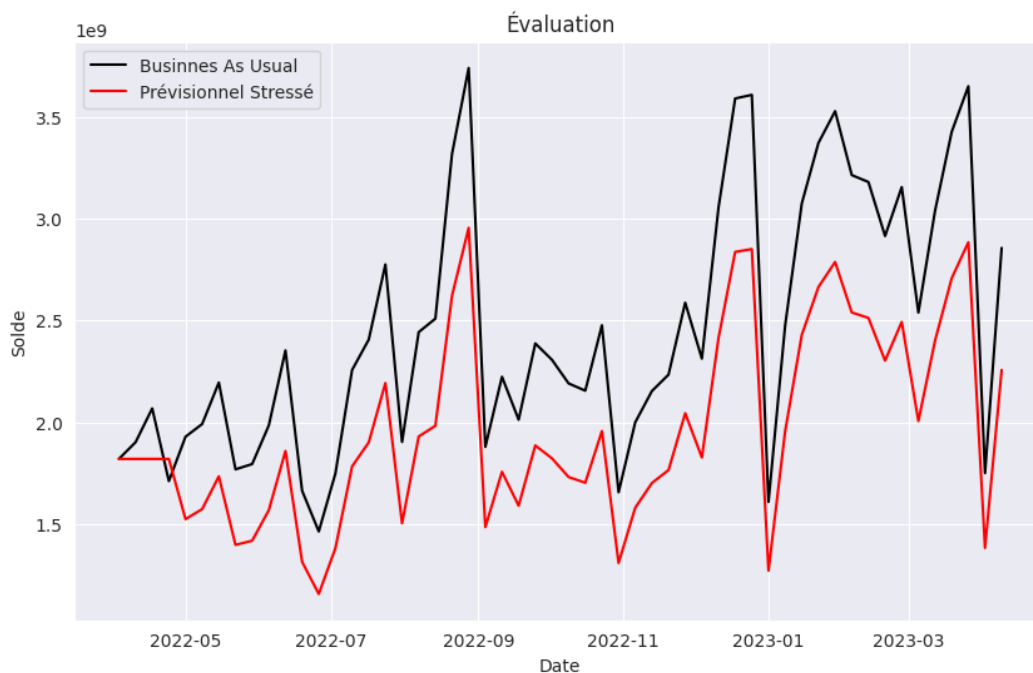


FIGURE 6.2 – Le prévisionnel BAU et le prévisionnel stressé du Trésor

Nous pouvons voir dans la figure [6.2](#) un décalage d'alimentation de **4 semaines** et une baisse du solde de **25%**.

### 6.3 Conclusion

Suite à notre analyse détaillée des scénarios de stress, nous avons pu observer de manière significative l'impact de ces derniers sur les actifs de trésorerie de la SGA, en particulier sur le trésor et la caisse. Dans chacun de ces scénarios de stress (considérés séparément et conjointement), la SGA devrait prendre en compte et estimer le solde BA et les coefficients de liquidité afin de déterminer si elle se trouve dans une situation de déficit de liquidité.

Lorsque le coefficient de liquidité est inférieur à 100%, cela signifie que la SGA fait face à un manque de ressources liquides pour faire face à ses échéances financières. Dans de telles circonstances, la SGA doit trouver des moyens de recapitalisation en ayant recours aux méthodes de refinancement mentionnées dans la partie [Etat de l'art](#).

# Chapitre 7

## Application utilisateur

### Introduction

Dans le but de permettre aux utilisateurs, qu'ils soient familiers avec la technologie ou non, de bénéficier pleinement de nos modèles, nous avons développé une application utilisateur qui simplifie leur utilisation. L'interface de cette application a été créée en utilisant le framework Streamlit. Dans cette partie de notre mémoire, nous allons explorer les fonctionnalités de notre application et présenter en détail l'outil de stress-test final. L'objectif principal de cette application est d'offrir une expérience utilisateur optimale, en mettant l'accent sur la convivialité et l'accessibilité.

### 7.1 Présentation de l'outil

Dans notre projet, nous avons élaboré un outil final de stress test en utilisant le framework Streamlit, déployé sur l'environnement de développement intégré PyCharm. Streamlit est une bibliothèque Python open-source qui facilite la création d'applications web interactives, spécialement conçues pour les projets en science des données et apprentissage automatique.

Ce framework simplifie considérablement le processus de conception et de partage d'applications centrées sur les données. Il offre une interface intuitive et un ensemble de widgets intégrés pour concevoir des éléments interactifs. Grâce à Streamlit, il est possible de créer des applications web directement à partir de scripts Python.

Streamlit propose une API simple qui permet de créer des composants interactifs, comme des curseurs, des cases à cocher et des menus déroulants. Ces composants peuvent être utilisés pour contrôler le comportement de l'application et visualiser les données. L'un des avantages majeurs de Streamlit est sa capacité à mettre à jour automatiquement l'application en temps réel lorsque le code est modifié. Cette caractéristique est particulièrement utile pour itérer rapidement et explorer nos données de manière efficace.

L'utilisation de PyCharm pour déployer Streamlit a apporté une expérience de développement fluide et productive. PyCharm est reconnu pour ses fonctionnalités robustes telles que l'autocomplétion intelligente, le contrôle qualité du code, et la facilité de débogage, ce qui a contribué à augmenter notre efficacité tout au long du processus de développement.

## Outil de prévisionnel et stress-test de la trésorerie

Sélectionnez un modèle et spécifiez une plage de prévisions.

Date de fin de prévision

2023/06/15

Choisissez un poste à prédire

Trésor

Choisissez le type du prévisionnel

Prévisionnel Business As Usual

Prédire

FIGURE 7.1 – Structure de l'outil de prévision

## 7.2 Fonctionnalités

Nous allons maintenant présenter les différentes fonctionnalités de notre application :

1. **Choix de la date de fin de prévision** :L'utilisateur devra renseigner le champ de date de fin de prévision, indiquant jusqu'à quel moment les prédictions seront générées. Il est important de noter que la date de début correspond à la dernière date disponible dans notre jeu de données initial, puisque notre modèle fonctionne avec des données séquentielles.

Ainsi, en sélectionnant la date de fin, l'application utilisera notre modèle pour effectuer des prédictions à partir de la dernière date disponible jusqu'à la date de fin spécifiée. Cette approche permet de prévoir les tendances futures en se basant sur les informations historiques jusqu'à la dernière observation connue. La figure [7.2](#) montre se que la fonctionnalité " Date de fin de prévision" génère.

## Outil de prévisionnel et stress-test de la trésorerie

Sélectionnez un modèle et spécifiez une plage de prévisions.

Date de fin de prévision

2023/06/15

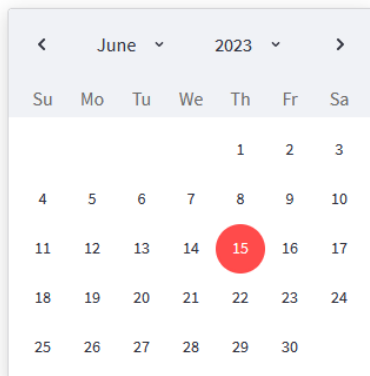


FIGURE 7.2 – Choix de la date de prévision

2. **Choix du poste à prédire** : Étant donné que nous avons à faire à quatre postes, nous donnons la possibilité à l'utilisateur de choisir quel poste il veut prédire parmi les quatre disponibles.

## Outil de prévisionnel et stress-test de la trésorerie

Sélectionnez un modèle et spécifiez une plage de prévisions.

Date de fin de prévision

2023/06/15

Choisissez un poste à prédire

Trésor

Trésor

Caisse

CCP

Scellés BA

FIGURE 7.3 – Choix du type de poste à prédire

3. **Choix du type de prévisionnel** : L'utilisateur aura le choix entre le prévisionnel nominal dit "Business As Usual" ou le prévisionnel sous scénario de stress.

# Outil de prévisionnel et stress-test de la trésorerie

Sélectionnez un modèle et spécifiez une plage de prévisions.

