

REPUBLIQUE DE CÔTE D'IVOIRE



Union – Discipline - Travail

MINISTERE DE L'ÉCONOMIE, DU PLAN ET DU DEVELOPPEMENT



ÉCOLE NATIONALE SUPERIEURE DE STATISTIQUE
ET D'ÉCONOMIE APPLIQUEE

ETUDE DE CAS

CONSTRUCTION D'UN OBSERVATOIRE DES
PRIX A LA CONSOMMATION EN COTE
D'IVOIRE PAR LES METHODES DE
MACHINE LEARNING

Réalisée par :

TCHIMTCHOUA NONO Mylene Elya

et

GANAME Abdoulaye Idrissa

Elèves Analystes Statisticiens 3^{ème} Année

Supervisée par :

Dr. COULIBALY Romaric

Directeur des Etudes

Juin 2025

Décharge

L'ENSEA n'entend donner aucune approbation, ni improbation aux opinions émises dans ce document. Ces opinions doivent être considérées comme propres à leurs auteurs.

Dédicace

Nous dédions cette étude de cas à nos familles respectives et à tous ceux et celles ayant participé de près ou de loin à notre accomplissement personnel et académique.

Remerciements

Avant de plonger dans les détails de cette expérience professionnelle, il est opportun d'entamer ce rapport par l'expression de ma gratitude envers ceux qui ont joué un rôle important durant cette période et même durant cette année de formation. Ces remerciements s'adressent à ceux qui ont contribué à rendre cette expérience enrichissante.

Nous voulons adresser nos plus sincères remerciements à notre encadrant, Dr. COULIBALY Romaric, qui nous a accompagné durant la période d'étude malgré ses lourdes charges. Son expertise, sa disponibilité, ses conseils et sa rigueur ont été déterminants dans la compréhension du sujet et de la démarche à utiliser. Nous lui sommes reconnaissants pour l'encadrement et les conseils.

Nous tenons également à remercier le Directeur Général de l'ENSEA d'Abidjan, Dr. KOUADIO Hugues pour le cadre convivial mis à disposition pour mes études et exprimer notre reconnaissance envers l'ensemble du corps enseignant de l'ENSEA, tout le personnel pour le merveilleux accueil et conseil qui m'ont permis de murir en tant qu'étudiant.

À M. HOUNDOGA Fréjus, Responsable de la division Analyste Statisticien (AS) pour ces précieux conseils, remarques et son suivi tout au long de notre formation. Nous souhaitons exprimer notre reconnaissance envers toutes les personnes qui ont consacré du temps à la lecture de ce document, même partiellement, dans le but de recueillir leurs recommandations et commentaires.

Enfin, nous ne saurions oublier nos familles, nos amis, nos proches, ainsi que toutes nos connaissances, qui de près ou de loin, nous ont apporté leurs connaissances, soutiens et motivations pour concevoir ce travail de dur labeur.

Nous leur adressons donc nos sincères remerciements pour leurs présences et conseils.

Avant-propos

L'Ecole Nationale Supérieure de Statistique et d'Economie Appliquée (ENSEA) est un établissement d'enseignement supérieur spécialisé dans la formation de statisticiens et d'économistes de haut niveau capable de répondre aux défis de ce monde en constante évolution.

La formation à l'ENSEA est subdivisée en plusieurs divisions dont celle des Analystes Statisticiens (AS). Cette formation est particulièrement marquée par une étude de cas pratique en troisième année sur des thématiques réelles. Ces études de cas constituent une occasion idéale pour les futurs statisticiens, d'appliquer les éléments d'analyse économique et statistique auxquels ils ont été formés. Cette étude de cas se présente sous la forme d'un mémoire et fait l'objet d'une évaluation par les enseignants.

Ce document représente le résultat d'un travail de recherche portant sur la conception d'un observatoire des prix à la consommation en Côte d'Ivoire à partir des méthodes de machine learning. Ce travail est d'autant plus intéressant qu'il rassemble des concepts informatiques complexes mais très prisé dans le monde professionnel et des connaissances économiques pour bien analyser l'évolution des prix, mais aussi, cela pourrait orienter les pistes de réflexion sur les questions de politique monétaire afin de maintenir la dynamique de croissance de la Côte d'Ivoire.

Ainsi, ce document présente le fruit de notre travail abouti bien que soumis à quelques difficultés. Nous avons rencontré des difficultés au niveau de la maîtrise des contours de notre thème, dans la démarche à adopter et dans l'accès aux données primaires sur lesquelles portent notre étude.

Notre satisfaction est d'être parvenus à surmonter ces difficultés et de pouvoir proposer ce modeste document.

Sommaire

Décharge	ii
Dédicace.....	iii
Remerciements.....	iv
Avant-propos	v
Sommaire	vi
Sigles et Abréviations	vii
Listes des tableaux	viii
Listes des graphiques – Figures et annexes	viii
Résumé.....	ix
Abstract	ix
Présentation de la structure	1
Introduction générale	2
Chapitre 1 : REVUE DE LITTERATURE	5
Chapitre 2 : Cadre méthodologique et Présentation des données	14
Chapitre 3 : Analyses des Résultats et Discussions	27
Conclusion et recommandations	41
Bibliographie.....	x
Annexes.....	xiii
Tables des matières	xv

Sigles et Abréviations

ADF	Augmented Dickey-Fuller
ANSTAT	Agence Nationale de la Statistique (Côte d'Ivoire)
ARIMA	AutoRegressive Integrated Moving Average
BCEAO	Banque Centrale des États de l'Afrique de l'Ouest
CLVC	Comité de Lutte contre la Vie Chère
COICOP	Classification of Individual Consumption According to Purpose
COPNI	Classification of the Purposes of Non-Profit Institutions Serving Households
COPP	Classification of the Purposes of Production
FAO	Food and Agriculture Organization
FMI	Fonds Monétaire International
HDX	Humanitarian Data Exchange
IHPC	Indice Harmonisé des Prix à la Consommation
INSEE	Institut National de la Statistique et des Études Économiques
KPI	Key Performance Indicator
LSTM	Long Short-Term Memory
PIB	Produit Intérieur Brut
RNN	Recurrent Neural Network
RMSE	Root Mean Squared Error
SARIMA	Seasonal ARIMA
SCN	Système de Comptabilité Nationale
STL	Seasonal and Trend decomposition using Loess
UEMOA	Union Économique et Monétaire Ouest Africaine

Listes des tableaux

Tableau 1: Liste des fonctions de consommation selon la NCOA.....	14
Tableau 2: critères de choix de modèles et leurs rôles	21
Tableau 3: Résultats des critères de sélection des modèles CAH et K-means	25

Listes des graphiques – Figures et annexes

Figure a: Page d'accueil de l'observatoire	28
Figure b: Module de l'analyse exploratoire des données	29
Figure c: Module de l'analyse des corrélations entre les séries d'indices	30
Figure d: Présentation de l'état de la base de données sur les indices de prix	31
Figure e: Module de choix du modèle de prédiction	31
Figure f: Exemple de choix de modèle	32
Figure g: Prédictions réalisées par chaque modèle	33
Figure h: Exemple de simulation de prix futur	33
Figure i: Module de classification.....	34
Figure j: Module de détection des anomalies et des alertes.....	36
Figure k: Module de traitement de la base de données.....	36
Figure l: Outil de collecte de données par webscrapping	37

Résumé

Dans un contexte inflationniste dans en Afrique de l'Ouest, et en Côte d'Ivoire en particulier, le suivi de la dynamique d'évolution des prix devient de plus en plus un enjeu majeur. L'analyse et la prédiction de l'inflation à partir des méthodes de machine learning s'inscrivent dans une logique de modernisation des dispositifs statistiques, avec pour ambition de produire un système interactif, automatisé et prédictif des prix à la consommation.

L'objectif principal de cette étude est de construire un observatoire des prix à la consommation en Côte d'Ivoire à partir des méthodes de machine learning en se basant sur les séries d'indices des prix sur la période de janvier 2017 à octobre 2024.

De la collecte des données sur les indices par webscraping, à la prédiction grâce à des modèles d'apprentissage, l'observatoire développé sous Streamlit permet une visualisation claire de la méthodologie adoptée et des résultats empiriques obtenus. Ces résultats révèlent une tendance haussière des prix alimentaires et de forts effets saisonniers dans les prix des transports et de l'énergie. Ces éléments confirment ainsi l'intérêt et la pertinence d'un tel dispositif.

Abstract

In an inflationary context in West Africa, and in Côte d'Ivoire in particular, monitoring price dynamics is increasingly becoming a major challenge. Analyzing and predicting inflation using machine learning methods is part of a broader effort to modernize statistical systems, with the goal of producing an interactive, automated, and predictive system for consumer prices.

The main objective of this study is to build a consumer price observatory in Côte d'Ivoire using machine learning methods based on price index series for the period January 2017 to October 2024.

From collecting index data through web scraping to predicting inflation using machine learning models, the observatory developed using Streamlit provides a clear visualization of the methodology adopted and the empirical results obtained. These results reveal an upward trend in food prices and strong seasonal effects in transportation and energy prices. These elements thus confirm the interest and relevance of such a system.

Présentation de la structure

L'École Nationale Supérieure de Statistique et d'Économie Appliquée (ENSEA) d'Abidjan, fondée en 1961, est un établissement public ivoirien d'enseignement supérieur et de recherche. Sa mission principale est de former des statisticiens pour les pays africains, notamment ceux d'expression française, ainsi qu'Haïti.

Historique et évolution

L'ENSEA visait initialement à former des agents techniques de la statistique. En 1963, elle élargit sa mission à la formation de cadres supérieurs. En 1969, en raison de son caractère international croissant, elle est érigée en établissement public d'enseignement supérieur doté de la personnalité civile et de l'autonomie financière. En 1982, ses objectifs sont étendus à la formation de cadres statisticiens au plus haut niveau, avec la création de la filière des Ingénieurs Statisticiens Économistes (ISE). Depuis lors, l'ENSEA a formé plus de 4 000 statisticiens issus de plus de vingt pays d'Afrique et d'Haïti.

Formations et filières

L'ENSEA propose plusieurs filières de formation initiale :

- Analystes Statisticiens (AS) : niveau licence
- Ingénieurs Statisticiens Économistes (ISE) : niveau master

Recherche et partenariats

L'ENSEA est reconnue comme un Centre d'Excellence Régional de l'UEMOA depuis 2005 et a été labellisée Centre d'Excellence Africain par la Banque mondiale en 2015. Elle est membre du Réseau des Écoles de formation Statistique Africaine (RESA), aux côtés de l'ISSEA de Yaoundé, de l'ENEAM de Cotonou et de l'ENSAE de Dakar.

Le personnel administratif de l'ENSEA compte une trentaine de membres qui attachent une grande importance à la recherche. L'institution encourage ses enseignants à mener des travaux de recherche pour renforcer leurs capacités et contribuer à la résolution de problèmes de société.

Avec plus de six décennies d'expérience, l'ENSEA d'Abidjan joue un rôle crucial dans la formation de statisticiens en Afrique. Son engagement envers l'excellence académique, la recherche appliquée et la coopération internationale en fait un acteur clé du développement statistique sur le continent.

Introduction générale

Contexte et justification

Le monde a récemment fait face à plusieurs crises et chocs notamment la crise de COVID 19 et le conflit géopolitique entre la Russie et l'Ukraine, qui ont impacté les prix de nombreux biens de consommation. Selon le rapport sur les perspectives de l'économie mondiale du FMI de janvier 2025, l'inflation globale devrait ralentir à 4,2 % en 2025, après avoir subi des hausses de 5,9% en 2024 et 6,8% en 2023. L'augmentation du niveau général des prix dans le monde entre donc dans une tendance baissière depuis que la crise russo-ukrainienne a exacerbé les tensions sur les marchés des céréales et de l'énergie. La hausse des prix à la consommation notamment sur les produits alimentaires et les services élémentaires (transport, énergie, etc.), réduit la capacité des ménages à satisfaire à leurs besoins de base. Autrement dit, l'inflation réduit le pouvoir d'achat des ménages et les oblige à réajuster leurs dépenses, à se tourner vers des produits de qualité moindre ou tout simplement à réduire la quantité de biens consommés. Ces ajustements peuvent entraîner dans les cas les plus extrêmes une insécurité alimentaire dans les pays faisant face à des fluctuations des prix à la hausse. Selon le rapport des Nations Unies sur l'état de la sécurité alimentaire et de la nutrition dans le monde de 2024, entre 8,9% et 9,4% de la population mondiale ont souffert de la faim en 2023, avec 152 millions de personnes en plus depuis 2019.

Le rapport de la BCEAO de Mars 2024 stipule que la contribution des prix des produits alimentaires à l'inflation totale de la sous-région s'est élevée à +1,7 point de pourcentage. Cela veut dire que les produits d'alimentation contribuent à hauteur de 1,7% à la hausse généralisée des prix dans la zone UEMOA et sont les plus soumis aux fluctuations à la hausse de leurs prix. C'est ainsi que la Côte d'Ivoire, comme la plupart des pays d'Afrique Subsaharienne, a connu une forte inflation ces dernières années, ce qui a eu pour conséquence d'impacter le pouvoir d'achat des ménages. Cette tendance à la hausse du niveau général des prix a ralenti depuis lors, et l'inflation en moyenne annuelle en Côte d'Ivoire se situe à (+ 2,8%) au mois de mars 2025, en baisse de 0,3 point de pourcentage par rapport au mois précédent. Cependant, certains secteurs, tels que le logement et les transports, ont continué de subir des hausses de prix, affectant particulièrement les populations vulnérables (ANSTAT).

L'inflation est un indicateur clé de la santé économique d'un pays. La collecte en temps réel des prix des denrées alimentaires et de leur évolution peut aider à améliorer les interventions

politiques et à assurer la sécurité alimentaire, en particulier pour les consommateurs les plus pauvres. Un suivi de l'évolution des prix permettrait de détecter rapidement les tendances inflationnistes et les postes de consommation source de ces tendances afin d'adapter les politiques économiques en conséquence. Les prédictions possibles des évolutions des indices de prix à la consommation sont de plus en plus réalisées à partir des méthodes de machine learning qui offrent de meilleurs résultats que de simples analyses statistiques (Siméon, 2022). Les décideurs peuvent ainsi, à partir de ces prévisions plus précises, intervenir plus rapidement sur le marché pour éviter que les ménages les plus vulnérables ne soient trop impactés par des hausses de prix. Les indices des prix à la consommation (IPC) sont couramment utilisés pour mesurer l'inflation et la déflation, servant ainsi d'outils essentiels pour les gouvernements et les banques centrales dans leurs prises de décision.

La flambée des prix des produits de consommation est un enjeu majeur en Côte d'Ivoire, affectant le pouvoir d'achat des ménages. Une surveillance efficace des prix peut contribuer à la stabilité financière, à la croissance économique et à la lutte contre la vie chère. Disposer d'un observatoire des prix permettrait de suivre l'évolution des tarifs des biens de consommation courante, d'en identifier rapidement les hausses anormales et d'orienter les politiques pour protéger les consommateurs.