Date de fin de prévision

2023/06/15

Choisissez un poste à prédire

Trésor

Choisissez le type du prévisionnel

Prévisionnel Business As Usual

Prévisionnel Business As Usual

Prévisionnel sous scénario de stress

FIGURE 7.4 – Choix du type de prévision

4. **Génération des prévisions** Une fois que l'utilisateur aura terminé de sélectionner la date de fin, le poste à prédire et le type de modèle à appliquer, il devra cliquer sur le bouton "Prédire" afin de générer les prévisions. Les prévisions seront présentées comme suit :

Prédire

Les prédictions générées sont les suivantes:

	Date	Prédiction
0	2023-04-09 00:00:00	1,931,553,792
1	2023-04-16 00:00:00	503,399,584
2	2023-04-23 00:00:00	732,344,512
3	2023-04-30 00:00:00	1,122,805,504
4	2023-05-07 00:00:00	732,344,512
5	2023-05-14 00:00:00	732,344,512
6	2023-05-21 00:00:00	1,263,715,584
7	2023-05-28 00:00:00	1,600,649,856
8	2023-06-04 00:00:00	1,657,804,672
9	2023-06-11 00:00:00	1,553,433,472

FIGURE 7.5 – Prédiction générées

Les prédictions générées seront aussi visibles dans un graphique :



FIGURE 7.6 – Graphique des prédictions

# Conclusion générale



En Algérie ou dans plusieurs pays, les banques sont confrontées à des défis et des incertitudes à cause des fluctuations économiques et des changements réglementaires. Les modifications réglementaires peuvent introduire de nouvelles règles, des restrictions ou des exigences supplémentaires pour les institutions financières, ce qui peut créer une certaine incertitude pour les acteurs du secteur bancaire. Ces incertitudes peuvent entraîner une baisse de la confiance des déposants dans le système bancaire, ce qui peut se traduire par des retraits massifs de dépôts comme dans le cas de la Silicon Valley Bank dernièrement car ses clients craignaient que la banque ne perde trop d'argent en raison de la dépréciation de son portefeuille d'actions car la Réserve Fédérale a resserré la politique monétaire.

Notre mémoire avait pour but d'anticiper des scénarios catastrophiques comme celui de la banque américaine en développant un outil de stress-test et en tirant parti des méthodes d'apprentissage automatique. Pour ce faire, nous avons commencé par explorer l'ALM de la SGA et plus particulièrement le risque de liquidité, étape indispensable pour le Business Understanding. Ensuite, nous avons établi un diagnostic complet en utilisant les méthodes d'analyse SWOT et PESTEL afin d'identifier les dysfonctionnements liés à l'implémentation d'un stress-test au sein de la SGA. Nous ne sommes focalisés surtout sur l'aspect prévisionnel et les méthodes utilisées par la SGA.

Une étude bibliographique a été réalisée pour examiner l'utilisation de l'IA et des techniques d'apprentissage automatique dans le contexte des stress-tests. Cette étude visait à établir un lien concret entre les aspects théoriques et leur application concrète dans le domaine des stress-tests.

Pour la partie pratique, nous avons commencé par développer des modèles de prévisions en s'appuyant sur la méthodologie CRISP-DM qui ont servi de base pour notre outil de stress. Il était important d'appliquer nos hypothèses sur un prévisionnel fiable. Nous avons testé plusieurs modèles sur les postes de la SGA qu'on devait prédire, pour ensuite, choisir le modèle le plus optimal possible. Par la suite, nous avons défini nos scénarios finaux après en avoir testé plusieurs et nous en avons gardé les plus probables, que nous avons appliqués à nos prévisions. Pour faciliter l'utilisation du stress-test, nous avons conçu une interface utilisateur conviviale et facile d'utilisation destinée à l'ALM.

Dans une perspective d'amélioration continue de l'outil, il est essentiel pour l'ALM de ne pas se contenter d'un seul scénario, mais plutôt d'envisager l'ajout de différents scénarios. Cette approche offre plusieurs avantages significatifs pour l'analyse et la gestion des risques liés à la liquidité. En permettant l'inclusion de plusieurs scénarios, l'ALM peut obtenir une vision plus complète et nuancée des impacts potentiels sur la liquidité. Chaque scénario peut représenter une situation spécifique, telle qu'une crise économique, une variation importante des taux d'intérêt, une dégradation de la qualité des actifs, ou toute autre circonstance susceptible d'affecter les flux de trésorerie et la disponibilité des liquidités. En explorant un éventail de scénarios, l'ALM peut évaluer de manière plus approfondie les variations possibles et les risques associés. Cela permet de mieux comprendre les facteurs clés qui influencent la liquidité et d'identifier les mesures nécessaires pour faire face à ces situations.

# Bibliographie

- [1] M. Levasseur, *Gestion de trésorerie*. Economica, 1979.
- [2] S. SAOULI and H. ZAID, “Étude analytique d’un stress test de liquidité durant la crise covid-19 «cas des banques publiques algériennes» analytical study of a liquidity stress testing during the covid-19 crisis «case of algerian public banks»,” *Revue Recherches et études en Développement*, vol. 9, no. 2, pp. 415–430, 2022.
- [3] N. Venard, “Gestion des risques bancaires et réglementation prudentielle,” *Revue d’économie financière*, no. 28, pp. 49–62, 1994.
- [4] B. Keizer, “La gestion des risques dans les banques,” *Revue d’économie financière*, no. 27, pp. 345–364, 1993.
- [5] F. Ait Amara and B. Boutalbi, “Gestion de la trésorerie d’une banque ; cas de la «cneq-banque : Agence laarbi ben-mhidi202, de tizi-ouzou»,” Ph.D. dissertation, Université Mouloud Mammeri, 2020.
- [6] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, “Financial time series forecasting with deep learning : A systematic literature review : 2005–2019,” *Applied soft computing*, vol. 90, p. 106181, 2020.
- [7] M. H. Pesaran, D. Pettenuzzo, and A. Timmermann, “Forecasting time series subject to multiple structural breaks,” *The Review of Economic Studies*, vol. 73, no. 4, pp. 1057–1084, 2006.
- [8] M. Jacobs Jr, “The validation of machine-learning models for the stress testing of credit risk,” *Journal of Risk Management in Financial Institutions*, vol. 11, no. 3, pp. 218–243, 2018.
- [9] N. Malik, P. V. Singh, and U. Khan, “Can banks survive the next financial crisis ? an adversarial deep learning model for bank stress testing,” *An Adversarial Deep Learning Model for Bank Stress Testing (June 30, 2018)*, 2018.
- [10] N. Milojević and S. Redzepagic, “Prospects of artificial intelligence and machine learning application in banking risk management,” *Journal of Central Banking Theory and Practice*, vol. 10, no. 3, pp. 41–57, 2021.
- [11] G. Gao, B. Mishra, and D. Ramazzotti, “Causal data science for financial stress testing,” *Journal of computational science*, vol. 26, pp. 294–304, 2018.
- [12] J. Cao, Z. Li, and J. Li, “Financial time series forecasting model based on ceemdan and lstm,” *Physica A : Statistical mechanics and its applications*, vol. 519, pp. 127–139, 2019.
- [13] B. Krollner, B. J. Vanstone, G. R. Finnie *et al.*, “Financial time series forecasting with machine learning techniques : a survey.” in *ESANN*, 2010.
- [14] C. Goodhart, “La gestion du risque de liquidité,” *Revue de la stabilité financière*, no. 11, pp. 41–46, 2008.

- [15] M. Costisor, “Le risque de liquidité dans le système bancaire,” Ph.D. dissertation, Université Paris-Est, 2010.
- [16] F. Razzak, “Deep learning for financial banking stress test analytics,” Ph.D. dissertation, Rutgers The State University of New Jersey, Graduate School-Newark, 2020.
- [17] C. Saiktishna, N. S. V. Sumanth, M. M. S. Rao, and J. Thangakumar, “Historical analysis and time series forecasting of stock market using fb prophet,” in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2022, pp. 1846–1851.
- [18] S. Chaturvedi, E. Rajasekar, S. Natarajan, and N. McCullen, “A comparative assessment of sarima, lstm rnn and fb prophet models to forecast total and peak monthly energy demand for india,” *Energy Policy*, vol. 168, p. 113097, 2022.
- [19] Y. Romanyuk, “Asset-liability management : An overview,” 2010.
- [20] K. Kosmidou and C. Zopounidis, “Bank asset liability management techniques : an overview,” *Fuzzy Sets In Management, Economics And Marketing*, pp. 255–268, 2001.
- [21] A. BENTHAMI, “La bancarisation et la liquidité des banques : cas du maroc,” *Revue Internationale des Sciences de Gestion*, vol. 2, no. 3, 2019.

# Annexes

# Chapitre A

## Actifs et Passifs

Les principaux postes d'un bilan bancaire :

- **Crédits** : Un crédit est une mise à disposition d'argent sous forme de prêt, consentie par un créancier (prêteur) à un débiteur (emprunteur). Pour le créancier, l'opération donne naissance à une créance sur l'emprunteur, en vertu de laquelle il pourra obtenir remboursement des fonds et paiement d'une rémunération (intérêt) selon un échéancier prévu.
- **Obligations** : Les obligations sont considérées comme un investissement car lorsque la banque investit dans une obligation, elle reçoit un paiement régulier de la part de son émetteur dit taux de rendement.  
Dans le cas où la banque émet une obligation car elle a besoin de lever des fonds, ceci est considéré comme un passif car il s'agit d'une dette. Nous distinguons de catégories d'obligations : Les obligations des entreprises et les obligations d'état dit aussi "Bons du Trésor".
- **Immobilisations** : Les immobilisations sont des actifs physiques tels que des terrains, des maisons, des immeubles, des locaux commerciaux ou industriels ou encore du matériel informatique et bureautique.
- **Dépôts clients** : Montants d'argent déposés par les clients de la banque, tels que les particuliers, les entreprises et les gouvernements. Les dépôts sont considérés comme des passifs car la banque doit les rembourser à tout moment.
- **Fonds propres** : Sommes versées par les associés ou actionnaires, augmentées par les profits générés annuellement par l'entreprise qui ne sont pas distribués en dividendes.
- **Emprunts** : Il s'agit d'emprunts auprès des autres banques. Dans le cas où la banque a besoin de liquidités à court terme, elle a recours aux prêts interbancaires.

# Chapitre B

Exploration, analyse de données et  
modélisation : CCP

### Evolution du solde CCP

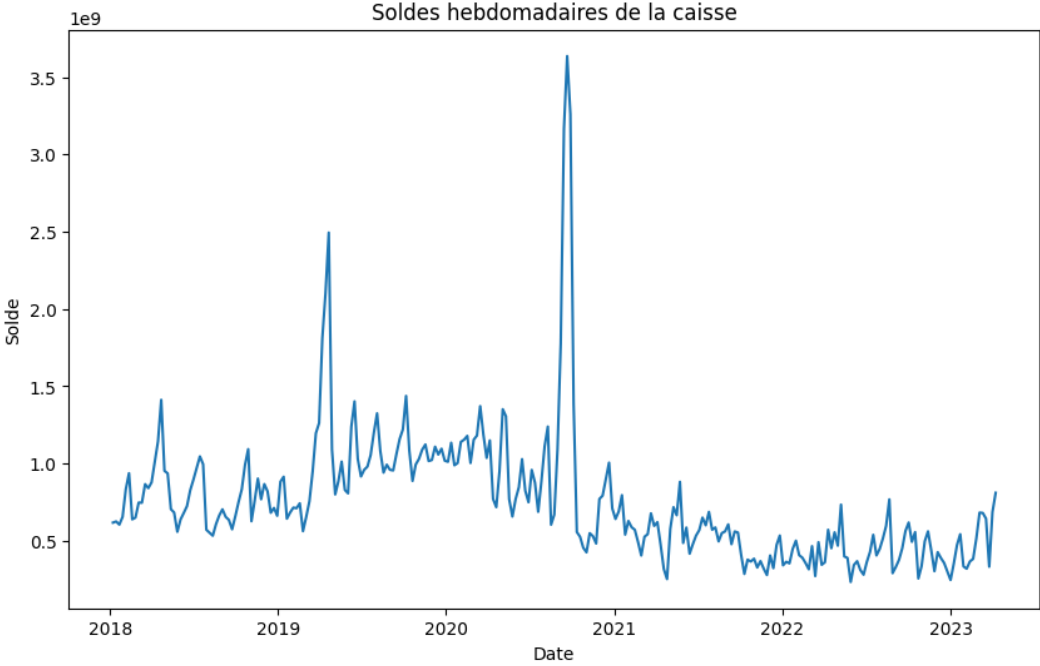


FIGURE B.1 – Evolution du solde CCP

### Histogramme du solde CCP

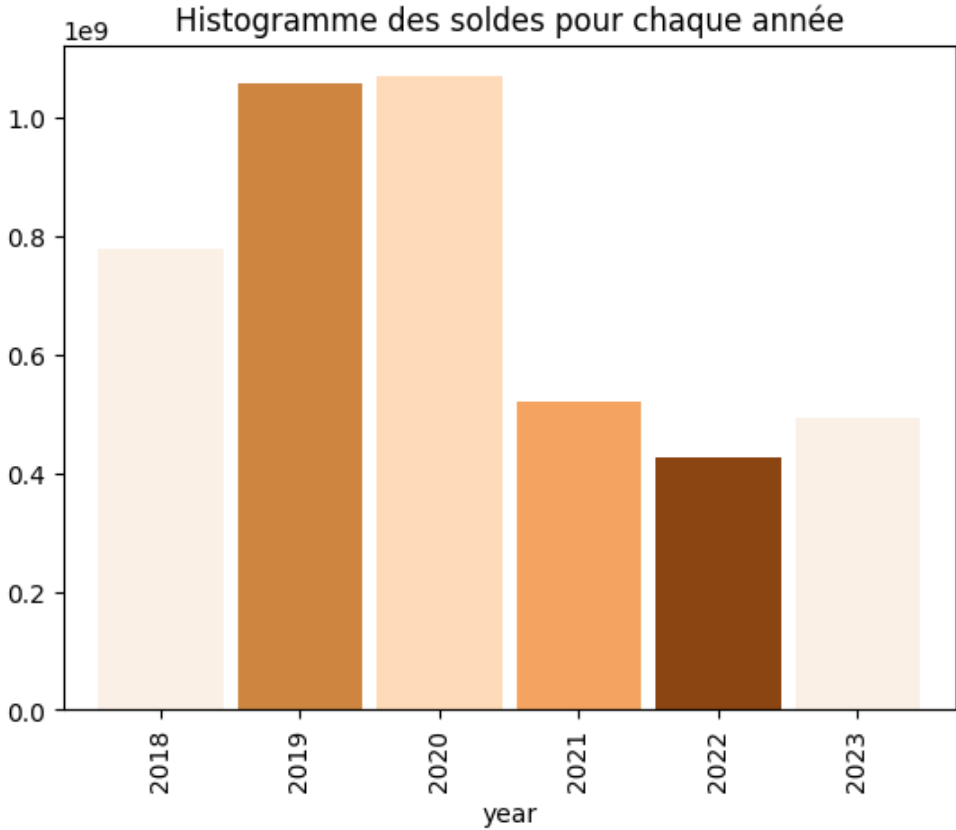


FIGURE B.2 – Histogramme du solde CCP



## Evolution du solde avec moyenne mobile et variance

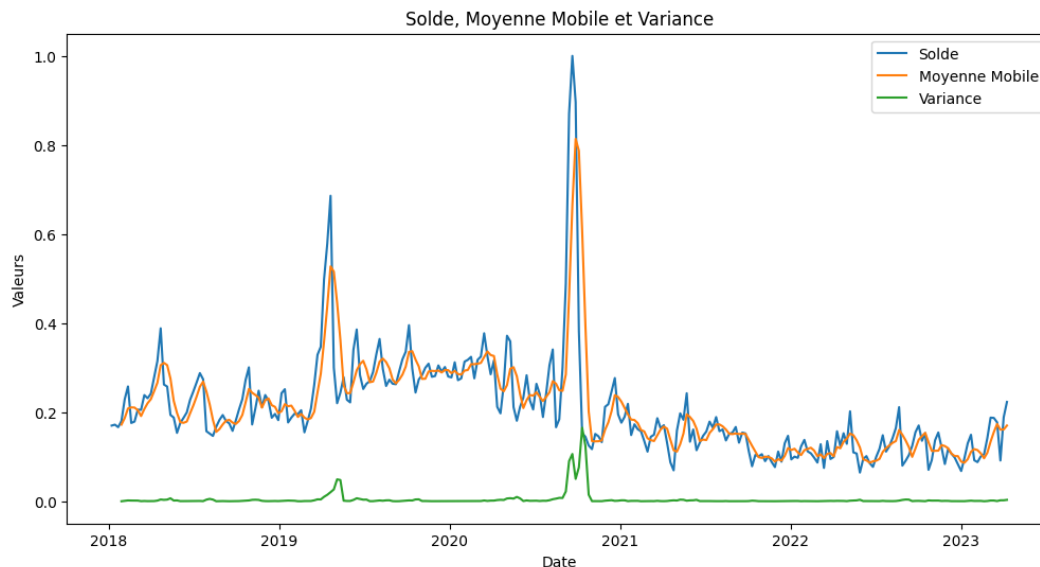


FIGURE B.3 – Evolution du solde avec moyenne mobile et variance

### Test de stationnarité

Les résultats du test ADF (Augmented Dickey-Fuller) sont les suivants :

- ADF Statistic : -4.611332031673571
- p-value : 0.00012309790212866367
- Num Lags : 3
- Num Observations Used For ADF Regression and Critical Values Calculation : 271
- Critical Values :
  - 1% : -3.4547128138328875
  - 5% : -2.8722649771800155
  - 10% : -2.5724850011573914

**Interprétation :** La statistique ADF (-4.611332031673571) est inférieure aux valeurs critiques à 1%, 5% et 10%, ce qui suggère que nous pouvons rejeter l'hypothèse nulle de non-stationnarité de la série temporelle. Le p-value (0.00012309790212866367) est également très faible, ce qui renforce la conclusion selon laquelle la série temporelle est stationnaire. Le nombre de retards utilisés dans la régression est de 3, et 271 observations ont été utilisées pour effectuer le test et calculer les valeurs critiques.

En résumé, les résultats du test ADF indiquent que la série temporelle est stationnaire, ce qui est important pour l'analyse et la modélisation des données.