Problématique

Les prix des biens de consommation ont subi plusieurs fluctuations ces dernières années à tel point de soulever la question de sécurité alimentaire dans le monde et particulièrement en Afrique. L'observation de l'évolution des prix sur le marché de la Côte d'Ivoire est essentielle pour anticiper sur les ajustements économiques à décider pour maintenir la dynamique économique du pays. Cela se fait principalement par l'indice harmonisé des prix à la consommation (IHPC). L'analyse de l'évolution de ces indices fournit une source d'information sur les politiques à adopter relativement à la régulation du marché. Les méthodes de machine learning sont des nouvelles méthodes innovantes qui permettent d'analyser et de prédire des valeurs de séries de données temporelles comme celles de l'IHPC, afin d'avoir de meilleures observations de l'évolution des prix et de la situation du marché. Ainsi, de ces observations faites sur la fluctuation des prix à la consommation, il serait pertinent de se demander :

Comment utiliser les méthodes de machine learning afin de construire un observatoire des indices de prix à la consommation en Côte d'Ivoire ?

Objectifs du projet

L'objectif de notre étude est de construire un observatoire dynamique des variations des prix à la consommation à partir des méthodes de machine learning afin d'éclairer les citoyens de la Côte d'Ivoire quant à l'évolution des prix et d'aider les autorités politiques dans leurs prises de décision.

Plus spécifiquement, il s'agira de :

- Collecter les données sur les prix des produits de consommation et leur évolution via des sources alternatives (internet)
- Construire un modèle de prévision des évolutions des indices des prix à la consommation à partir des méthodes de machine learning
- Élaborer une classification des séries d'indices, afin d'identifier des profils de comportement similaires
- Concevoir un dispositif permettant d'observer de manière régulière l'évolution des prix à la consommation.

Hypothèses de l'étude

Dans le cadre de cette étude, nous allons considérer indices des prix à la consommation selon les fonctions de consommation. Nous nous intéressons aux produits qui composent le panier de biens standard défini par la nomenclature COICOP formée de 12 fonctions de consommation et utilisé par l'ANSTAT dans le cadre du calcul de l'IHPC. La collecte de données historiques sur les indices se fera sur internet, sur le site de l'ANSTAT. Nous utiliserons des méthodes de webscraping pour collecter les données sur internet et nous utiliserons les méthodes de machine learning pour analyser ces prix et construire un observatoire.

Plan de travail

Pour mener à bien notre étude, nous présentons notre travail en trois chapitres. Le premier chapitre examine un panorama de projets et d'études existants – en Côte d'Ivoire, en Afrique et ailleurs – qui se sont intéressés à la construction d'observatoires ou de systèmes de suivi des prix et la revue de littérature. Le second chapitre présente les sources de données et l'approche méthodologique. Enfin, le troisième chapitre consiste en l'analyse et la présentation des résultats et de l'observatoire construit à partir des méthodes de machine learning.

Chapitre 1 : REVUE DE LITTÉRATURE

Dans ce chapitre nous allons présenter l'évolution des prix de manière générale dans le monde et en Afrique, ensuite nous allons présenter des exemples d'observatoire de prix dans le monde et enfin nous allons parcourir plusieurs études portant sur les méthodes de collecte de prix, les modèles de machine learning adaptés à la construction d'un observatoire de prix.

I. Définitions des concepts

Prix : Selon l'INSEE (2023), le prix d'un bien ou service représente « sa valeur exprimée en unités monétaires sur un marché donné à une période donnée ». En économie, il constitue un signal fondamental du marché, reflétant l'équilibre entre l'offre et la demande. Le prix joue ainsi un double rôle : il influence les comportements de consommation et oriente les décisions de production. Dans le cadre de la politique économique, le suivi de l'évolution des prix, notamment à travers l'indice des prix à la consommation (IPC), permet de mesurer l'inflation, d'ajuster les politiques salariales ou monétaires, et de préserver le pouvoir d'achat des ménages (FMI, 2004). Les prix ne sont pas fixes dans le temps : ils peuvent varier sous l'effet de facteurs internes (coûts de production, fiscalité) ou externes (cours mondiaux, taux de change, crises géopolitiques). Ainsi, l'analyse dynamique des prix constitue un enjeu majeur pour la régulation économique et sociale.

Consommation : La consommation désigne l'acte d'utilisation d'un bien ou d'un service dans le but de satisfaire un besoin. Selon la définition du Système de Comptabilité Nationale (SCN 2008), « la consommation finale représente la valeur des biens et services utilisés pour la satisfaction directe des besoins individuels ou collectifs des ménages ou des administrations publiques ». L'OCDE (2008) précise que « la consommation est la dépense de ressources pour satisfaire un besoin humain immédiat, par opposition à l'investissement ». L'analyse de la consommation est centrale pour l'élaboration des politiques publiques, car elle permet d'identifier les postes budgétaires les plus sensibles (alimentation, logement, transport) et de détecter d'éventuelles vulnérabilités sociales.

Observatoire : Un observatoire désigne une structure de veille, de collecte, d'analyse et de diffusion de données dans un domaine spécifique de manière régulière et systématique.

L'UNESCO (2019) définit un observatoire comme un « mécanisme institutionnel visant à produire une veille stratégique sur une thématique à fort enjeu, par le biais de données statistiques, de rapports analytiques et d'alertes ». L'intérêt d'un observatoire réside dans sa capacité à fournir des indicateurs fiables et continus, facilitant la prise de décision, la formulation des politiques publiques et le suivi-évaluation.

Machine Learning : Le machine learning ou apprentissage automatique est une branche de l'intelligence artificielle qui se définit comme « l'ensemble des méthodes permettant à un système d'apprendre à partir de données, sans avoir été explicitement programmé pour chaque tâche » (Mitchell, 1997). Selon l'OCDE (2021), il s'agit d'une technologie qui « permet aux algorithmes d'identifier des modèles complexes dans des données massives afin de faire des prédictions, des classifications ou des détections automatiques ». Le machine learning repose sur des techniques statistiques et computationnelles, telles que la régression, les forêts aléatoires ou les réseaux de neurones, et s'applique à des domaines variés comme la finance, la santé ou l'économie. Contrairement aux modèles économétriques classiques, les modèles de machine learning met davantage l'accent sur la performance prédictive que sur l'interprétation des paramètres.

II. Revue empirique de l'évolution des prix

Le ralentissement de l'activité économique observée lors de la crise COVID suivie du conflit russo-ukrainien, a contribué à faire grimper les prix à la consommation dans plusieurs pays (FMI, 2023). Ces deux chocs ont eu un effet commun, le ralentissement du commerce international. Forbes M. (2018) montre que la mondialisation a entraîné l'augmentation des flux commerciaux. Cela a modifié la dynamique de l'inflation à travers les relations de plus en plus étroites de fixation des prix des entreprises en fonction de l'évolution de l'offre et de la demande et en modifiant l'impact des taux de change sur l'inflation. Elle explique que les modèles permettant de prédire doivent inclure en plus des variables nationales, d'autres facteurs mondiaux tels que le volume des échanges commerciaux.

L'évolution à la hausse ou à la baisse des prix à la consommation a un effet sur le niveau de production et sur la croissance économique (théorie de l'offre et de la demande). Lorsque les prix sont élevés, les consommateurs ont tendance à réduire leur niveau de consommation, et à l'augmenter lorsque les prix baissent. Ce qui influence fortement le niveau de production des

entreprises qui anticipent sur la demande des consommateurs d'après la théorie keynésienne. C'est ainsi que Ndoricimpa (2017) montre qu'il existe un seuil d'inflation (6,7%) pour lequel la croissance du continent africain est affectée négativement, et cela que le pays soit à revenu intermédiaire ou à faible revenu. Une inflation basse est bénéfique à la croissance économique de l'Afrique.

Cependant, Diarrassouba et al. (2024), analysent la hausse généralisée des prix des produits vivriers en Côte d'Ivoire, plus précisément à Yopougon de 2005 à 2009 avec des hausses de certains produits (riz) allant à plus de 100%. En réponse à ces hausses des prix des produits alimentaires, les ménages ajustent leur comportement de consommation et on observe seulement 39% des ménages enquêtés qui prennent régulièrement un petit déjeuner en 2009 contre 73,2% en 2005. Ces études permettent de justifier la construction d'un observatoire des prix à la consommation pour surveiller le niveau de l'inflation globale afin de ne pas nuire à la croissance des pays d'Afrique.

La plupart des pays de l'Afrique sont des exportateurs de matières premières très souvent liées aux produits de consommation quotidienne tels que le pétrole pour le carburant. Chuku, et al. (2018) montrent que les fluctuations des prix des matières premières impactent différemment la stabilité macroéconomique des états africains. Certains Etats en raison de leurs dépendances aux exportations de matières premières sont plus affectés que d'autres, bien que ces effets dépendent des réponses de politique budgétaires et monétaires mises en place pour faire face à ces fluctuations.

III. Vue d'ensemble sur les observatoires de prix à la consommation

Le suivi en temps réel des prix à la consommation est un enjeu crucial pour toute économie. De nombreux observatoires de prix existent dans le monde.

1. Au niveau mondial

Nous avons d'abord le Billion Price Project (BPP) lancé par le MIT et Harvard. Le BPP a démontré qu'il est possible de suivre l'inflation en temps réel en collectant quotidiennement des millions de prix sur les sites d'e-commerce dans plus de 70 pays. Ce projet à l'origine académique a permis d'automatiser le relevé des prix en ligne de milliers de produits, construisant ainsi des indices de prix quotidiens hautement corrélés aux indices officiels de prix (IPC).

Nous avons ensuite l'observatoire de la formation des prix et des marges des produits alimentaires de la France. Dans le but de moderniser l'agriculture et la pêche dans une période de volatilité des prix, l'observatoire des prix et des marges des produits alimentaires de France a été mis sur pied par la loi n° 2010-874 du 27 juillet 2010. L'observatoire ne se surajoute, ni ne se substitue, à l'Autorité de la concurrence et a une mission d'information économique à caractère statistique. Les données de prix au détail utilisées par l'OFPPMA sont issues du panel d'une société de sondage (Kantar Worldpanel) qui fournit, par article de la nomenclature du panel, les quantités et les montants des achats des panélistes (12 000 ménages). Le prix moyen qui en est déduit est donc une moyenne pondérée par les volumes d'achat. Actuellement, les suivis sectoriels de l'OFPPMA portent sur une dizaine de filières (produits laitiers de vache, de brebis et de chèvre ; viande bovine ; viande porcine fraîche et salaisonnerie ; viande ovine ; volailles ; pain ; pâtes alimentaires ; fruits et légumes ; produits de la pêche et de l'aquaculture).

Enfin nous avons également l'Observatoire des prix belge créé en 2009 dans le cadre des compétences de l'Institut des comptes nationaux, a pour missions l'analyse de l'évolution des prix (à la consommation), des niveaux des prix, du fonctionnement de marché et des marges. Elle se distingue par son approche tripartite :

- Surveillance trimestrielle comparative de l'inflation avec les pays voisins
- Études sectorielles (énergie, télécoms, chaîne alimentaire)
- Analyses transversales des dynamiques de marché

Ses rapports, validés par un comité scientifique pluridisciplinaire, intègrent des comparaisons internationales systématiques des niveaux de prix.

2. En Afrique

En Côte d'Ivoire, le suivi officiel des prix à la consommation est assuré par l'Indice Harmonisé des Prix à la Consommation (IHPC), mis en place en 1998 dans le cadre de l'UEMOA. Depuis cette date, l'Institut National de la Statistique (INS) publie chaque mois l'IHPC, qui constitue la mesure officielle et comparative de la variation des prix dans le pays et la sous-région. Cet indice harmonisé repose sur des relevés mensuels des prix de divers biens et services dans des marchés et commerces des principales agglomérations. Par exemple, le Bulletin Mensuel des Prix à la Consommation des Ménages de la Ville de Man N°50 de février 2016, présente les prix moyens relevés sur les marchés locaux et leur variation mensuelle et annuelle, ce qui reflète l'évolution du « panier de la ménagère ».

En parallèle des indices officiels, des initiatives spécifiques ont vu le jour pour suivre les prix de produits particuliers dans plusieurs pays ouest-africains. Par exemple au Mali, l'Observatoire du Marché Agricole (OMA) collecte hebdomadairement les prix des principaux produits vivriers et agricoles sur un réseau d'environ 35 marchés à travers le pays. Les données recueillies (prix du mil, sorgho, maïs, riz, etc.) sont diffusées via des radios locales pour informer producteurs, commerçants et décideurs sur l'évolution des cours. Ces relevés réguliers servent non seulement à constater la variation intra-annuelle des prix (par exemple, flambée en période de soudure), mais alimentent aussi des analyses prédictives : l'OMA du Mali a montré qu'en observant les niveaux de prix en début de campagne de commercialisation, on peut estimer la production céréalière à venir à l'aide d'un modèle économique, les prix reflétant les anticipations des acteurs sur l'offre disponible.

Hormis l'IHPC, le gouvernement de Côte d'Ivoire a mis en place un Comité de Lutte contre la Vie Chère (CLVC) ces dernières années, visant à surveiller et maîtriser les prix des produits de grande consommation. Ses travaux incluent des relevés de prix sur le terrain et la publication périodique d'une liste de prix indicatifs de certains produits alimentaires et biens essentiels. Toutefois, le CLVC ne dispose pas encore d'une plateforme publique consolidant ces données sous forme d'un indice ou d'un tableau de bord dynamique accessible au grand public. L'idée d'un Observatoire national des prix à la consommation, intégrant des techniques modernes de collecte et d'analyse, constituerait donc une innovation en Côte d'Ivoire – s'inspirant à la fois des méthodes de l'INS (rigueur statistique) et des initiatives régionales (réactivité, focalisation sur certains produits sensibles).

IV. Revue de littérature sur les méthodes de collecte de prix

La collecte des données est la première pierre d'un observatoire de prix. Plusieurs approches peuvent être combinées pour obtenir une couverture large et à haute fréquence des tarifs des produits de consommation :

- Enquêtes et relevés de terrain traditionnel : C'est la méthode classique des instituts nationaux de statistique. Des enquêteurs formés se rendent dans un échantillon de points de vente (marchés, commerces, supermarchés) et notent les prix des produits selon une liste prédéfinie. C'est le cas de l'ANSTAT de la Côte d'Ivoire. Cette approche assure une bonne qualité statistique (contrôle des produits comparables, respect d'une méthodologie

homogène), mais elle est coûteuse et peu réactive (les données sont souvent disponibles avec un délai mensuel). Néanmoins, elle demeure indispensable pour des références officielles et pour des produits qui ne disposent pas d'autres sources de données.