## Décomposition de la série temporelle

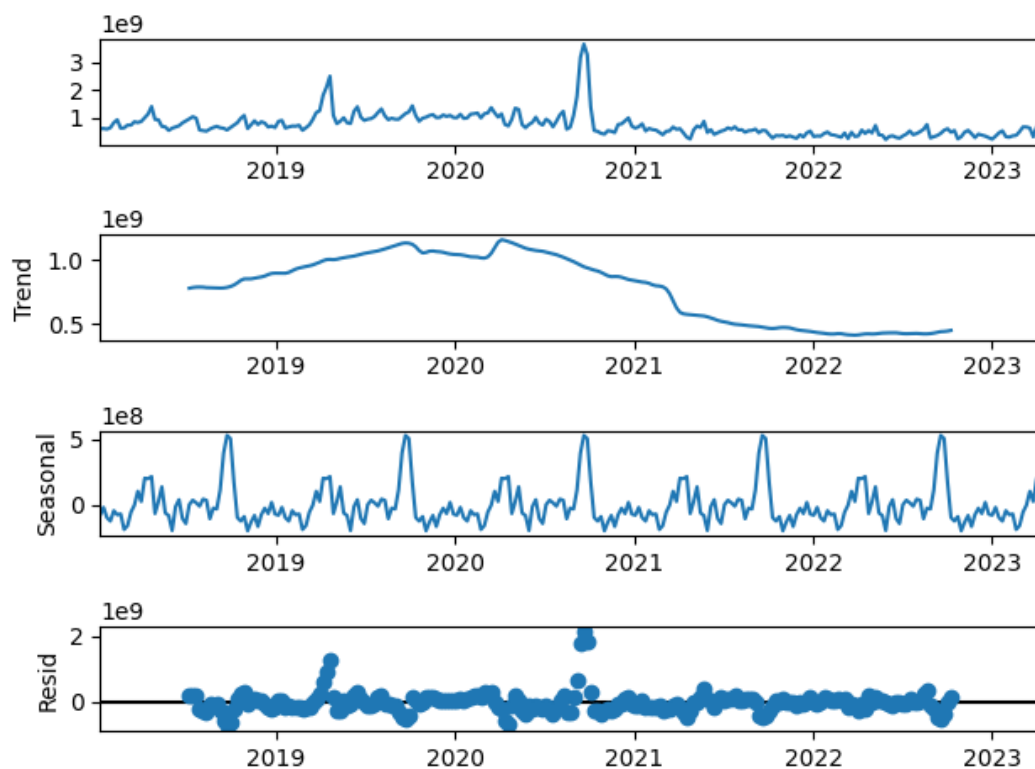


FIGURE B.4 – Décomposition de la série temporelle du solde CCP

## Autocorrélation et autocorrélation partielle

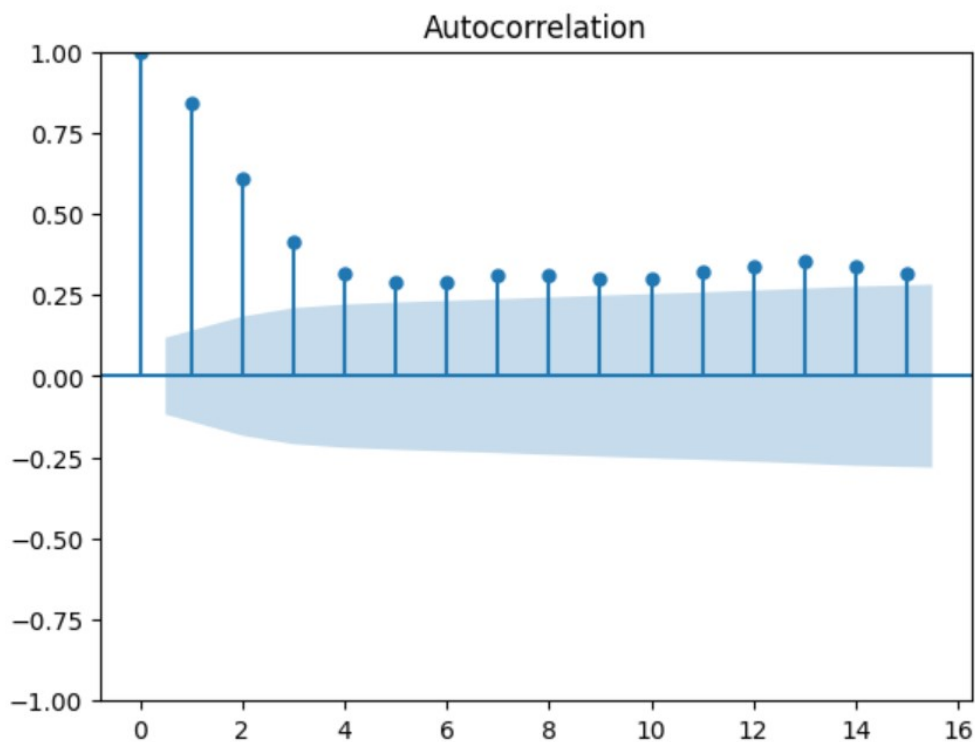


FIGURE B.5 – Autocorrélation de la série temporelle du solde CCP

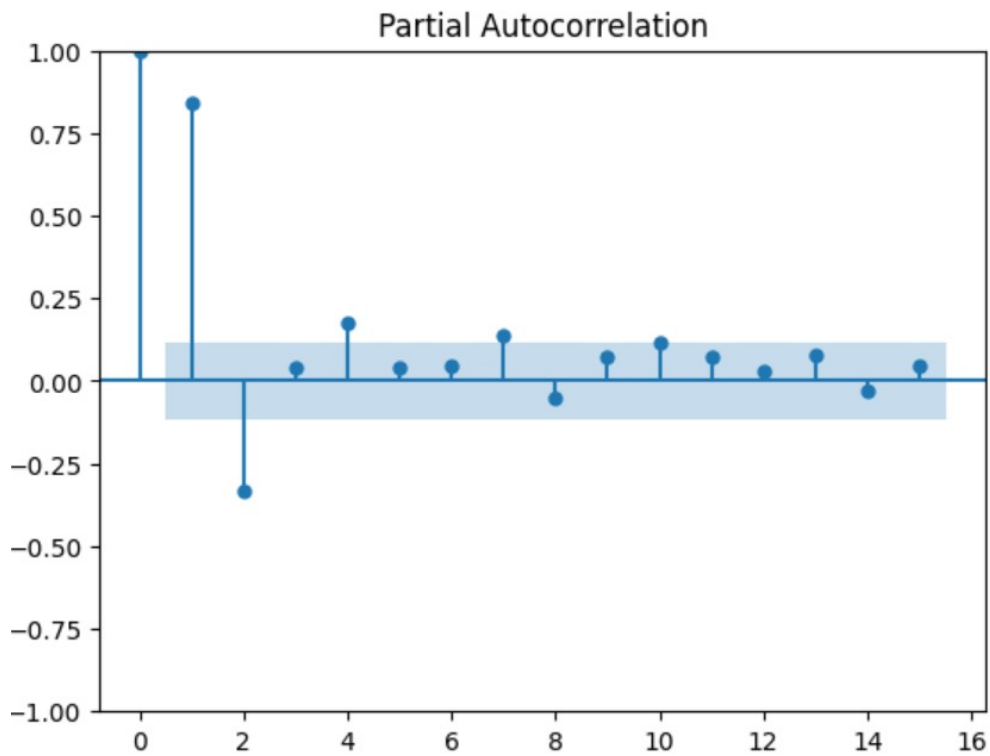


FIGURE B.6 – Autocorrélation partielle de la série temporelle du solde CCP

### Prévisions du modèle sélectionné (LSTM) sur le solde CCP

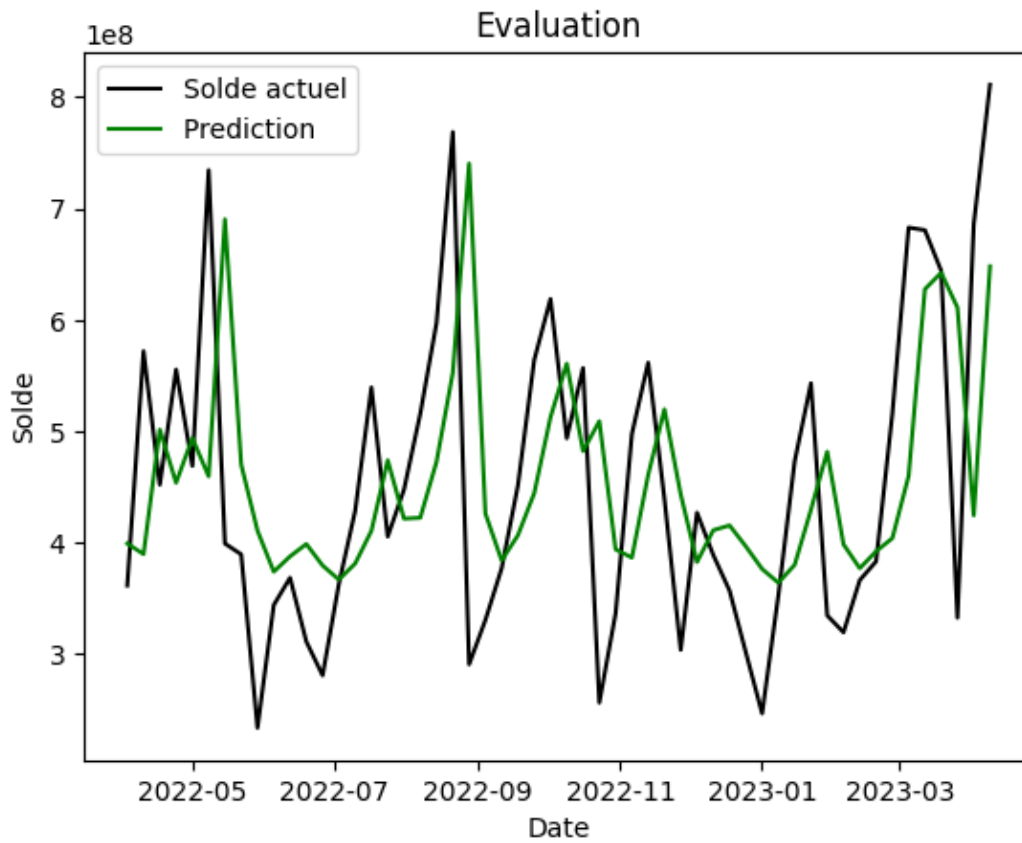


FIGURE B.7 – Prévisions (LSTM) et valeurs réelles du solde CCP

# Chapitre C

## Exploration, analyse de données et modélisation : Trésor

### Evolution de la série temporelle par année

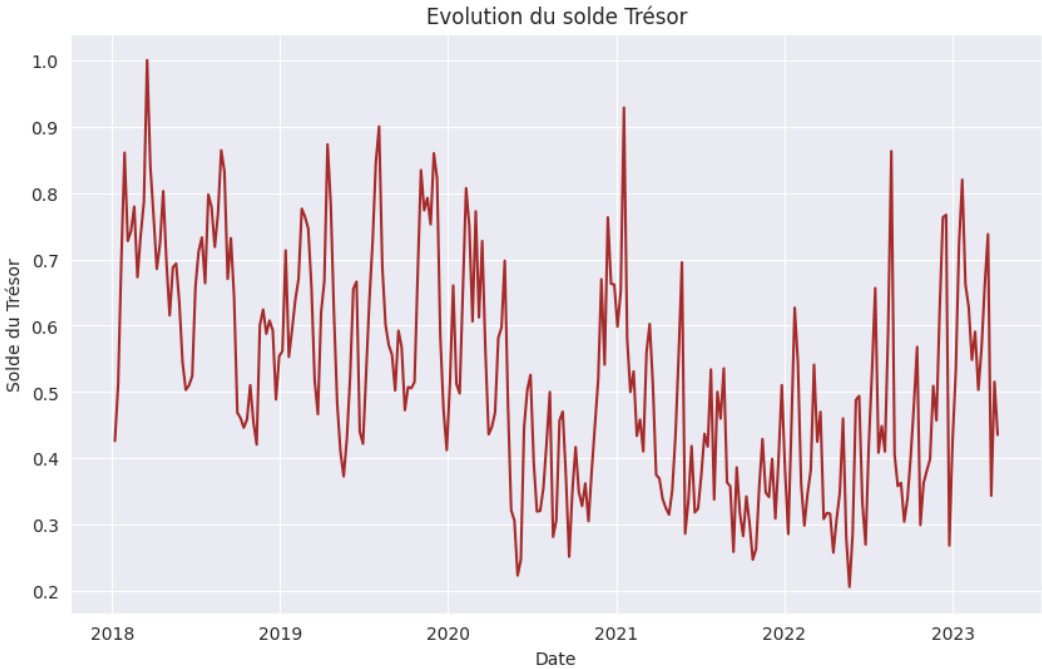


FIGURE C.1 – Évolution du compte Trésor à travers le temps

Nous pouvons voir une légère tendance à la baisse du solde depuis l’année 2018 jusqu’à maintenant. Un histogramme pourrait nous aider à y voir plus clair.

### Histogramme du solde du trésor public

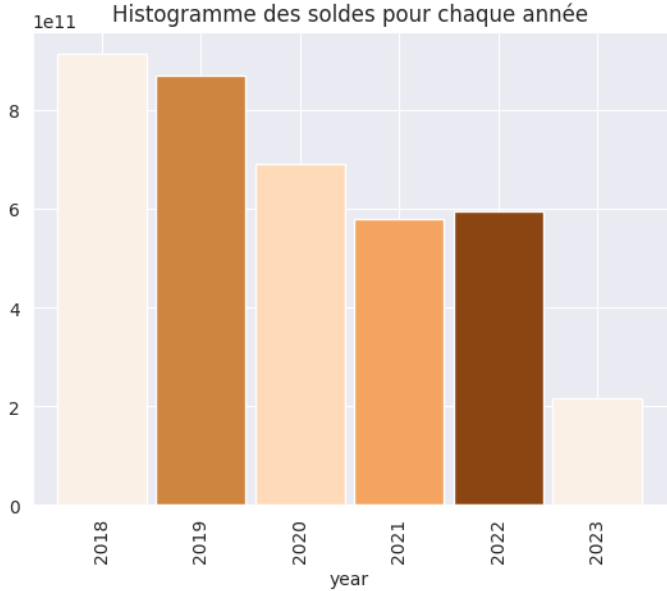


FIGURE C.2 – Histogramme annuel du Trésor

Nous pouvons clairement voir une baisse du solde, le compte Trésor public étant débité que lorsque le seuil de nivellement est atteint. Cette baisse est expliquée par : Le changement de seuil de nivellement. On voit que les seuils ont diminué à travers les années. La structure ‘Cash Management’ est responsable de cette tâche.

## Test de stationnarité

Les résultats du test ADF (Augmented Dickey-Fuller) pour le trésor sont les suivants :

- ADF Statistic : -2.70768694284632