- Données administratives : Il peut s'agir des relevés de prix effectués par les services du ministère du Commerce ou de données publiées sur des portails d'open data. Certaines villes ou pays mettent en ligne des prix observés. Au niveau international, des organisations comme la Banque mondiale, la FAO ou le PAM publient quelques jeux de données de prix librement accessibles.
- Webscrapping de prix (extraction de prix en ligne) : Cette méthode consiste à utiliser des scripts informatiques parcourant automatiquement des sites web marchands pour en extraire les informations de prix des produits. Avec la montée du commerce en ligne, la collecte de prix sur internet se développe de plus en plus. Elle peut fournir un flux quotidien de milliers de prix sans mobiliser d'enquêteurs tout en automatisant le système de collecte mis en place. Toutefois, elle nécessite des compétences techniques (programmation ou usage d'outils spécialisés) et tous les produits ne sont pas sur internet (surtout pour les marchés traditionnels).
- Crowdsourcing et données collaboratives : Concrètement, il s'agit de solliciter un réseau de personnes pour qu'elles transmettent les prix observés près de chez elles (magasin du quartier, marché local). Le recours aux citoyens ou à des contributeurs locaux équipés de smartphones, via des applications mobiles dédiées, des chatbots ou même des SMS, permet de démultiplier les points de collecte. L'avantage est d'obtenir des données de terrain très locales et en temps réel, y compris dans des zones où Internet fait défaut. Le défi est de contrôler la qualité des données et d'atteindre une couverture suffisante.
- Enquêtes par mobile (SMS/voix) : Proche du crowdsourcing mais plus pilotée, cette approche consiste à utiliser le téléphone pour interroger périodiquement des vendeurs ou ménages sur les prix. Dans le contexte humanitaire et des pays en développement, le Programme Alimentaire Mondial (PAM) gère une vaste base de données des prix alimentaires dans le monde, alimentée par ses bureaux pays. Dans certaines zones fragiles, le PAM a adopté dès 2013 un système de collecte par mobile (projet mVAM), obtenant ainsi des informations régulières même en zones de conflit ou d'accès difficile. Elle suppose néanmoins d'identifier au préalable des répondants de confiance et de les former aux questions. Les données de prix ainsi collectées sont publiées sous forme de bulletins et partagées en open data (plateforme HDX), fournissant aux gouvernements et humanitaires une alerte précoce en cas de flambée des prix locaux.

V. Méthodes de machine learning adaptées pour l'analyse des prix

La prévision des prix à la consommation repose sur diverses approches analytiques notamment le machine learning. Le machine learning ou en français « apprentissage machine » est une branche de l'intelligence artificielle. Elle utilise plusieurs méthodes de prévision des données parmi lesquelles des méthodes statistiques et des méthodes d'apprentissage.

1. Méthodes statistiques de prévision

Nous avons :

- Les modèles ARIMA : Ces modèles analysent les données historiques pour identifier des tendances et des motifs temporels. Le modèle ARIMA (AutoRegressive Integrated Moving Average) est particulièrement utilisé pour sa capacité à modéliser des séries non stationnaires en combinant des composantes autorégressives, de différenciation et de moyenne mobile.
- La méthode de Holt-Winters : c'est une méthode de lissage exponentiel qui modélise les composantes de la série temporelle à savoir la tendance, la saisonnalité et le niveau de la série. Ce modèle est simple à implémenter et est très efficace pour les séries présentant de la saisonnalité. Cependant, il ne prend pas en compte la saisonnalité changeante.
- Les modèles VAR : Ils établissent des relations statistiques entre les prix et d'autres variables explicatives, permettant ainsi de prévoir les évolutions futures.

2. Méthode d'apprentissage automatique

Les algorithmes de machine learning sont classés en deux catégories principales : apprentissage supervisé et apprentissage non supervisé.

a. *Apprentissage supervisé*

Il est utilisé lorsque les données d'entraînement comprennent des étiquettes ou des résultats connus. Son principe repose sur l'apprentissage d'un modèle sur des données connues capable de prédire la bonne valeur cible d'un nouveau jeu de données. Les principaux algorithmes incluent :

- Le modèle de forêt aléatoire introduit par Breiman (2001) est l'un des algorithmes les plus efficaces pour les tâches de classification et de régression. Ce modèle repose sur le principe de bagging (Breiman, 1996), un système d'agrégation qui génère plusieurs ensembles de données par bootstrap à partir de l'ensemble d'entrée d'origine. Chaque ensemble de

données générées correspond à une prédiction basée sur le modèle CART (Breiman et al., 1984) et toutes les prédictions sont agrégées dans un résultat final.

- Régression linéaire : On fait la régression lorsque la valeur cible à prédire est continue. Elle permet par exemple de calculer la prédiction du prix de vente d'une maison en fonction de plusieurs critères.
- Le modèle Prophet : ce modèle est utilisé dans le cas du traitement des séries temporelles. Il a été développé par Meta/Facebook et repose sur un modèle additif pour décomposer la série afin de mieux la prévoir. Cette méthode est très flexible, robuste et facile à ajuster (lorsque la saisonnalité change). Néanmoins, il est peu performant sur peu de données (sur des séries courtes) et nécessite la prise en compte des effets de calendrier.

b. Apprentissage non supervisé

Il est appliqué lorsque les données ne comportent pas d'étiquettes, visant à découvrir des structures ou des motifs cachés. Le principe ici est l'apprentissage d'un modèle capable d'extraire les régularités présentes dans les données afin de les regrouper pour mieux visualiser ou appréhender la structure de l'ensemble des données. Les techniques courantes comprennent :

- K-Means : l'algorithme de segmentation K-Means [28] en se basant sur les images pixelisées d'intensité b pour i parcourant la base d'apprentissage. La segmentation par la méthode des K-Means est une méthode itérative qui, quel que soit son point de départ converge vers une solution. La solution obtenue n'est pas nécessairement la même selon l'initialisation. Le but des K-Means étant de minimiser la variance à l'intérieur de chaque groupe.
- Clustering hiérarchique : On fait la classification lorsque la valeur cible à prédire est discrète. Elle admet de calculer la prédiction de l'état succès/échec d'un approvisionnement par rapport à un stock et compte tenu de différents facteurs.

c. Approches basées sur l'apprentissage profond (Deep Learning)

Le deep learning, et en particulier les réseaux de neurones profonds, offre des capacités avancées pour la prévision des séries temporelles complexes :

- Réseaux de neurones profonds : Inspirés du fonctionnement du cerveau humain, ces réseaux sont constitués de multiples couches de neurones artificiels. Grâce à un entraînement sur de vastes ensembles de données, ils peuvent modéliser des relations non linéaires complexes. Un réseau neuronal artificiel (ou simplement réseau de neurones) est un modèle prédictif qui reproduit le fonctionnement du cerveau. Les réseaux de neurones

constituent un sous-domaine 34 du machine learning pour réaliser des tâches identiques, mais lorsque l'espace probabiliste géré est plus complexe. Chen et al. (2001) et Nakamura (2006) traitent de la prévision de l'inflation à l'aide de réseaux neuronaux. Leurs études montrent que la qualité de prévision de ces réseaux neuronaux surpasse les modèles linéaires classiques pour des horizons plus courts. Suhartono (2005), quant à lui, compare les performances des réseaux neuronaux, ARIMA et ARIMAX dans la prévision de l'inflation en Indonésie. Il conclut que les réseaux neuronaux donnent une prévision d'inflation plus précise que les approches classiques (ARIMA et ARIMAX).

- Modèles LSTM (Long Short-Term Memory) : Ces réseaux de neurones récurrents sont conçus pour capturer les dépendances à long terme dans les données temporelles, ce qui les rend particulièrement adaptés à la prévision des séries chronologiques telles que les prix du pétrole ou l'inflation. Les réseaux LSTM sont un sous-type de réseaux de neurones récurrents (RNN) capables d'apprendre des dépendances à long terme. Ils ont été introduits par Hochreiter et Schmidhuber (1997) et perfectionnés par Graves et Schmidhuber (2005). Les RNN introduisent la notion de temps dans les réseaux de neurones traditionnels et sont utilisés pour modéliser les comportements dépendants du temps. Cependant, comme l'ont montré Bengio et al. (1994), les RNN sont incapables de gérer les dépendances à long terme en pratique, souffrant de « mémoire à court terme ».

d. Comparaison des performances des modèles en Côte d'Ivoire

Une étude comparative de Koffi S. (2022) menée sur la prévision de l'indice harmonisé des prix à la consommation (IHPC) en Côte d'Ivoire entre janvier 2012 et mai 2022 a évalué les performances des modèles ARIMA, Holt-Winters et LSTM. Les résultats ont montré que :

- Le modèle LSTM a obtenu les prévisions les plus précises, avec une amélioration de plus de 90%.
- Le modèle ARIMA a amélioré les prévisions de 68%.
- Le modèle Holt-Winters a enregistré une amélioration de 61%.

Ces résultats suggèrent que les techniques d'apprentissage profond, en particulier les modèles LSTM, offrent une supériorité notable sur les méthodes traditionnelles pour la prévision de l'inflation en Côte d'Ivoire.

Chapitre 2 : Cadre méthodologique et Présentation des données

I. Présentation des données

Les données utilisées dans la suite de nos travaux ont été collectées par des méthodes de webscrapping. Nous présenterons les sources des données qui nous ont permis de constituer la base de données sur les indices de prix à la consommation.

1. Source de données

Les données analysées proviennent du site officiel de l'agence nationale de la statistique (ANSTAT) de la Côte d'Ivoire. Ce sont les indices de prix à la consommation calculés pour chacune des 12 fonctions de consommation de la Nomenclature de Consommation Ouest Africaine (NCOA). Ces indices sont harmonisés car la méthodologie utilisée pour leur production est identique dans les huit pays membres de l'Union Economique et Monétaire Ouest Africaine (UEMOA) et basée sur la NCOA. Les différentes fonctions de consommation utilisées dans nos analyses sont :

Tableau 1: Liste des fonctions de consommation selon la NCOA

Code NCOA	Fonction de consommation
01	Produits alimentaires et boissons non alcoolisées
02	Boissons alcoolisées, tabac et stupéfiants
03	Articles d'habillement et chaussures
04	Logement, eau, gaz, électricité et autres combustibles
05	Meubles, articles de ménage et entretien courant du foyer
06	Santé
07	Transports
08	Communications
09	Loisirs et culture
10	Enseignement
11	Hôtels, cafés et restaurants
12	Biens et services divers

Source : UEMOA

Chaque fonction peut être décomposée en postes élémentaires de consommation (niveau 4 COICOP), utilisés pour des analyses plus détaillées.

Les indices de prix à la consommation calculés par l'ANSTAT sont mensuels et sont calculés à partir de prix collectés sur le terrain par des agents enquêteurs sur les produits composant

chaque poste de consommation. Les prix sont relevés régulièrement (mensuellement, généralement) sur le terrain, dans différents types de points de vente (marchés, supermarchés, boutiques, etc.) et, de plus en plus, en ligne. Les relevés couvrent tout le territoire ou les principales agglomérations selon la couverture souhaitée. Ils reflètent ainsi l'évolution des prix de chacune des différentes fonctions de consommation.

La collecte des données nécessaire à nos analyses se porte sur les indices calculés par l'ANSTAT en raison de l'indisponibilité des prix des éléments constitutifs de certaines fonctions de consommation en ligne et la difficulté à collecter par webscrapping des données sur les prix de certains sites populaires de e-commerce. Par exemple, pour la fonction bien et services divers qui comprends les postes de consommation de protection sociale et d'assurance dont les prix (le coût) ne sont pas accessibles par internet, cela est difficile d'utiliser des méthodes de webscrapping pour collecter ces données.

Cela justifie que nous ayons choisi de collecter les indices des prix à la consommation de toutes les fonctions à partir du site de l'ANSTAT qui a pour mission de produire ces statistiques suivant une période mensuelle. Les indices collectés par webscrapping couvrent la période de janvier 2017 à 2024. La méthode de webscrapping est adaptée à ce type de collecte et a permis de constituer une base de données de plusieurs indices des prix à la consommation indicés par le temps.

2. Variables de l'étude

Les variables utilisées dans la suite de nos analyses sont les indices mensuels par fonction et par poste de consommation, l'indice harmonisé des prix à la consommation (IHPC), l'inflation (en pourcentage) et le taux d'inflation en glissement annuel. Toutes ces variables sont indicées mensuellement.

II. Choix technologiques et justification

La mise en place de l'observatoire des prix repose sur des choix technologiques rigoureux visant à garantir la fiabilité scientifique, la rapidité de prototypage, la reproductibilité des résultats ainsi que l'extensibilité du système. Ces décisions, guidées par les besoins spécifiques du projet, s'inscrivent dans une logique de compromis entre robustesse, simplicité et performance.

1. Choix du langage et d'écosystème

Le langage principal utilisé est Python (version 3.11), choisi pour son écosystème particulièrement riche en matière de science des données, incluant des bibliothèques éprouvées telles que pandas, scikit-learn, statsmodels ou Prophet. Ce langage, déjà maîtrisé dans le cadre du cursus, permet un développement rapide tout en bénéficiant d'une communauté très active et d'une documentation abondante. Bien que ses performances brutes soient inférieures à celles de langages compilés comme C ou Julia, cette faiblesse est largement compensée par l'usage de bibliothèques optimisées (NumPy, Cython...).

La gestion de l'environnement a été assurée via Conda, combinée à un fichier requirements.txt permettant de figer les versions utilisées. Cette approche assure une séparation claire entre les phases de développement et de production, tout en facilitant la reproductibilité. Néanmoins, des conflits entre bibliothèques, notamment autour de Prophet et pystan, ont exigé un épinglage minutieux des versions.

2. Collecte et pré-traitement des données

La collecte des données s'appuie sur une combinaison de Selenium (avec ChromeDriver) et de BeautifulSoup, adaptée à la nature dynamique des pages de l'ANSTAT qui nécessitent une interaction avec du JavaScript. Cette solution permet de simuler les clics, télécharger des fichiers et récupérer des informations de manière fine. Le prétraitement des données est effectué via pandas (version 2.2), standard de facto pour la manipulation tabulaire, avec une intégration fluide à d'autres outils de visualisation ou d'analyse.

3. Analyse statistique & Machine Learning

Pour l'analyse statistique et le machine learning, plusieurs bibliothèques ont été mobilisées. statsmodels a permis la mise en œuvre de modèles traditionnels comme ARIMA, SARIMA, ou Holt-Winters, ainsi que de tests stationnarité (ADF, KPSS). Pour les tâches de classification, réduction de dimension ou évaluation, scikit-learn a fourni une API homogène et efficace. En complément, Prophet a été utilisé pour les prévisions à composantes additives, intégrant saisonnalité et effets calendaires de manière interprétable. Enfin, la détection de ruptures dans les séries a été prise en charge par la bibliothèque ruptures, qui implémente l'algorithme PELT de manière performante.

4. Visualisation interface

L'interface utilisateur repose sur Streamlit (version 1.33), permettant de transformer rapidement les scripts Python en applications web interactives. Cette solution a été préférée à des frameworks plus lourds comme Django ou FastAPI, jugés trop complexes pour un prototype. La visualisation s'appuie sur Plotly (Express et GraphObjects), offrant des graphiques interactifs, exportables en PNG ou HTML, incluant des graphiques en cascade, des heatmaps et des visualisations interactives d'ACP.

5. Déploiement et exploitation

Concernant le déploiement, le projet est entièrement conteneurisé grâce à Docker, garantissant une portabilité parfaite entre les environnements local, serveur ou cloud. Le principe « Build once, run anywhere » a été scrupuleusement respecté. L'automatisation des tâches récurrentes, comme les mises à jour mensuelles du jeu de données, a été confiée à des scripts Python planifiés via cron.

6. Alternatives examinées mais non retenues

Plusieurs alternatives ont été examinées mais finalement écartées. Scrapy, par exemple, bien qu'excellent pour le scraping massif, ne gèrait pas les pages dynamiques d'ANSTAT aussi bien que Selenium. Le couple Django/React, bien que plus puissant pour une application complexe, aurait nécessité un effort de développement considérable sans valeur ajoutée suffisante dans le cadre d'un MVP. Enfin, les approches de type deep learning (LSTM, NeuralProphet) ont été rejetées en raison du faible volume de données disponibles, rendant ces modèles instables et potentiellement surajustés.

Finalement, l'architecture technologique retenue repose sur des briques robustes, éprouvées et bien documentées. Elle offre un compromis optimal entre rigueur scientifique, facilité de développement, ouverture de l'écosystème et capacités de déploiement. Elle permet ainsi de disposer d'un observatoire évolutif, maintenable et extensible, tout en restant accessible aux équipes locales qui auront à terme la charge de son exploitation.

III. Approche méthodologique

En se basant sur la revue de littérature et le contexte de notre étude, nous allons utiliser des méthodes de webscraping pour collecter les indices de prix à la consommation daté de janvier

1963 à octobre 2024, ensuite réaliser une analyse descriptive des données collectées et des prévisions à partir des modèles de machine learning et enfin nous allons construire une plateforme de visualisation qui sera l'observatoire.

1. La collecte de données par webscrapping

La collecte des données par webscrapping a été réalisée sur le site de l'ANSTAT. Nous avons utilisé le package Selenium avec le langage de programmation python afin d'automatiser la collecte en ligne des indices de prix à la consommation. Il a été spécifié dans l'algorithme que la collecte des fichiers dans l'onglet « indicateurs » du site de l'ANSTAT ne devait concerner que les indices de prix à la consommation et ignorer tous les autres indices. Ainsi, au total, 120 fichiers relatifs aux indices mensuels des prix à la consommation ont été extraits via webscraping. Ces fichiers, structurés selon la nomenclature NCOA, ont été agrégés en une base consolidée après nettoyage et harmonisation. La base finale des indices de prix à la consommation comprend : 108 postes, 9 fonctions de consommation sur les 12 du système harmonisé de l'UEMOA et les données sur les indices IHPC, inflation et inflation en glissement annuel.

2. Traitement des données

La base de données sur les indices de prix à la consommation étant constituée, un traitement des données a été réalisé afin de rendre les données collectées exploitables.

La base est formée de 124 colonnes et retrace mensuellement les indices de prix à la consommation de 1964 à 2023. Cependant, un nombre important de valeurs manquantes est observé entre la période janvier 1963 à décembre 2016. Nous avons donc filtré la base à partir de janvier 2017 pour pouvoir travailler avec des données existantes. Néanmoins, une valeur manquante demeure en 2018 pour la variable inflation en glissement en annuel (infla_gliss).

Nous avons remarqué dans la base un problème de valeur pour la variable indiquant le mois car les valeurs 2, 8 et 12 représentant respectivement février, août et décembre, étaient toutes remplacées par 0. Nous avons corrigé cette erreur en imputant les bonnes valeurs pour chacun des mois.

IV. MODELISATION

Dans le cadre de la construction d'un observatoire des prix à la consommation à partir des indices collectés via webscrapping, plusieurs méthodes de prévision de machine learning basées sur des séries temporelles ont été retenues. Ces méthodes permettent de modéliser l'évolution des indices dans le temps, de détecter des patterns saisonniers et de prédire leur dynamique future. Les principaux modèles considérés dans nos analyses sont les suivants : ARIMA, S+ARIMA, HOLT-WINTERS et PROPHET. La littérature a validé les modèles ARIMA et SARIMA pour l'analyse et les prévisions de données temporelles portant sur les prix et leurs indices. Les méthodes HOLT-WINTERS et PROPHET ont été utilisées pour leur capacité à prévoir les séries à partir de leurs différentes composantes.

1. Le modèle ARIMA (AutoRegressive Integrated Moving Average)

Il utilise les valeurs passées et présentes des séries non stationnaires pour analyser leurs évolutions et prédire leurs futures valeurs. Un modèle ARIMA(p,d,q) est défini comme suit :

$$\Phi_p(L).(1-L)^d Y_t = \theta_q(L).\varepsilon_t$$

Avec :

- L : opérateur de retard, $LY_t = Y_{t-1}$
- $\Phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p$: partie autorégressive (AR)
- $\theta_q(L) = 1 + \theta_1 L + \dots + \theta_q L^q$: partie moyenne mobile (MA)
- d : degré de différenciation
- ε_t : bruit blanc (résidu aléatoire)

Si la variable n'est pas stationnaire (c'est-à-dire $d > 0$), alors la variable est rendue stationnaire par différenciation dans l'estimation du modèle. Ce modèle prend en compte la non-stationnarité de la variable mais ne traite pas directement la saisonnalité. Il faut utiliser le modèle SARIMA dans ce cas.

2. Le modèle SARIMA (Seasonality AutoRegressive Integrated Moving Average)

Il s'agit d'un complément du ARIMA qui prend en compte la composante saisonnière des séries à analyser afin d'affiner les résultats. SARIMA (p,d,q,s) inclut les composantes saisonnières et on a :

$$\Phi_p(L)\Phi P(Ls)(1-L)d(1-Ls)DY_t = \theta_q(L)\theta Q(Ls)\varepsilon_t$$

Avec :

- s : période saisonnière (ex. : $s=12$ pour données mensuelles)
- P, D, Q : ordres des composantes saisonnières AR, I, MA

3. Le modèle PROPHET

Ce modèle a été développé par META (Facebook). C'est un modèle additif qui décompose la série analysée en trois composantes, la tendance, la saisonnalité et les effets calendaires. Ce modèle s'adapte très bien pour les séries économiques irrégulières. Il se présente sous la forme :

$$Y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Où :

- $g(t)$: tendance (linéaire ou logistique avec ruptures de pente)
- $s(t)$: saisonnalité (annuelle, hebdomadaire ou personnalisée)
- $h(t)$: effets de jours fériés ou événements
- ε_t : bruit (résidu) supposé gaussien

Ce modèle utilise des fonctions de Fourier pour estimer et prédire les valeurs des séries temporelles.

4. Critères d'évaluation des modèles de prévision

Conformément aux recommandations méthodologiques de la littérature sur les séries temporelles (Armstrong, 2001), la performance de chaque modèle est jugée à l'aide de trois indicateurs complémentaires :

Tableau 2: critères de choix de modèles et leurs rôles

Indicateur	Formule	Rôle dans l'évaluation
RMSE – Root Mean Squared Error	$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$	Mesure l'écart-type des résidus ; pénalise fortement les erreurs importantes, ce qui en fait un bon révélateur de la variabilité de la prévision (utile pour détecter les pics d'erreur).
MAE - Mean Absolute Error	$MAE = \frac{1}{n} \sum_{t=1}^n y_t - \hat{y}_t $	Fournit l'erreur moyenne en unités natives ; robuste aux valeurs proches de zéro et facile à interpréter par les décideurs.
MAPE – Mean Absolute Percentage Error	$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t}$	L'écart moyen en pourcentage entre la valeur observée et la prévision. Chaque erreur est ramenée à l'échelle ce qui rend la métrique indépendante des unités.

- **Complémentarité** : RMSE est sensible aux grosses erreurs, MAE reflète l'écart moyen sans pondération quadratique, MAPE ramène tout à un pourcentage lisible.
- **Robustesse** : l'utilisation conjointe de RMSE et MAE limite le biais dû à la pénalisation disproportionnée des gros écarts.

En pratique, les modèles sont classés prioritairement sur le RMSE (critère de variance), puis départagés au besoin par le MAE et le MAPE. Cette hiérarchisation garantit que l'option retenue minimise d'abord les erreurs critiques tout en restant globalement précise et compréhensible.

V. Classification

Afin de mieux comprendre la dynamique des postes de consommation suivis dans l'observatoire, nous avons appliqué une méthode de classification basée sur des indicateurs statistiques extraits des séries temporelles. Cette approche est inspirée de celle privilégiée par Wang et al. (2006). L'objectif principal est de regrouper les postes ayant des comportements similaires, ce qui facilite la prise de décision stratégique, notamment dans le ciblage des

politiques de prix, l'identification des postes sensibles à l'inflation, ainsi que la surveillance des comportements atypiques.

1. Variables utilisées pour la typologie

Pour chaque série d'indice de prix à la consommation par poste, nous avons extrait les caractéristiques suivantes :

- Trend (*pente de tendance*) : pente d'une régression linéaire ajustée sur la série, indiquant la direction générale du poste (hausse, baisse ou stabilité à long terme).
- Season (*amplitude saisonnière*) : amplitude moyenne de la composante saisonnière obtenue par décomposition STL (Seasonal and Trend decomposition using Loess).
- Volatility (*écart-type des résidus*) : mesure de la variabilité de court terme après retrait des composantes saisonnières et de tendance.
- YoY (*croissance annuelle moyenne*) : moyenne des taux de croissance en glissement annuel, indiquant le rythme d'évolution structurelle du poste.
- CV (*coefficient de variation*) : rapport entre l'écart-type et la moyenne de la série, utilisé pour caractériser la dispersion relative.

Ces variables résument les dynamiques structurelles et cycliques des séries d'indices, et constituent une base robuste pour la classification.

2. Algorithmes

a. Clustering K-means

Le K-means est un algorithme de partitionnement simple et élégant (Iqbal, 2021) dont l'objectif est de minimiser la variance intra-cluster, c'est-à-dire la somme des carrés des distances euclidiennes entre chaque observation et le centre (« centroïde ») de son groupe. Le fonctionnement se déroule en cinq étapes :

- Choix du nombre de clusters K (prédéterminé).
- Initialisation aléatoire de K centroïdes.
- Affectation de chaque point x_i au centroïde le plus proche :

$$\forall i, K_i = \arg \min_k \|x_i - u_k\|_2$$
- Mise à jour des centroïdes :

$$\mu_k \leftarrow \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$

v. Itération des étapes 3–4 jusqu'à convergence (les labels ne changent plus).

Formellement, on minimise la fonction de coût :

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} ||x_i - \mu_k||^2$$

Cette méthode présente toutefois des limites, à savoir :

- Une sensibilité aux outliers : un point extrême peut déplacer significativement un centroïde.
- Une dépendance à l'initialisation : différents seeds peuvent conduire à des partitions différentes.
- Elle fait l'hypothèse de clusters sphériques et de taille comparable, parfois irréalistes.

Pour atténuer ces faiblesses, on peut recourir à des variantes comme **K-medoids**, qui remplace chaque centroïde par un véritable point de la donnée (médóide), renforçant la robustesse sur les valeurs extrêmes.

b. Classification ascendante hiérarchique (CAH)

La CAH est une méthode agglomérative qui construit une structure arborescente (dendrogramme) sans présupposer le nombre de clusters à l'avance (Hastie et al., 2009). Elle part de n clusters unitaires (chaque observation étant un cluster), puis elle fusionne itérativement les deux clusters les plus similaires, selon un critère de distance (souvent euclidienne) et une méthode de linkage (simple, complet, moyen, etc.). Le processus permet de construire un dendrogramme, qui sera coupé à un niveau qui donne le nombre de clusters souhaité.

Cette méthode présente les avantages de ne pas nécessiter de définir le nombre de clusters initialement et d'offrir une visualisation claire du nombre de clusters grâce au dendrogramme. Elle présente néanmoins des limites telles que la complexité des résultats lorsque qu'on a de très grands jeux de données et aussi le fait que les premières fusions conditionnent toute la hiérarchie, rendant parfois la structure rigide.

3. Critères de choix

Dans les travaux de Lallich et Lenca (2015) sur les indices de qualité en clustering, une liste de 27 indices est présentée. Cependant, nous n'utiliserons que trois des métriques exposées dans nos travaux.

a. Indice de Silhouette

L'indice de Silhouette $S(i)$ quantifie, pour chaque observation i , la capacité de son cluster à la regrouper de manière cohérente tout en la distinguant des autres clusters. Il s'obtient par :

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Où :

- $a(i)$ est la distance moyenne entre le point i et tous les autres points de son propre cluster.
- $b(i)$ est la plus petite distance moyenne entre i et tous les points de l'un des autres clusters (le plus "proche").

Il est interprété comme suit :

- $S(i) \approx 1$: i est très bien assorti à son cluster et éloigné des autres.
- $S(i) \approx 0$: i se situe à la frontière entre deux clusters.
- $S(i) < 0$: i est mal classé (il serait plus proche d'un autre cluster).

On retient le score global en moyenne sur tous les points : plus il est élevé, meilleures sont la cohésion et la séparation.

b. Indice de Davies–Bouldin

L'indice de Davies–Bouldin (DB) évalue la qualité de tous les clusters en comparant, pour chaque paire (i, j) , la somme des dispersions intra-cluster à la distance entre les centroïdes. Il est donné par :

$$DB = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \frac{\sigma_i + \sigma_j}{d(c_i, c_j)}$$

Où :

- K est le nombre de clusters,
- σ_i est la dispersion (écart-type ou rayon moyen) du cluster i ,
- c_i et c_j sont les centroïdes des clusters i et j ,

- $d(c_i, c_j)$ la distance entre ces centroïdes (typiquement euclidienne).

Un faible DB indique que les clusters sont à la fois compacts (peu de dispersion intra-cluster) et bien séparés (grande distance inter-cluster).

c. Indice de Calinski–Harabasz

Aussi appelé Critère de Variance, l'indice de Calinski–Harabasz (CH) mesure le rapport entre la variance inter-cluster et la variance intra-cluster. Il s'obtient grâce à cette formule :

$$CH = \frac{SSB/(K - 1)}{SSW/(n - K)}$$

Où :

- *SSB (Sum of Squares Between)* est la somme des carrés entre les clusters,
- *SSW (Sum of Squares Within)* est la somme des carrés au sein des clusters,
- *n* le nombre total d'observations,
- *K* le nombre de clusters.

Un CH élevé reflète des clusters denses et bien séparés.