- p-value : 0.07271751184464079
- Num Lags : 12
- Num Observations Used For ADF Regression and Critical Values Calculation : 262
- Critical Values :
  - 1% : -3.455558114028747
  - 5% : -2.872635586277424
  - 10% : -2.572682677000175

**Interprétation :** La statistique ADF (-2.70768694284632) est supérieure aux valeurs critiques à 1%, 5% et 10%, ce qui suggère que nous pouvons accepter l'hypothèse nulle de non-stationnarité de la série temporelle. Le p-value (0.07271751184464079) est également supérieure au seuil de 0.05, ce qui renforce la conclusion selon laquelle la série temporelle est stationnaire. Le nombre de retards utilisés dans la régression est de 15, et 259 observations ont été utilisées pour effectuer le test et calculer les valeurs critiques.

En résumé, les résultats du test ADF indiquent que la série temporelle n'est pas stationnaire, ce qui est important pour l'analyse et la modélisation des données.

Étant donné que l'une des hypothèses de base de ARIMA est que la série temporelle doit être stationnaire et que notre série présente une tendance, nous allons lui appliquer une différenciation d'ordre 1 afin de voir si la tendance a été éliminée.

La différenciation de premier ordre a été appliquée comme suit :

$$Y'_t = Y_t - Y_{t-1}$$

Nous présentons la décomposition saisonnière avant la différenciation où la tendance est plus visible :

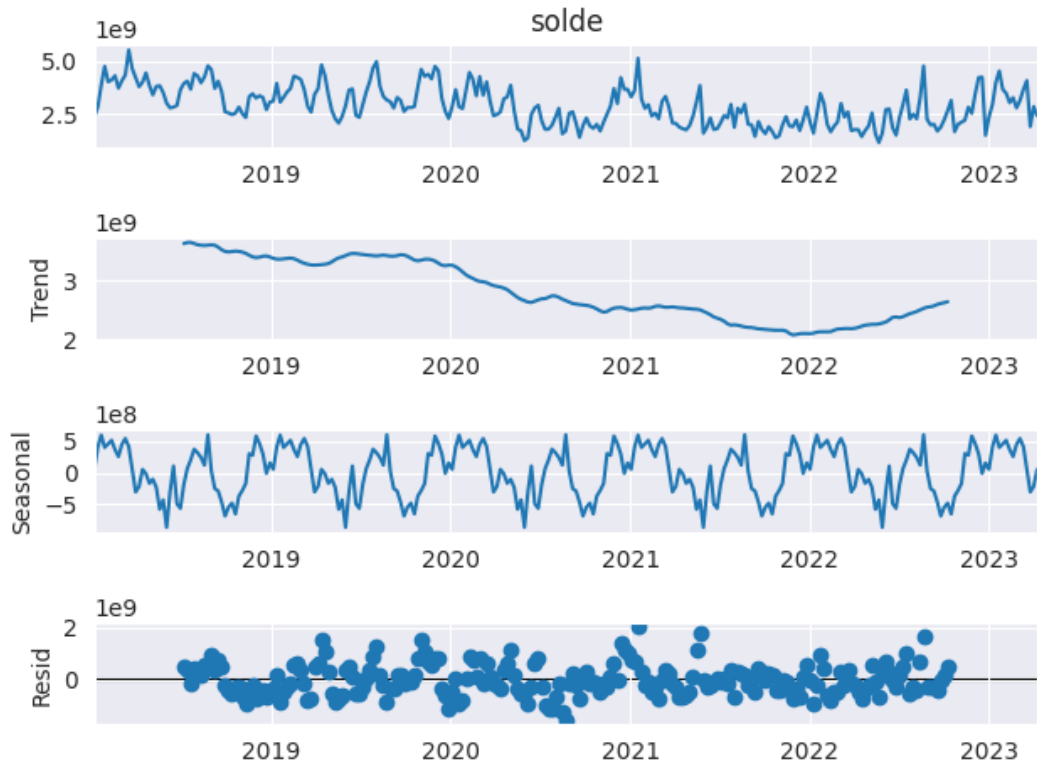


FIGURE C.3 – Décomposition de la série avant différenciation

Après avoir appliqué la différenciation, nous obtenons cette décomposition saisonnière :

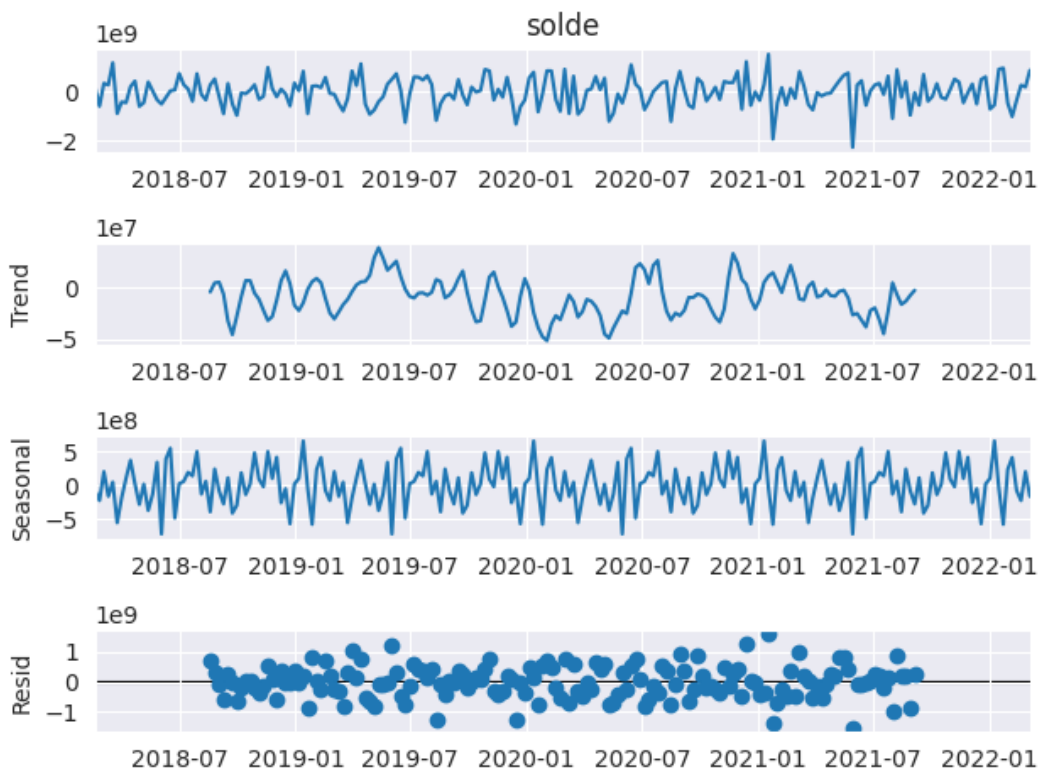


FIGURE C.4 – Décomposition de la série après différenciation

## Autocorrélation et autocorrélation partielle

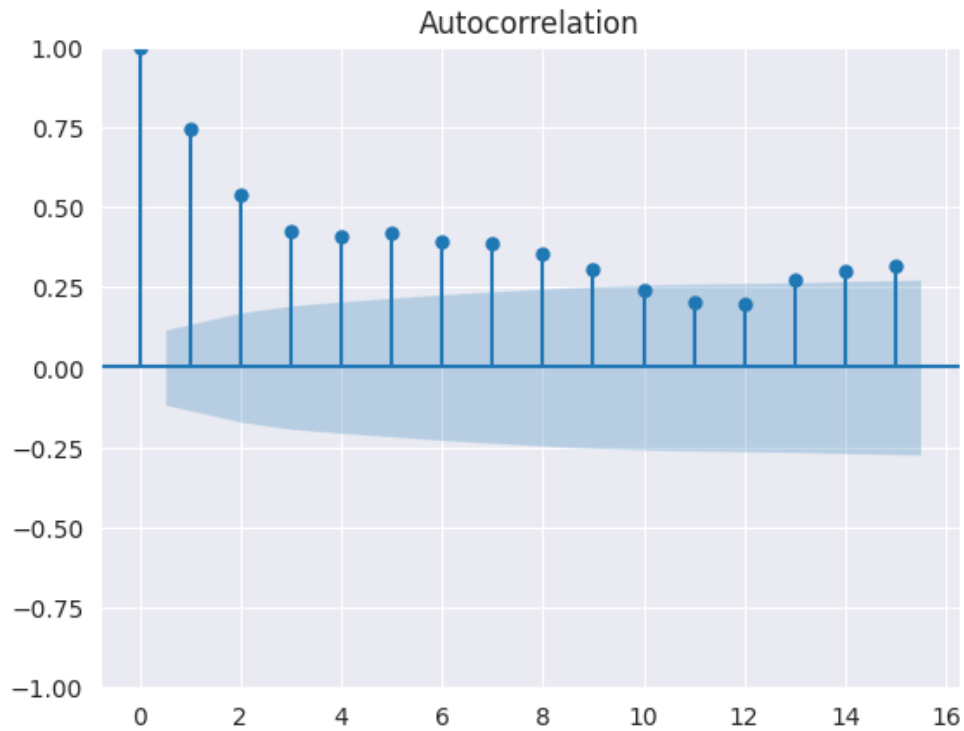


FIGURE C.5 – Autocorrélation de la série temporelle du compte Trésor

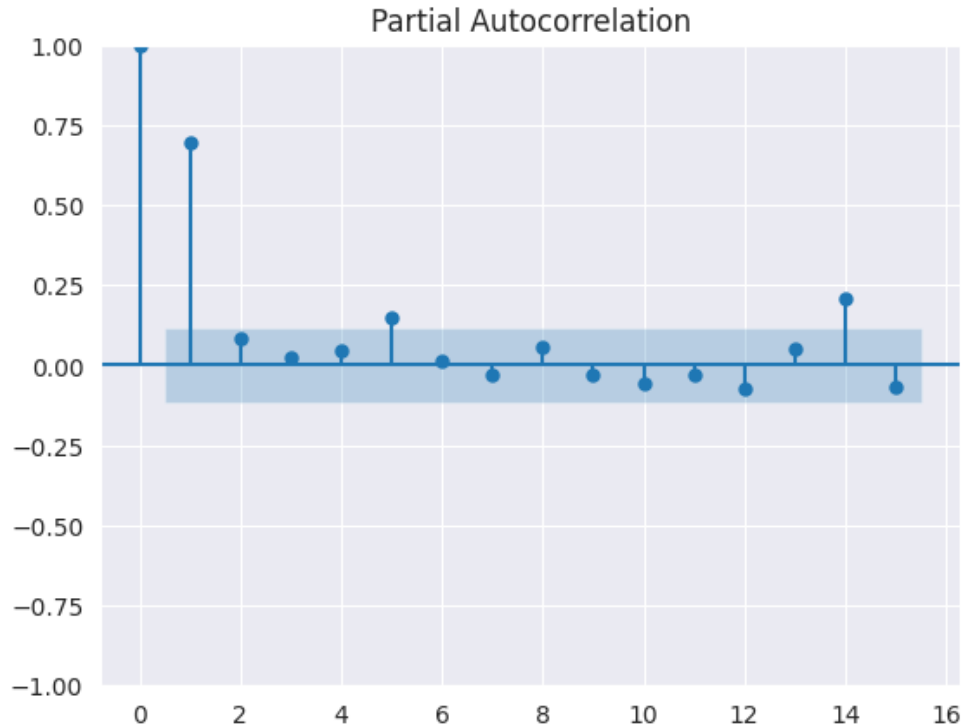


FIGURE C.6 – Autocorrélation partielle de la série temporelle du compte Trésor

**Interprétation :** La série temporelle présente une forte corrélation avec elle-même, indiquant une autocorrélation significative et une faible autocorrélation partielle.