4. Choix du meilleur modèle

En combinant ces trois indices, on obtient une évaluation robuste. Pour choisir la méthode de partitionnement la plus appropriée, nous avons comparé K-means, de K=3 en nous basant sur la méthode du coude (Voir annexe 1), et CAH de 4 clusters choisi aussi par la méthode du coude (Voir annexe 2) sur trois indices d'évaluation :

Tableau 3: Résultats des critères de sélection des modèles CAH et K-means

Méthode	Score de Silhouette	Calinski–Harabasz	Davies–Bouldin
K-means	0,4776	115,0209	0,8813
CAH	0,4495	104,8057	0,9218

Source : nos résultats sous python, les auteurs

Dans notre cas, K-means obtient le meilleur score de Silhouette et de l'indice Calinski–Harabasz. Ainsi, K-means a été retenu pour la typologie finale, car il offre le meilleur compromis entre cohésion et séparation des groupes.

5. La méthode de HOLT-WINTERS

Elle consiste en un lissage exponentiel triple des séries qui intègre une tendance, une saisonnalité et une moyenne lissée pour chacune des séries. Elle permet de modéliser trois composantes clés d'une série temporelle selon que le modèle est soit additif, soit multiplicatif.

A chaque période t , cette méthode met à jour les différentes composantes et on a :

$$\ell_t = \alpha(Y_t - st - m) + (1 - \alpha)(\ell_{t-1} + bt - 1)$$

$$bt = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)bt - 1$$

$$st = \gamma(Y_t - \ell_t) + (1 - \gamma)st - m$$

avec:

- ℓ_t : composante de niveau
- bt : composante de tendance
- st : composante saisonnière
- m : période de la saisonnalité (ex. : 12 mois)
- $\alpha, \beta, \gamma \in [0,1]$: coefficients de lissage

La reconstitution de la série se fait selon le modèle que suit la série.

Le modèle additif se présente comme suit : $Y_t = \ell_t + bt + st + \varepsilon_t$

Le modèle multiplicatif se présente comme suit : $Y_t = (\ell_t + bt) \times st + \varepsilon_t$

6. Construction de la plateforme de visualisation

Nous avons construit un outil de visualisation de l'observatoire des prix à la consommation en Côte d'Ivoire avec la bibliothèque Python Streamlit. Streamlit est une bibliothèque open-source en Python, c'est-à-dire que ses caractéristiques peuvent être vues et modifiées par d'autres utilisateurs, qui permet de créer des applications web interactives pour la visualisation de données, le machine learning et l'analyse de données. Son usage n'est possible qu'avec le langage de programmation python mais il est simple à utiliser et permet un déploiement facile en local via un Docker ou en ligne via un cloud ou un serveur.

Les données utilisées étant produites suivant une période mensuelle par l'ANSTAT, l'observatoire crée à travers la plateforme pourra être actualisée de manière mensuelle, selon la fréquence de publication sur le site de l'ANSTAT. Un bouton pour collecter à nouveau les nouvelles données est prévu à cet effet pour actualiser les informations fournies par l'observatoire à tout moment.

Chapitre 3 : Analyses des Résultats et Discussions

Ce chapitre présente les résultats obtenus à l'issue du processus de développement et de mise en œuvre de l'application logicielle dédiée à l'observatoire des prix à la consommation en Côte d'Ivoire. Nous détaillerons ici les fonctionnalités principales de l'application finale et aborderons les implications de l'application, tout mettant en lumière quelques analyses.

L'observatoire construit est accessible sous le lien : <https://observatoireprix-civ.streamlit.app/>

I. Stationnarité

Pour appliquer les modèles ARIMA, SARIMA et PROPHET, il est nécessaire d'avoir des séries temporelles stationnaires. Selon Lardic et Mignon (2002), une série temporelle est dite stationnaire si ses propriétés statistiques fondamentales, telles que la moyenne, la variance et la covariance, restent constantes au cours du temps.

Nous avons testé la stationnarité de toutes les séries de notre base grâce au test de Dickey-Fuller Augmenté (ADF). Ce test a pour hypothèse nulle le fait que la série présente une racine unitaire et n'est donc pas stationnaire. Les résultats du test ADF pour chacune des séries est disponible dans le module « Choix » de notre observatoire.

Toutes nos séries ne sont pas toutes stationnaires, elle nécessite donc d'être différenciées tout en tenant compte de la saisonnalité. Un test de saisonnalité est également réalisé à travers la décomposition STL pour identifier les séries présentant de la saisonnalité. La différenciation saisonnière garantit que les variations saisonnières des séries sont correctement modélisées, tandis que les termes capturent efficacement les dépendances périodiques. Cela assure que les modèles obtenus sont adaptés non seulement à la nature dynamique des séries temporelles, mais également à leur caractère saisonnier.

Dans notre application, la non-stationnarité est automatiquement prise en compte par les modèles ARIMA/SARIMA (différenciation) et par le modèle Prophet (décomposition). Nous notons toutefois qu'en l'absence de saisonnalité claire, le modèle Holt-Winters peut être moins efficace.

La détermination des ordres p et q pour le modèle ARIMA repose sur l'analyse des graphiques des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF). Ces graphiques

aident à identifier les termes autorégressifs et les termes moyenne mobile qui décrivent au mieux la structure des données. Ces paramètres ont ensuite été utilisés comme bornes dans la fonction `auto_sarima`, qui sélectionne automatiquement les valeurs optimales minimisant les critères d'information AIC et BIC. Toute cette paramétrisation est prise en compte dans la conception de l'observatoire de prix réalisé.

Pour chaque série de notre base, les modèles ARIMA, SARIMA, Holt-Winters et PROPHET ont été implémenté, en fonction de leur capacité à minimiser le Residuals Mean Square Error (RMSE) tout en capturant les dynamiques spécifiques des séries temporelles. L'analyse des résultats montre que, plusieurs séries sont mieux modélisées par le modèle SARIMA par rapport aux autres modèles (car il fournit le RMSE le plus petit). Cela indique que la prise en compte des variations saisonnières améliore significativement la précision des prévisions.

II. Architecture générale de l'application

Figure a: Page d'accueil de l'observatoire



Source : L'application développée

Cette architecture permet une manipulation efficace de l'application afin de réaliser plusieurs analyses sur les indices de prix. La barre de navigation verticale sur le côté gauche de l'application constitue le point central d'accès aux différents modules fonctionnels de l'application. Ces modules comprennent la plage d'analyse, la visualisation des données et analyses, et les outils de collecte et de traitement instantané des données. Elle présente aussi

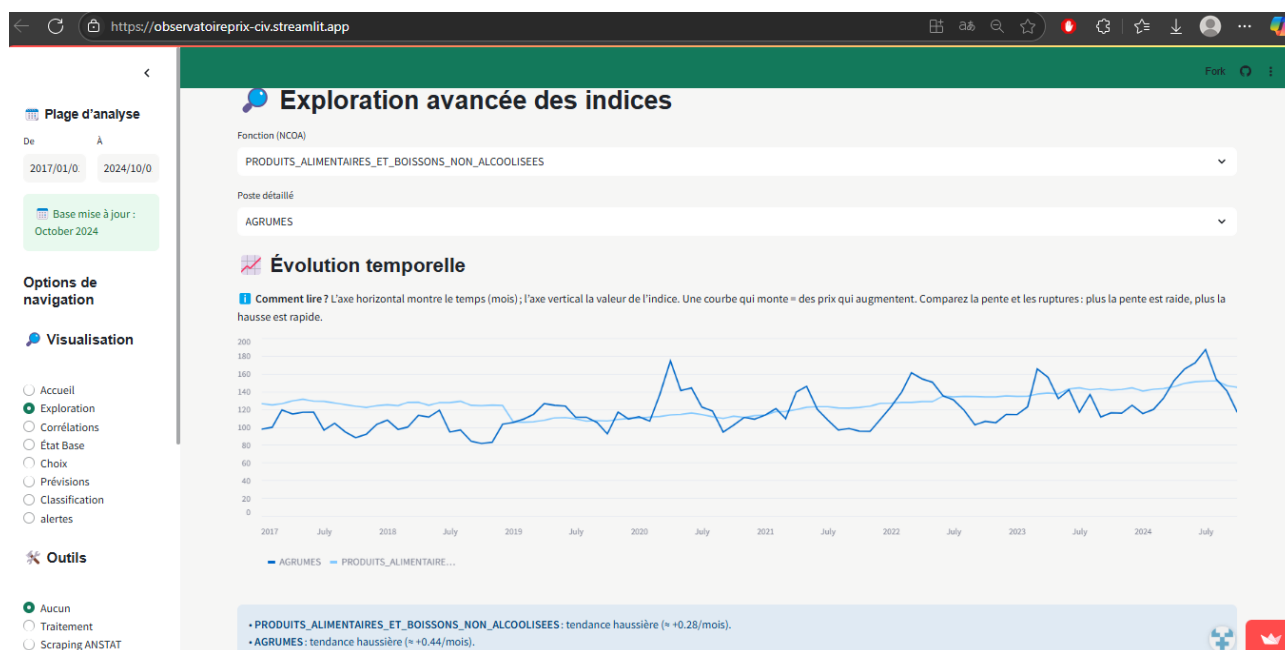
dans sa partie centrale un tableau de bord exposant les valeurs des indicateurs IHPC et inflation en glissement annuel du mois le plus récent dans les données collectées, la période couverte et le nombre de postes suivis. Nous observons aussi les 5 postes qui au cours du dernier mois ont subi une forte hausse et une forte baisse de leurs prix. Ainsi, les agrumes ont vu leurs prix baisser en octobre 2024 d'environ 17%.

Pour mieux comprendre la manière dont cette architecture se traduit en fonctionnalités pratiques, il est opportun d'examiner en détail les différents modules qui composent l'application. Chacun de ces modules répond à des besoins spécifiques identifiés lors de la phase de conception et contribue à l'atteinte des objectifs du projet. Nous commencerons par le module des outils de collecte et traitement instantané des données, qui constitue la première étape du processus de gestion des informations au sein de l'observatoire. Ensuite, nous verrons le module de visualisation et d'analyses de données. Enfin, le module de prédiction sera exposé.

1. Exploration

Cette section permet une visualisation interactive des indices de prix par poste de consommation (alimentation, transport, etc.).

Figure b: Module de l'analyse exploratoire des données



Source : L'application développée

Ce module permet de visualiser les données brutes après exploration pour en dégager avant l'analyse, certaines caractéristiques. L'utilisateur peut y observer l'évolution temporelle des

indices par fonction et ensuite par poste de consommation et observer les tendances de chacune des séries. Nous voyons que la fonction de consommation des produits alimentaires a une tendance haussière sur toute la période de janvier 2017 à octobre 2024.

2. Corrélations

Cette section explore les corrélations entre les différentes fonctions de consommation, postes de consommation, IHPC et en outre toutes les séries.

Figure c: Module de l'analyse des corrélations entre les séries d'indices



Source : L'application développée

Elle met en évidence les postes dont les indices évoluent de manière similaire et de manière inverse avec l'IHPC. Nous observons ainsi que la fonction de consommation « communication » a une forte corrélation positive de l'ordre de 0.79 tandis que la fonction de consommation « loisirs et culture » est faiblement corrélée à l'IHPC. Ces résultats permettent d'identifier les postes et les fonctions de consommation responsables indirectement du niveau de l'inflation.

3. État Base

Figure d: Présentation de l'état de la base de données sur les indices de prix



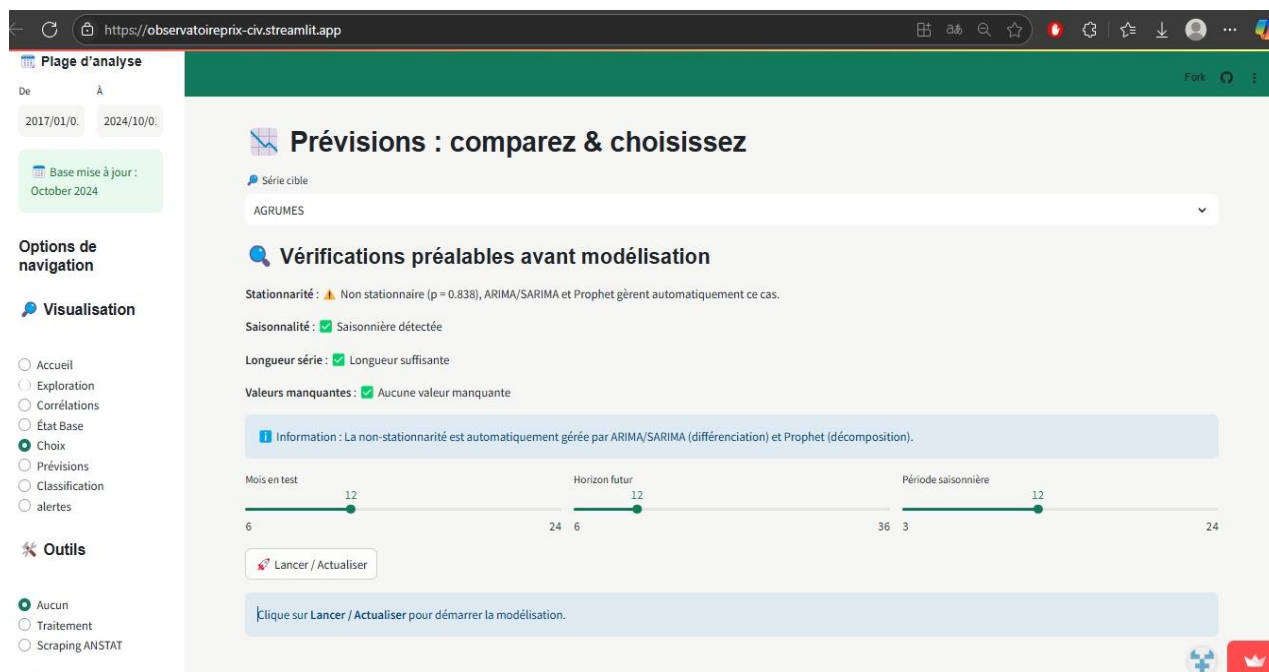
Source : L'application développée

Cette section fournit une synthèse de la base de données en fournissant le nombre de périodes, le nombre de variables collectées, la période couverte et la présence ou non de valeurs manquantes.

4. Choix

Cette interface permet la sélection des séries et du modèle à utiliser. Elle présente les résultats des tests de stationnarité et de saisonnalité mais aussi des caractéristiques de la série choisie.

Figure e: Module de choix du modèle de prédiction

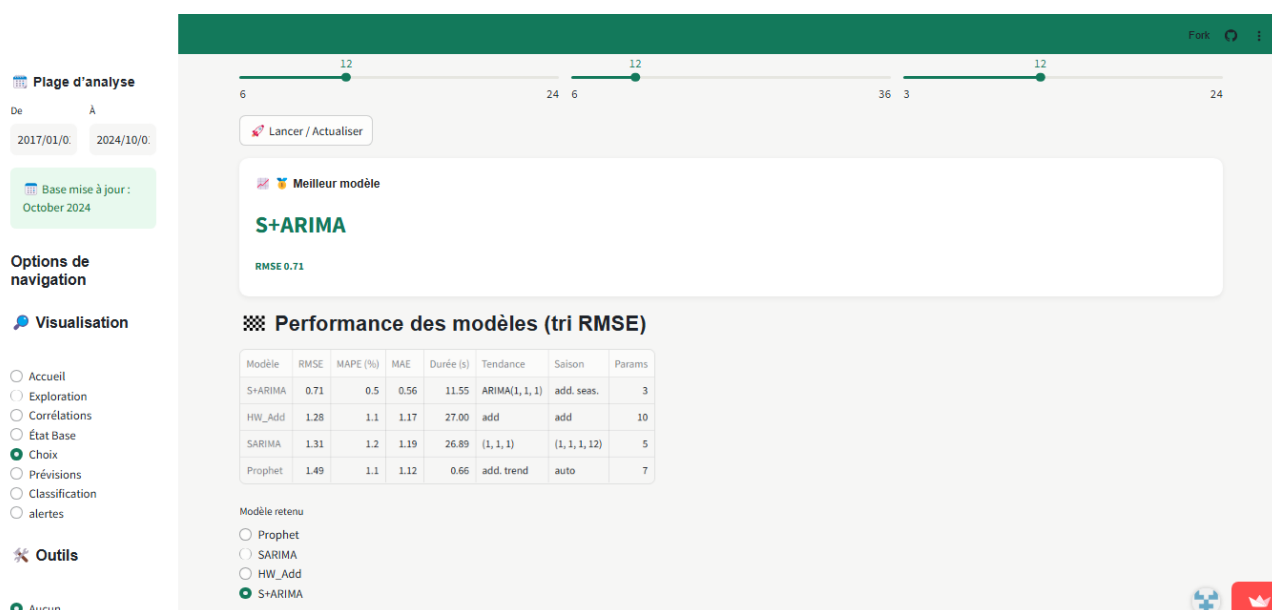


Source : L'application développée

Lorsque tous les tests sont bons, l'horizon de prévision peut être réglé allant jusqu'à 36 mois, soit 3 ans plus tard. Une fois tous les réglages faits, la modélisation peut être lancée via le bouton « Lancer/Actualiser ».

On obtient de nouvelles surfaces dans l'onglet (voir ci-dessous). Nous pouvons voir que l'application propose le modèle S+ARIMA pour modéliser et prédire la série des « Agrumes » avec un RMSE faible de 0,71. Il est toutefois possible de choisir un autre modèle pour analyser la série et la prédire.

Figure f: Exemple de choix de modèle

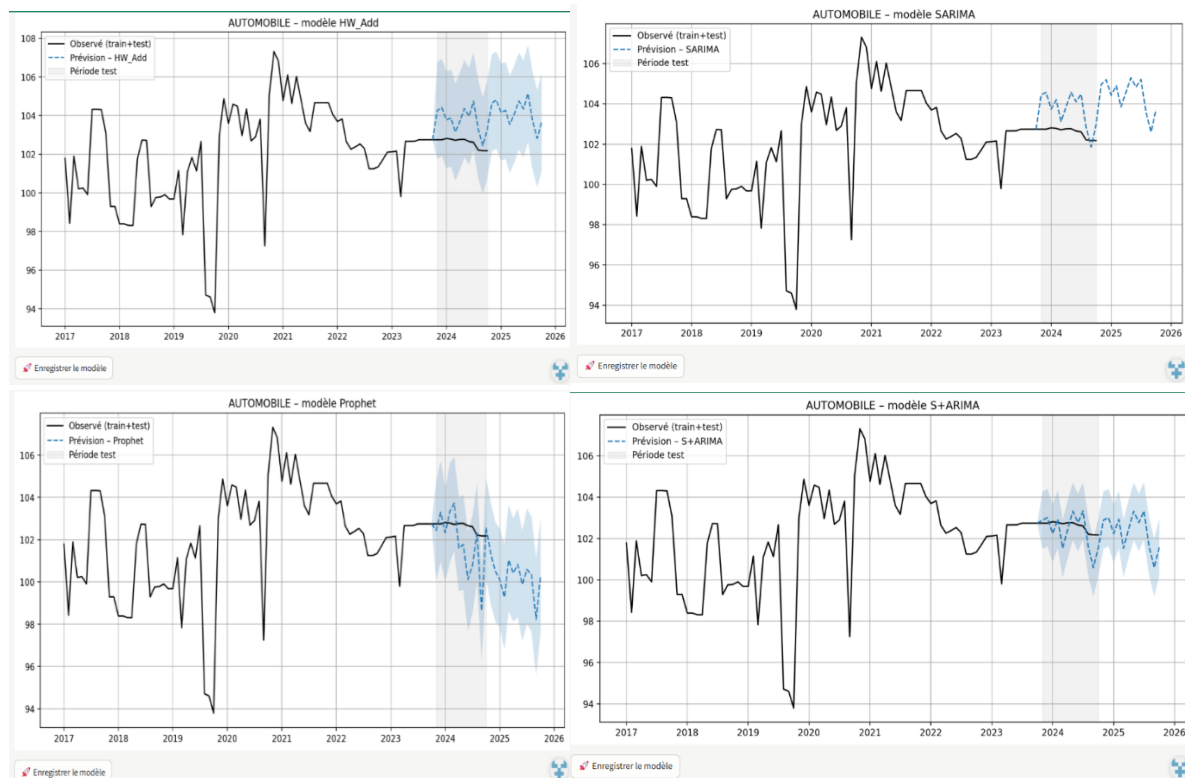


Source : L'application développée

5. Prévisions

Dans la section prévision, nous obtenons les projections faites pour la série sélectionnée au niveau de l'onglet « Choix ». Ainsi nous avons quatre projections pour chacun des modèles implémentés.

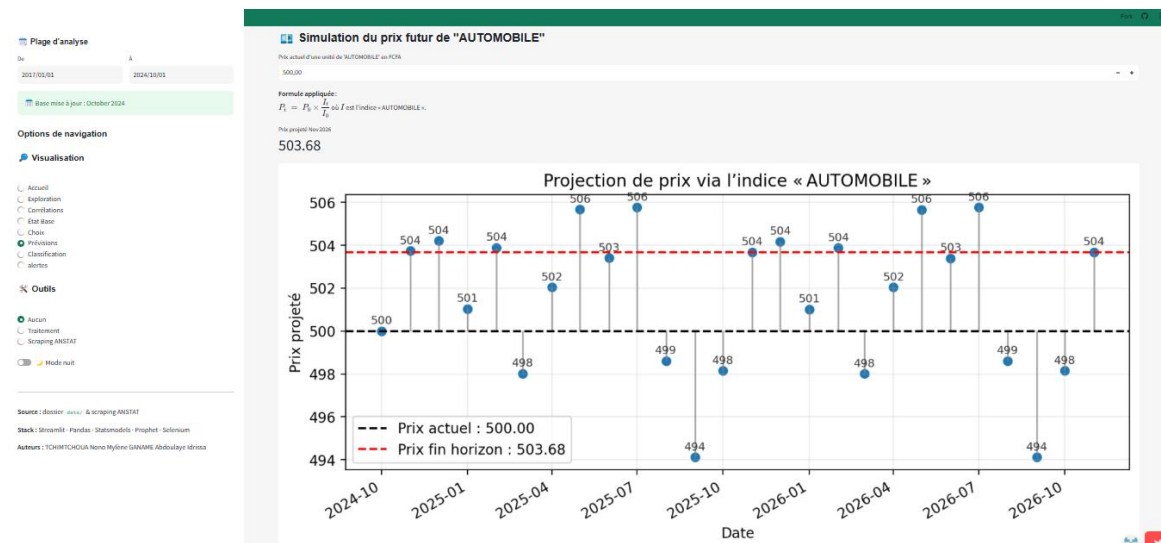
Figure g: Prédictions réalisées par chaque modèle



Source : L'application développée

Il est également possible de simuler le prix futur moyen d'un poste de consommation en entrant un prix connu dans l'application. Sur cette figure, nous avons entré « 500 » francs comme prix connu du poste « Automobile » dans la dernière période à jour (octobre 2024) et nous obtenons que le prix de ce poste évolue à la hausse tout le dernier trimestre de 2024 avant de ralentir en janvier 2025.

Figure h: Exemple de simulation de prix futur



Source : L'application développée

6. Classification

Cet onglet permet de visualiser le regroupement des postes de consommation selon leur comportement de prix par notre algorithme de K-Means retenu.

Figure i: Module de classification



Source : L'application développée

On obtient, des visualisations du comportement de chaque groupe selon leur tendance, saisonnalité et leur volatilité. Les groupes formés sont :

i. Cluster 0 : « Biens et services stables à faible volatilité »

Ce groupe regroupe principalement des postes dont l'indice évolue de façon linéaire et régulière. Sa pente de tendance moyenne est très faible (0,07 point d'indice par mois), ce qui signifie qu'il n'existe pas de dynamique haussière ou baissière forte sur le long terme. La saisonnalité est modérée (0,67 sur l'échelle de saisonnalité), traduisant des variations cycliques peu prononcées au cours de l'année. Enfin, la volatilité est très faible (écart-type mensuel $\approx 1,25\%$), attestant d'une stabilité remarquable dans les fluctuations de prix.

En pratique, ces caractéristiques font de ce cluster un « pôle de stabilité » où les postes (logement, enseignement, services de santé, loisirs, etc.) évoluent de manière prévisible, sans pics saisonniers ni chocs ponctuels. Il constitue une référence : toute déviation significative de ces profils, par exemple une hausse soudaine, doit alerter les analystes et guider les actions de veille.

ii. Cluster 1 – « Dépenses essentielles à forte saisonnalité et haute volatilité »

Ce groupe regroupe essentiellement des postes dont l'évolution des prix est marquée à la fois par une tendance haussière modérée (0,35 point d'indice par mois), une saisonnalité très prononcée (1,76 sur l'échelle de saison) et une volatilité élevée (écart-type mensuel $\approx 3,37\%$).

Concrètement, on y trouve des postes comme les carburants, les loyers, certains produits alimentaires de base (céréales, huiles, légumineuses), ainsi que les transports. Ces catégories sont soumises à des pics saisonniers et à des fluctuations fortes. Cela implique qu'il faut une surveillance renforcée de ces postes et une attention particulière à la conjoncture externe susceptible d'accentuer ou d'atténuer ces cycles.

En résumé, le cluster 1 représente les postes "essentiels" dont les prix montent régulièrement, mais avec de fortes oscillations saisonnières et des chocs ponctuels qu'il convient de traiter comme des signaux d'alerte prioritaires.

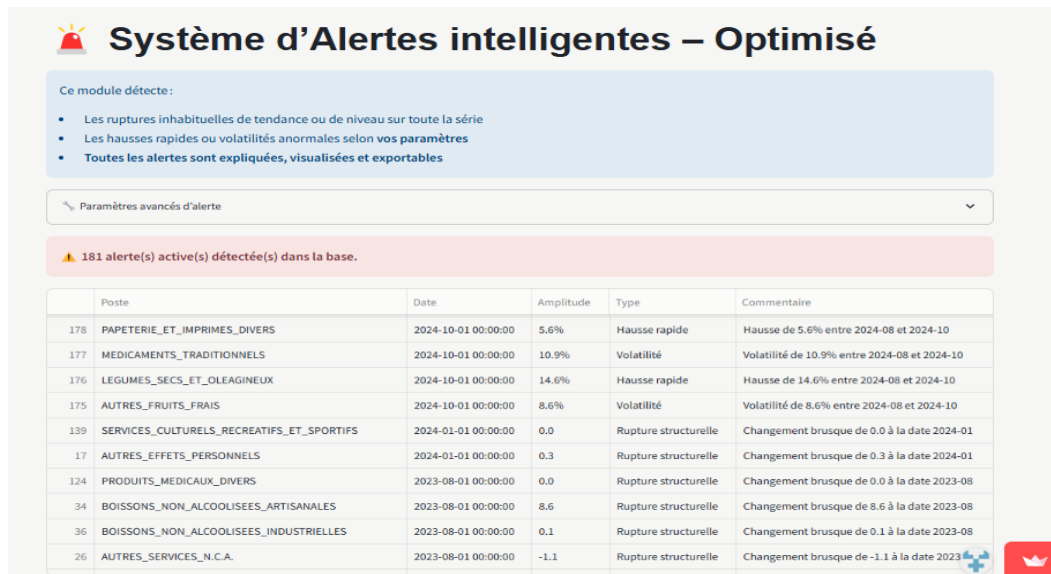
iii. Cluster 2 – « Produits frais & hyper-saisonniers »

Ce groupe se caractérise par une forte tendance haussière (0,81 point d'indice par mois), une saisonnalité extrêmement prononcée (7,40) et une volatilité très élevée (écart-type mensuel $\approx 7,31\%$). On y retrouve essentiellement des produits périssables : légumes frais, fruits, tubercules, poissons frais ou séchés. La courbe de tendance montre une accélération notable de l'indice du groupe à partir de 2020, traduisant une hausse structurelle portée par la demande et les perturbations d'approvisionnement. Les pics observés sont dus à des périodes de récolte ou de forte consommation (juin-juillet, mois de ramadan) et les creux signalent la fin de saison ou encore des difficultés logistiques. Ces postes nécessitent donc une veille renforcée et des stratégies d'anticipation pour atténuer l'impact des variations sur l'indice global de prix à la consommation.

7. Alertes

Ce module détecte les anomalies dans les données : les ruptures structurelles, les hausses rapides et des extrêmes volatilités en identifiant les périodes concernées. Ce module est utile pour la surveillance des hausses brutales.

Figure j: Module de détection des anomalies et des alertes

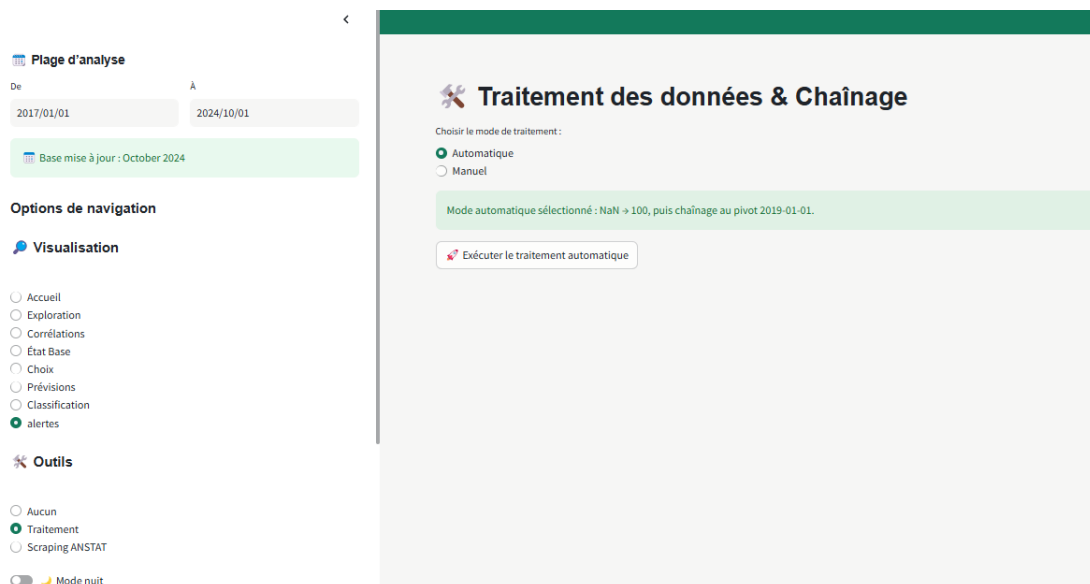


Source : L'application développée

8. Traitement

Ce module donne l'opportunité de traiter les données en tenant compte du chaînage. Il s'agit du changement de l'année de base dans le calcul des indices de prix à la consommation. Le traitement peut être automatique ou manuel.