**Résultat du modèle obtenu :**

Comme vu dans le tableau évaluation du compte Trésor 5.7 le modèle le plus optimal est les RNN. Dans la figure C.7, nous montrons la série original et la série prédite à partir du modèle RNN :

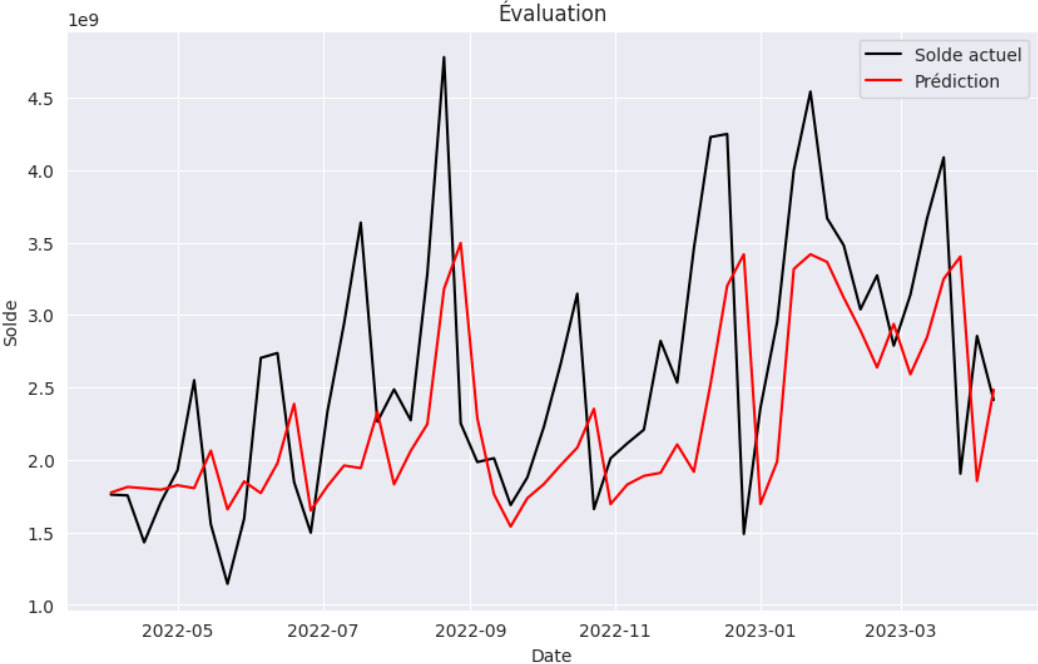


FIGURE C.7 – Série temporelle original et prédictions

# Chapitre D

## Exploration, analyse de données et modélisation : Scellés BA

Nous rappelons d'abord que le poste scellés BA comprend les sacs scellés en attente de comptage à la BA. Dans cette annexe, nous allons explorer et comprendre nos données.

## Evolution de la série temporelle par année

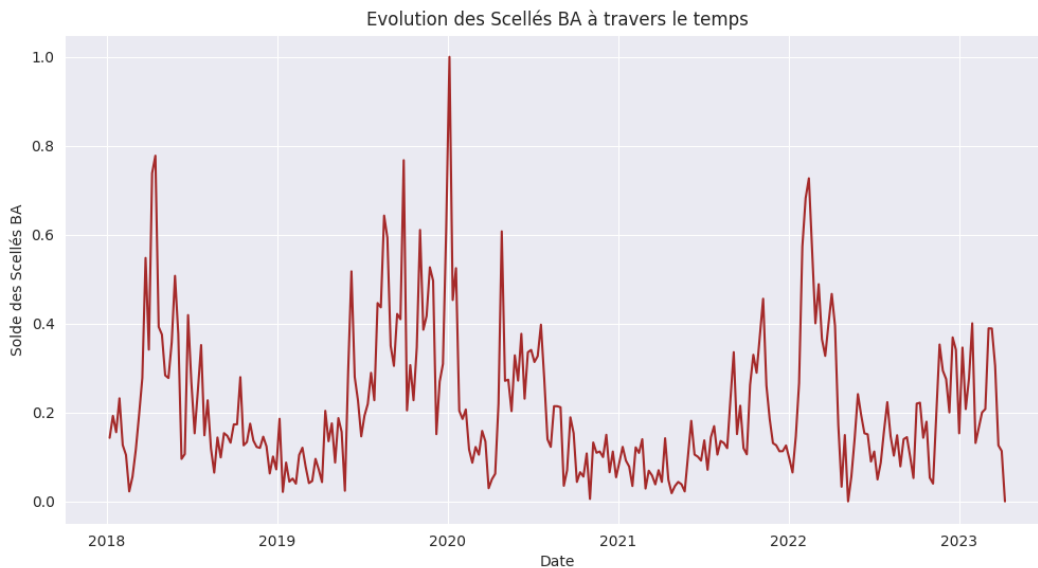


FIGURE D.1 – Évolution des Scellés BA à travers le temps

Les scellés BA dépendent des agents responsables du comptage dès la réception des sacs scellés. Nous pouvons voir un pic exceptionnel en 2020 qui est expliqué par la période de COVID où l'Algérie a mis en place des mesures de confinement et des restrictions pour prévenir la propagation du virus telles que les arrêts de travail où les réductions d'activité.

## Test de stationnarité

Les résultats du test ADF (Augmented Dickey-Fuller) sont les suivants :

- ADF Statistic : -5.508154482307871
- p-value : 2.000025210657827e-06
- Num Lags : 15
- Num Observations Used For ADF Regression and Critical Values Calculation : 259
- Critical Values :
  - 1% : -3.4558530692911504
  - 5% : -2.872764881778665
  - 10% : -2.572751643088207

**Interprétation :** La statistique ADF (-3.53983650490495) est inférieure aux valeurs critiques à 1%, 5% et 10%, ce qui suggère que nous pouvons rejeter l'hypothèse nulle de non-stationnarité de la série temporelle. Le p-value est également faible en prenant un intervalle de confiance de 95%, ce qui renforce la conclusion selon laquelle la série temporelle est stationnaire. Le nombre de retards utilisés dans la régression est de 16, et 258 observations ont été utilisées pour effectuer le test et calculer les valeurs critiques.

En résumé, les résultats du test ADF indiquent que la série temporelle est stationnaire, ce qui est important pour l'analyse et la modélisation des données. Nous pouvons confirmer la stationnarité par l'ACF et PACF :

### Autocorrélation et autocorrélation partielle

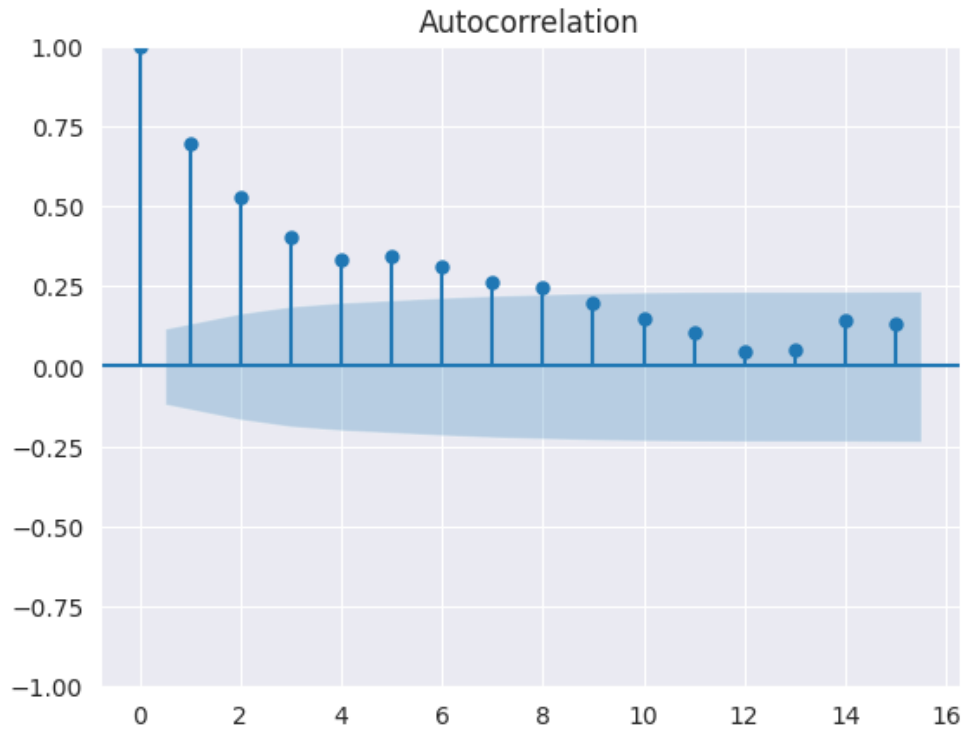


FIGURE D.2 – Autocorrélation de la série temporelle des scellés BA

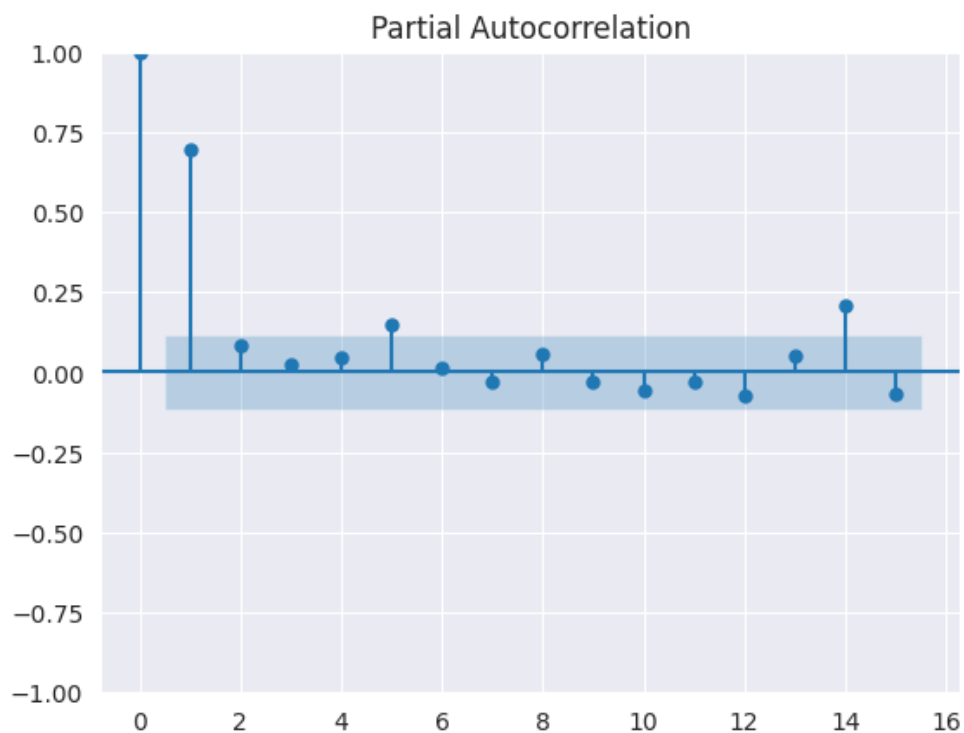


FIGURE D.3 – Autocorrélation partielle de la série temporelle des scellés BA

**Interprétation :** La série temporelle présente une forte corrélation avec elle-même, indiquant une autocorrélation significative et une faible autocorrélation partielle.

### Résultats du modèle obtenu :

Le modèle retenu pour les scellés BA est le BI-LSTM. Dans cette annexe, nous allons parler de la variante de LSTM qui est BI-directional LSTM.

Dans le modèle bidirectionnel, notre entrée circule dans les deux directions, ce qui différencie le bi-LSTM du LSTM classique. Avec le LSTM classique, nous pouvons faire circuler l'entrée dans une seule direction, soit en arrière, soit en avant. Cependant, avec le modèle bidirectionnel, nous pouvons faire circuler l'entrée dans les deux directions afin de conserver les informations du futur et du passé.

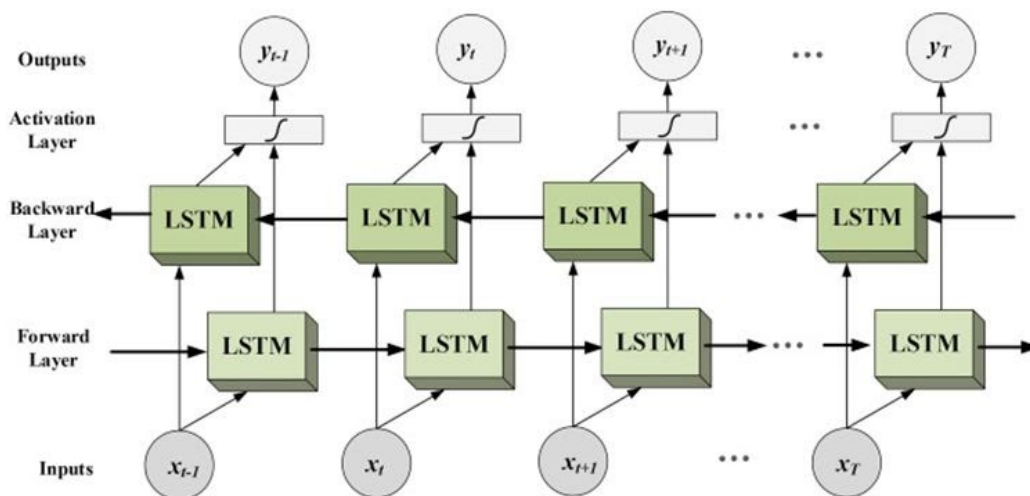


FIGURE D.4 – Architecture d'un BI-LSTM

La figure [D.5](#) suivante montre les résultats du modèle obtenu et la série temporelle originale :

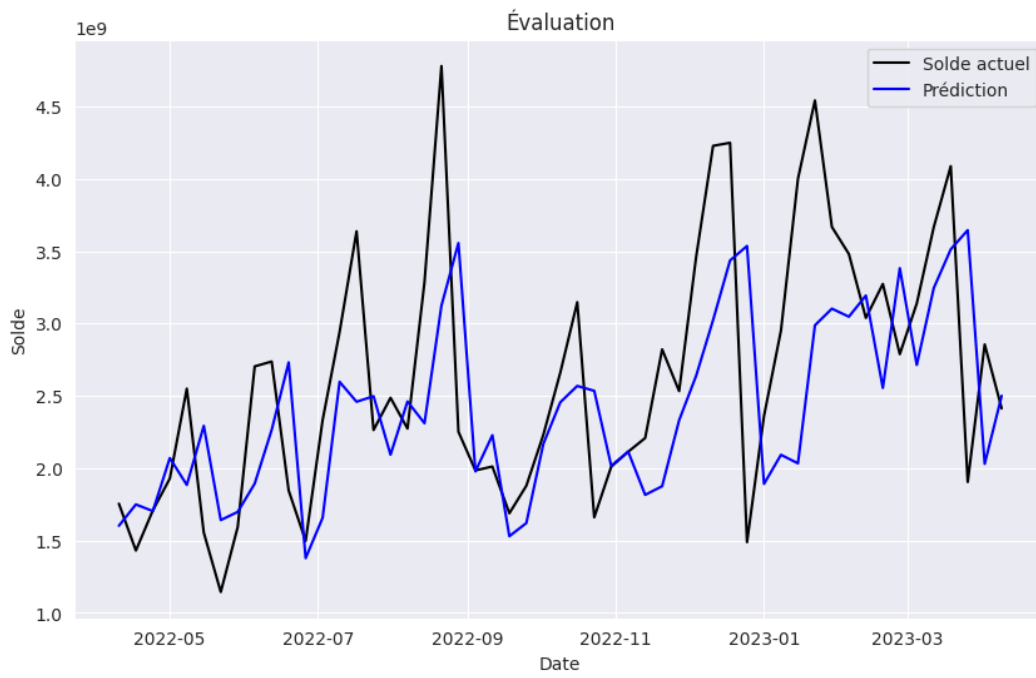


FIGURE D.5 – Série temporelle original et prédictions

# Chapitre E

## Présentation des outils utilisés

## SAP BusinessObject

Nous avons utilisé pour l'extraction de données SAP BusinessObjects. SAP BusinessObjects est une suite de logiciels d'intelligence d'affaires (Business Intelligence) proposée par SAP, une société allemande spécialisée dans les solutions logicielles pour les entreprises. SAP BusinessObjects offre une gamme complète d'outils et de fonctionnalités pour la collecte, l'analyse et la visualisation des données, permettant aux organisations de prendre des décisions éclairées basées sur des informations pertinentes et précises.



## DBeaver

Un autre outil utilisé pour l'extraction est Dbeaver. DBeaver est un outil de gestion de bases de données universel et gratuit. Il fournit une interface graphique conviviale pour se connecter, naviguer, consulter et gérer différentes bases de données, qu'il s'agisse de bases de données relationnelles (comme MySQL, PostgreSQL, Oracle, SQL Server, etc.) ou de bases de données NoSQL (comme MongoDB, Cassandra, Redis, etc.).



## Python

Étant des data scientist, notre choix s'est porté sur le langage de programmation Python. Python est un langage de programmation interprété, polyvalent et de haut niveau.

Il a été créé par Guido Van Rossum et a été publié pour la première fois en 1991. Depuis lors, Python est devenu l'un des langages de programmation les plus populaires au monde en raison de sa simplicité, de sa lisibilité et de sa flexibilité.



## Google Colab

Le travail en collaboration est important lorsqu'on travaille à deux. Il s'avère que l'un des principaux avantages de Google Colab est la collaboration en temps réel. Google Colab est un environnement de développement basé sur le cloud qui permet d'exécuter du code Python dans un navigateur web. Il est principalement utilisé pour l'apprentissage automatique, l'analyse de données et la recherche.





## PyCharm

Pour la partie implémentation de l'outil, nous nous sommes penchés vers l'IDE Python PyCharm. En plus d'une interface utilisateur intuitive et conviviale qui facilite le développement et la navigation dans le code Python, il permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web avec Django.