Figure k: Module de traitement de la base de données



Source : L'application développée

9. Scraping

Cette section permet de lancer la collecte en ligne instantanée des indices de prix à la consommation sur le site de l'ANSTAT. Une fois les données collectées, une fusion est nécessaire pour mettre à jour la base de données déjà disponible dans l'application.

Figure 1: Outil de collecte de données par webscrapping

The screenshot displays the 'Scraping ANSTAT' web application interface. The left sidebar contains navigation links: Accueil, Exploration, Corrélations, État Base, Choix (selected), Prévisions, Classification, and alertes. Below this are 'Outils' (Aucun, Traitement, Scraping ANSTAT) and a 'Mode nuit' toggle. The main content area shows the 'Scraping ANSTAT' header, a 'Lancer le Scraping' button, and a status message: 'Démarrage du processus de scraping...'. A table lists extracted links with columns for index, description, and URL.

0	1	2	3
0	1	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION SERVICES FINANCIERS	https://www.an
1	2	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION MATÉRIEL DE TÉLÉPHONIE ET DE TÉLÉCOPIE	https://www.an
2	3	INDICE DE PRIX DE PRODUCTION DANS L'INDUSTRIE (IPPI) PAR DIVISION REGROUPEE INDUSTRIES ÉLECTRONIQUES, INFORMATIQUES, ÉLECTRIQUES ET DE MACHINES	https://www.an
3	4	INDICE DU CHIFFRE D'AFFAIRES DANS L'INDUSTRIE PAR DIVISION REGROUPEE INDUSTRIES ÉLECTRONIQUES, INFORMATIQUES, ÉLECTRIQUES ET DE MACHINES	https://www.an
4	5	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION EQUIPEMENT PHOTOGRAPHIQUE, CINÉMATOGRAPHIQUE, OPTIQUE, AUTRE BIEN DURABLE À FONCTION RÉCREATIVE ET CULTURELLE	https://www.an
5	6	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION MATÉRIEL DE TRAITEMENT DE L'INFORMATION ET SUPPORTS D'ENREGISTREMENT DE L'IMAGE ET DU SON	https://www.an
6	7	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION CÉRÉALES NON TRANSFORMÉES	https://www.an
7	8	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION BOEUF	https://www.an
8	9	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION MOUTON - CHÈVRE	https://www.an
9	10	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION PORC	https://www.an
10	11	INDICE HARMONISÉ DES PRIX À LA CONSOMMATION PAR POSTE DE CONSOMMATION MOUTON - CHÈVRE	https://www.an

Below the table, it indicates 'Téléchargement de : SERVICES FINANCIERS'.

The second screenshot shows the same interface after clicking 'Lancer le Scraping'. It displays a list of downloaded files with checkboxes indicating success: 'Fichier rangé : /app/excels_anstat_new/SERVICES_FINANCIERS.csv', 'Fichier rangé : /app/excels_anstat_new/MATERIEL_DE_TELEPHONIE_ET_DE_TELECOPIE.csv', 'Fichier rangé : /app/excels_anstat_new/EQUIPEMENT_PHOTOGRAPHIQUE_CINEMATOGRAPHIQUE_OPTIQUE_AUTRE_BIEN_DURABLE_A_FONCTION_RECREATIVE_ET_CULTURELLE.csv', 'Fichier rangé : /app/excels_anstat_new/MATERIEL_DE_TRAITEMENT_DE_L_INFORMATION_ET_SUPPORTS_D_ENREGISTREMENT_DE_L_IMAGE_ET_DU_SON.csv', 'Fichier rangé : /app/excels_anstat_new/CEREALES_NON_TRANSFORMEES.csv', 'Fichier rangé : /app/excels_anstat_new/BOEUF.csv', 'Fichier rangé : /app/excels_anstat_new/MOUTON_CHEVRE.csv', and 'Fichier rangé : /app/excels_anstat_new/PORC.csv'.

Source : L'application développée

III. ANALYSE DU MARCHÉ DE CONSOMMATION DE LA CÔTE D'IVOIRE

Les prix ont évolué de façon croissante depuis 2017 en Côte d'Ivoire. Nous allons analyser les tendances de la consommation en Côte d'Ivoire sur la période de janvier 2017 à octobre 2024. L'observatoire développé sous Streamlit permet une visualisation interactive de l'évolution des indices de prix à la consommation en Côte d'Ivoire par poste et fonction de consommation selon la nomenclature COICOP. L'analyse des séries temporelles issues de cette plateforme révèle plusieurs dynamiques économiques importantes sur la période récente.

1. La structure du marché de consommation

Le paysage de consommation ivoirien a fortement évolué ces dernières années. L'indice de prix de la fonction « Produits alimentaires et boissons non alcoolisées » affiche une tendance haussière marquée sur l'ensemble de la période observée. Cette dynamique traduit un renchérissement durable du panier alimentaire du ménage et cela résulte de la hausse des coûts logistiques et d'importation et des effets climatiques sur la production agricole nationale (sécheresses, inondations, etc.). Cela reflète une inflation de type « importée » et structurelle, impactant directement le pouvoir d'achat des ménages.

Les fonctions « Transport » et « Logement, eau, gaz, électricité » présentent une saisonnalité visible et une volatilité accrue, c'est-à-dire que leurs prix varient très fréquemment. Des hausses nettes sont observées en fin d'année (effets de fêtes) et en période de réajustement des prix du carburant, notamment à la suite de la libéralisation partielle des prix en 2022. Ces hausses traduisent l'impact des choix politiques tarifaires et les répercussions internationales sur les prix énergétiques. Elles rappellent le rôle de la régulation gouvernementale dans l'amortissement des chocs inflationnistes.

À l'inverse, les fonctions telles que « Loisirs et culture », « Articles d'habillement et chaussures » et « Biens et services divers » conservent une évolution plus modérée de leurs indices des prix sur plusieurs trimestres. Cela pourrait vouloir signifier que les ménages accordent moins de poids aux dépenses non essentielles telles que les loisirs ou le vestimentaire. Il s'agit alors d'un ajustement de la demande des ménages, qui dépriorisent ces dépenses en période de pression sur les produits de première nécessité.

De manière plus globale, l'application montre une accélération visible des indices de prix à partir de 2020, année marquée par la pandémie de COVID-19. Cette poussée inflationniste s'est prolongée jusqu'en 2023, exacerbée par le conflit entre la Russie et l'Ukraine en 2022.

Cette augmentation globale et soutenue des prix est due à la perturbation des chaînes logistiques mondiales, à la hausse des prix internationaux des matières premières et à la dépréciation locale du franc CFA par rapport à l'euro.

2. Implications économiques

La hausse soutenue des prix des produits alimentaires, qui constituent une part importante du panier de consommation des ménages, impose une pression directe sur le pouvoir d'achat, notamment pour les couches les plus vulnérables. Cela aggrave les inégalités de consommation et peut conduire à une baisse du niveau de vie moyen, surtout dans les zones urbaines où les produits alimentaires sont d'avantage commercialisés que produits par les ménages.

Les fonctions de consommation moins essentielles telles que les loisirs ou l'habillement, connaissent une baisse de la demande. Cela reflète des arbitrages économiques opérés par les ménages, qui tendent à orienter leurs dépenses vers des biens de première nécessité. Un ralentissement des activités dans ces secteurs dits « secondaires ou tertiaires » est alors observé, réduisant ainsi leur contribution au produit intérieur brut (PIB).

L'inflation observée est fortement due à des facteurs externes tels que les perturbations engendrées par le conflit russo-ukrainien. Cela montre que l'économie ivoirienne est fortement dépendante de l'extérieur, ce qui finalement réduit la marge de manœuvre du gouvernement pour contrôler l'inflation.

Une inflation persistante sans hausse équivalente des revenus peut engendrer une spirale inflationniste si les agents économiques anticipent une hausse continue des prix. En effet, si les consommateurs anticipent une hausse des prix, ils vont chercher à consommer immédiatement (hausse de la demande), ce qui va faire augmenter les prix selon la loi de l'offre et de la demande. Ce qui peut conduire à des déséquilibres économiques, une perte de confiance en la monnaie et des tensions sociales si les inégalités se creusent.

IV. Perspectives d'évolution de l'application

L'observatoire actuel constitue une avancée significative pour la surveillance de l'évolution des prix des biens et services de consommation en Côte d'Ivoire en se concentrant sur la période de janvier 2017 à octobre 2024. Toutefois, le potentiel de cette plateforme pourrait être encore plus mis en avant pour s'adapter aux besoins des utilisateurs.

1. Extension aux autres produits

A ce jour, l'observatoire ne se limite qu'à la visualisation des indices de prix à la consommation. Cependant, d'autres produits nécessitent une observation en temps réels de leurs prix tels que les matières premières.

En étendant la collecte et l'analyse à ces autres produits, l'observatoire offrira une vue complète du comportement des prix des biens depuis la production jusqu'à la consommation. Cela pourrait donner aux décideurs publics une meilleure vue d'ensemble sur les causes des tendances inflationnistes.

2. Ajout de filtres géographiques

L'intégration d'un filtre par région, district ou ville serait une bonne évolution apportée à l'observatoire. Une visualisation graphique des localités soumises à de fortes hausses de prix rendraient l'analyse de l'inflation plus profonde et précise. Cela permettrait de cibler les zones affectées par des tensions inflationnistes.

Conclusion et recommandations

Dans un contexte marqué par une instabilité des prix à l'échelle mondiale et régionale, la présente étude s'est attachée à concevoir un outil innovant de surveillance économique : un observatoire dynamique des prix à la consommation en Côte d'Ivoire, fondé sur les méthodes de machine learning. L'objectif principal de ce travail était de répondre à la nécessité pour les décideurs publics et les citoyens de disposer d'un outil de suivi des évolutions de l'Indice Harmonisé des Prix à la Consommation (IHPC), afin de mieux comprendre la dynamique inflationniste et de soutenir la prise de décision en matière de politique économique.

Pour répondre aux besoins identifiés, nous avons formulé l'hypothèse selon laquelle les indices des prix à la consommation publiés par l'ANSTAT peuvent, une fois collectés via webscrapping et analysés à l'aide de techniques de machine learning, permettre de construire un observatoire fonctionnel et prédictif, capable de détecter les alertes concernant l'inflation, de classer les postes sensibles et de fournir des prévisions fiables.

À travers une architecture modulaire, cet observatoire de l'évolution des indices prix à la consommation permet une bonne gestion, un bon traitement, une bonne analyse et une bonne modélisation des données à des fins de prévision. Ces modules de visualisation et de prévision (ARIMA, SARIMA, HOLT-WINTERS, PROPHET) renforce la capacité de cet observatoire à servir d'outil d'aide à la décision et d'instrument d'information.

Les résultats empiriques issus de cet observatoire ont confirmé la faisabilité et l'utilité économique d'un tel projet. L'observatoire ainsi développé permet une visualisation interactive des séries d'indices, des classifications des postes selon leurs dynamiques, des détections d'anomalies et des prévisions. D'un point de vue économique, les analyses montrent une hausse structurelle des prix alimentaires et une forte saisonnalité liée aux transports et à l'énergie. Cela révèle encore l'importance d'une surveillance de la dynamique des prix dans l'économie et souligne l'intérêt stratégique de mobiliser les technologies d'apprentissage automatique (intelligence artificielle) dans l'analyse et l'éclairage de l'opinion publique et des autorités politiques.

La réalisation de cette étude a toutefois été contrainte à une difficulté majeure, l'accès difficile aux données primaires sur les prix unitaires des produits sur les plateformes de e-commerce. Cette difficulté a justifié le recours aux indices agrégés publiés par l'ANSTAT, permettant de garantir une cohérence avec les normes statistiques officielles (COICOP, IHPC).

Dans le but d'améliorer les travaux réalisés pour ce projet, nous invitons les autorités statistiques et les gestionnaires de plateforme de e-commerce à offrir une ouverture plus large sur les microdonnées de prix afin d'enrichir la granularité des analyses. Également, une extension de la couverture de l'observatoire à des données géographiques permettrait une analyse approfondie et ciblée des politiques à mettre en œuvre en matière de prix et d'inflation dans l'économie.

En définitive, cette étude montre que les données sur les prix, couplées à l'intelligence artificielle, peuvent devenir un levier important pour la transparence, la réactivité économique et la protection du pouvoir d'achat. Ce prototype d'observatoire, construit sous Streamlit, constitue une base prometteuse pour le développement d'un système national de suivi automatique des prix en Côte d'Ivoire.

Bibliographie

AFRISTAT. (2020). Projet de guide méthodologique pour l'élaboration des indices du coût de la construction pour les pays d'Afrique subsaharienne.

Arcade Ndoricimpa, 2017, Thresold effects of inflation on economic growth in Africa: evidence from a dynamic panel threshold regression approach.

Banque mondiale, Data Science for Public Policy, 2020.

BCEAO (2023). Bulletin de la politique monétaire et de l'inflation dans l'UEMOA.

BIERNAT Eric, LUTZ Michel, « Data Science fondamentaux et étude des cas Machine Learning avec Python et R ». Eyrolles. 2015. page 93.

Boyer, Philippe. (Mai 2014). Objectifs, méthodes, résultats et projets de l'observatoire de la formation des prix et des marges des produits alimentaires. Séance libre de l'Académie d'Agriculture de France du 21 mai 2014 : la formation des prix alimentaires, Académie d'Agriculture de France, PARIS, France.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. Wadsworth Int. Group, 37(15), 237-251.

Chucku Chuku, Simpasa Anthony and Oduor Jacob, 2018, Macroeconomic consequences of commodity price fluctuations in African economies

Diakalia Diarassouba, N'guessan Marie-Thérèse, Koffi Simplicie Yao, 2015, évaluation de l'inflation des prix des produits vivriers dans la commune de Yopougon.

DJABALLAH, Mohammed Abdelmoumen. (2021). Prédiction de la consommation d'énergie d'une maison intelligente à l'aide de l'apprentissage profond. Mémoire de maîtrise en Informatique.

Guide de l'IPC, FMI, 2004

Hastie, T., Tibshirani, R., et Friedman, J., The Elements of Statistical Learning : Data Mining, Inference, and Prediction, 2nd Edition, Springer, 2009.

Inflation Forecasting in Côte D'Ivoire: A Comparative Analysis of the Arima, Holt-Winters, and Lstm Models Koffi, Siméon, Aout 2022

ANSTAT Côte d'Ivoire. Note mensuelle sur l'IHPC.

INSEE, Glossaire économique, 2023

Iqbal et al., Machine Learning : Algorithms, Real-World Applications and Research Directions, SN Computer Science, 2021.

James, G., Witten, D., Hastie, T., et Tibshirani, R., An Introduction to Statistical Learning, Springer, 2013.

Jean, Éric. (2015). Les enjeux liés à la collecte de données en ligne : Le cas d'une recherche auprès de gestionnaires. La Revue des Sciences de Gestion, Direction et Gestion n° 272 – Épistémologie.

Kouassi, Adoumou Hugues. (2021). Prévion de l'inflation au Canada. Mémoire de maîtrise en économie. Québec, Canada.

Kristin J Forbes, 2018, Has the globalization changed the inflation process?, BIS Working Papers

Lallich, S., & Lenca, P. (2015). Indices de qualité en clustering. Présentation, Orange Labs, Issy Les Moulineaux, France.

Nations Unies, Division de statistique. (Année non spécifiée). Manuel de Classification des fonctions (COFOG, COICOP, COPNI, COPP). (Basé sur le Système de Comptabilité Nationale de 1993 et la Classification internationale type de l'éducation de 1997).

Objectifs, méthodes, résultats et projets de l'observatoire de la formation des prix et des marges des produits alimentaires, Philippe Boyer, octobre 2020

Observatoire des prix. (2024). Rapport annuel 2024.

Rakotomalala, Ricco. (8 janvier 2016). Tanagra Data Mining. (Texte associé ou revue).

Real-time monitoring of food price policy interventions during the first two years of COVID-19, Sarah Consoli, Juan José Egas Yerovi, Matteo Machiorlatti, Cristian Morales Opazo

Statbel. (2018). Le webscraping, la collecte et le traitement de données en ligne pour l'indice des prix à la consommation, SPF Économie Belgique.

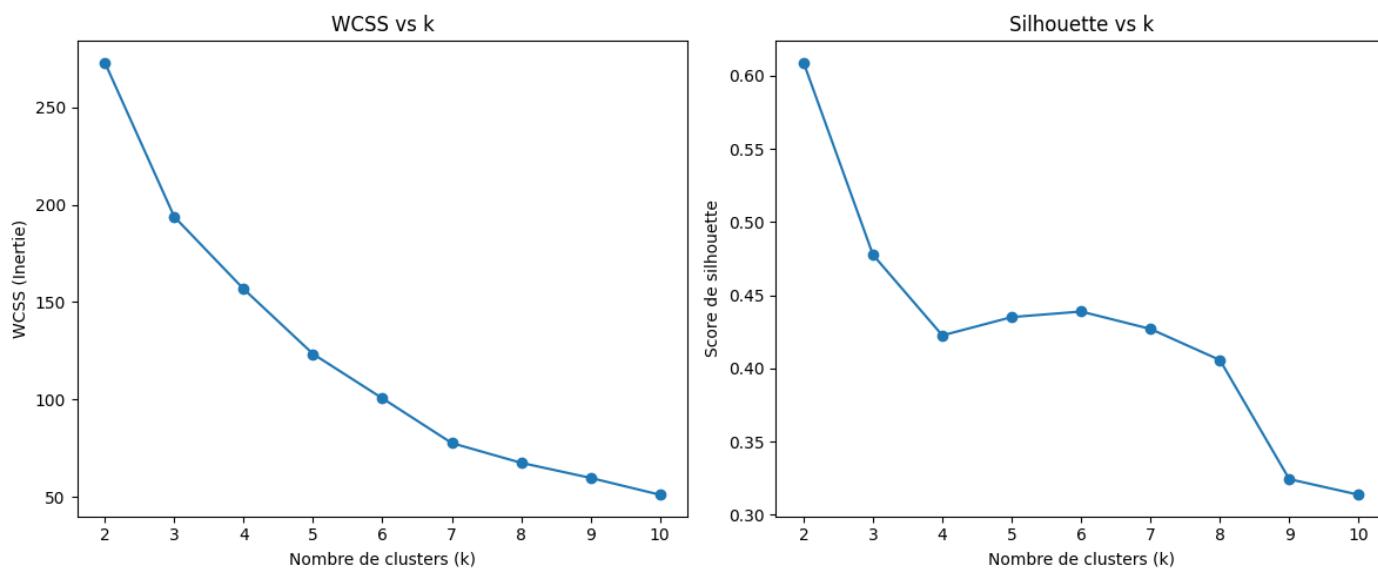
Till Wohlfarth. Machine-learning pour la prédiction des prix dans le secteur du tourisme en ligne. Analyse de données, Statistiques et Probabilités [physics.data-an]. Télécom ParisTech, 2013. Français. NNT : 2013ENST0090 tel-01310537

Wang, Xiaozhe & Smith-Miles, Kate & Hyndman, Rob. (2006). Characteristic-Based Clustering for Time Series Data. Data Min. Knowl. Discov.. 13. 335-364. 10.1007/s10618-005-0039-x.

Yakine & Tassnim Hala. (2023). Analyse de l'évolution des prix du pétrole : Étude comparative des modèles ARIMA et ARMA pour la prévision.

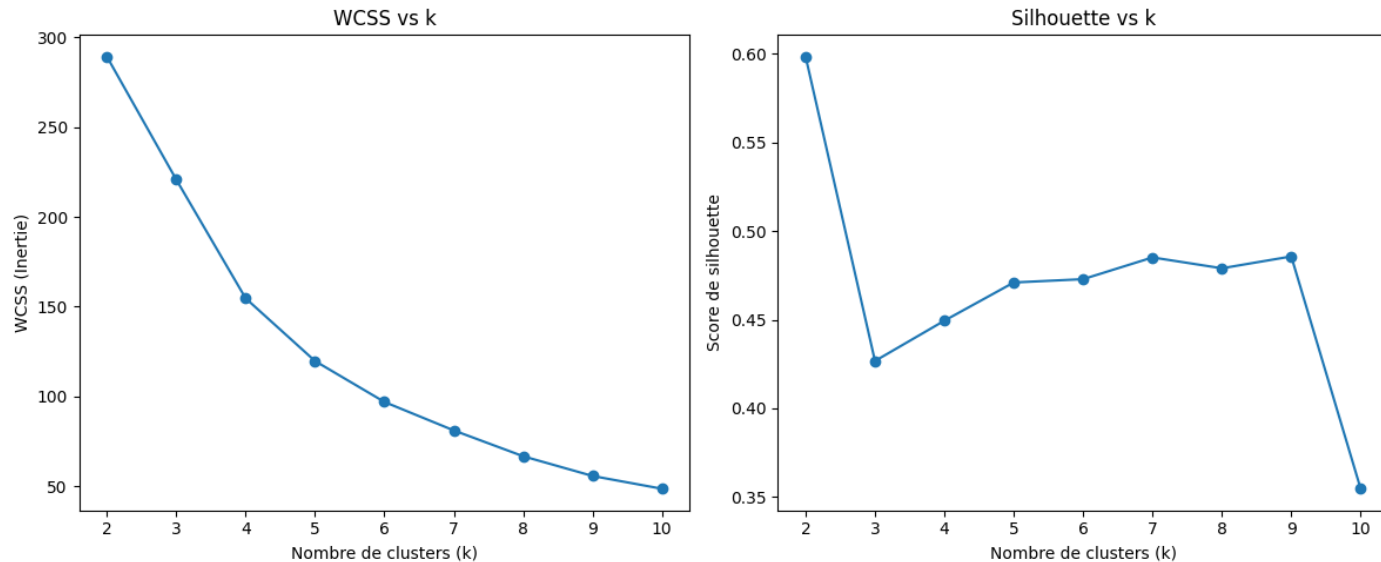
Annexes

Annexe 1 : Le coude et la silhouette pour la méthode Kmeans



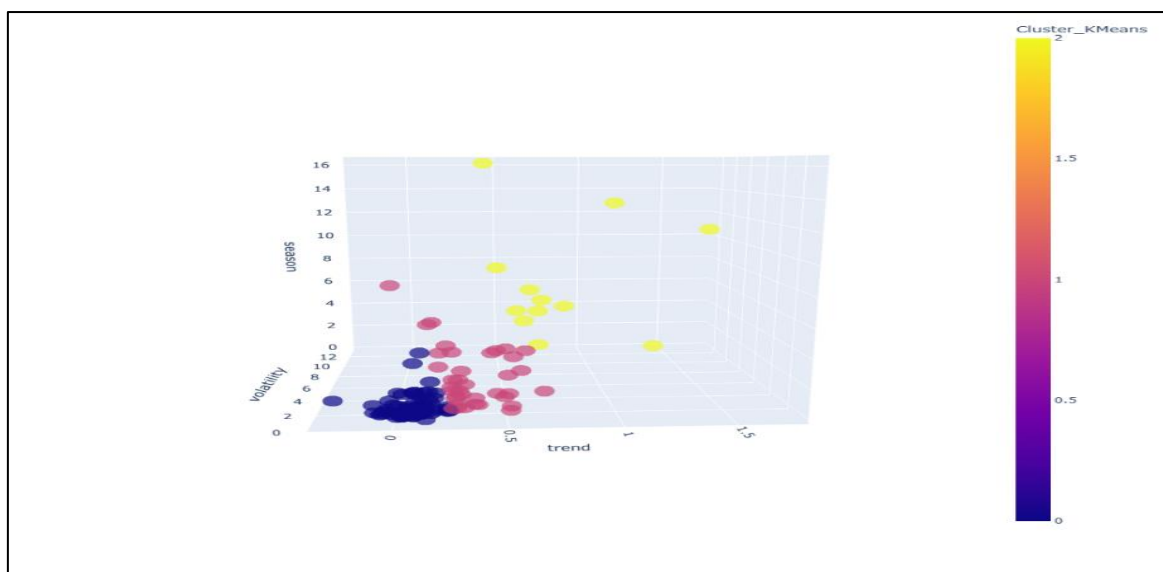
Source : Nos calculs, avec les données de l'ANStat

Annexe 2 : Le coude et la silhouette pour la méthode CAH



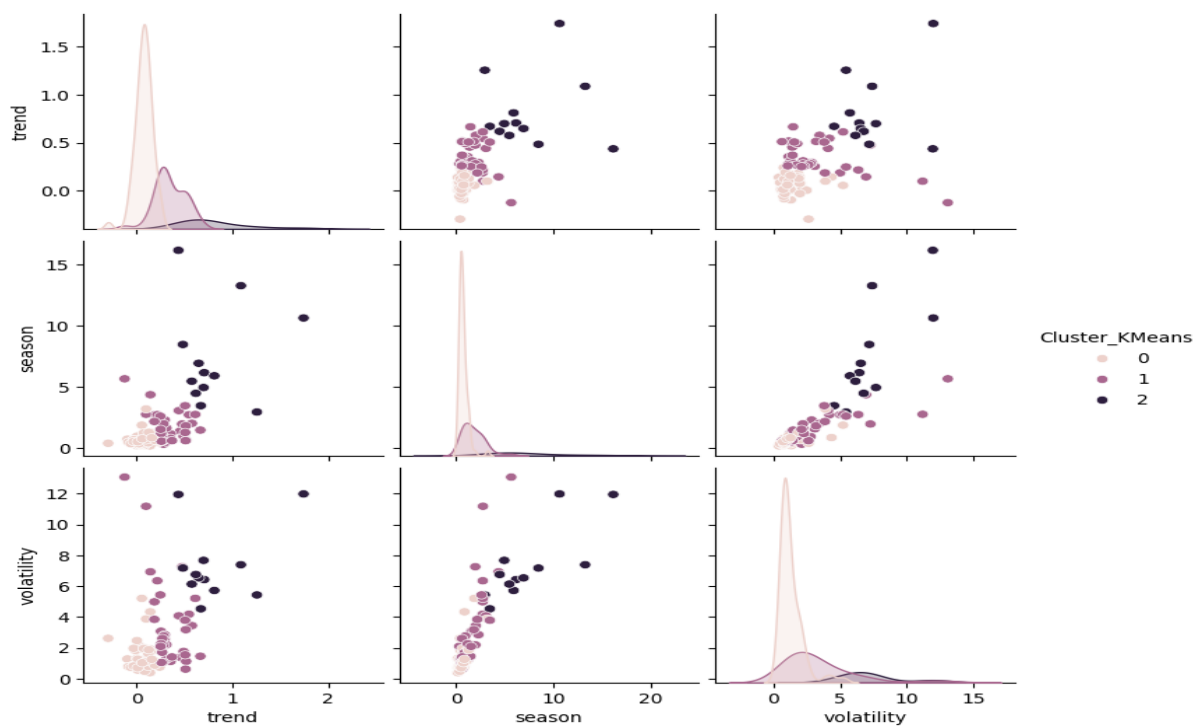
Source : Nos calculs, avec les données de l'ANStat

Annexe 3 : Représentation des clusters pour le Kmeans en 3D (Trend, Saisonnalité et Volatilité)



Source : Nos calculs, avec les données de l'ANStat

Annexe 4: Profil des clusters par trend, saisonnalité et volatilité



Source : Nos calculs, avec les données de l'ANStat

Tables des matières

Décharge.....	ii
Dédicace	iii
Remerciements	iv
Avant-propos	v
Sommaire	vi
Sigles et Abréviations.....	vii
Listes des tableaux	viii
Listes des graphiques – Figures et annexes.....	viii
Résumé.....	ix
Abstract	ix
Présentation de la structure	1
Introduction générale.....	2
Chapitre 1 : REVUE DE LITTERATURE.....	5
I. Définitions des concepts.....	5
II. Revue empirique de l'évolution des prix	6
III. Vue d'ensemble sur les observatoires de prix à la consommation	7
1. Au niveau mondial	7
2. En Afrique	8
IV. Revue de littérature sur les méthodes de collecte de prix	9
V. Méthodes de machine learning adaptées pour l'analyse des prix	11
1. Méthodes statistiques de prévision.....	11
2. Méthode d'apprentissage automatique	11
a. Apprentissage supervisé.....	11
b. Apprentissage non supervisé.....	12
c. Approches basées sur l'apprentissage profond (Deep Learning)	12

d. Comparaison des performances des modèles en Côte d'Ivoire	13
Chapitre 2 : Cadre méthodologique et Présentation des données	14
I. Présentation des données.....	14
1. Source de données	14
2. Variables de l'étude	15
II. Choix technologiques et justification	15
III. Approche méthodologique	17
1. La collecte de données par webscrapping	18
2. Traitement des données	18
IV. MODELISATION	19
1. Le modèle ARIMA (AutoRegressive Integrated Moving Average).....	19
2. Le modèle SARIMA (Seasonnality AutoRegressive Integrated Moving Average) .	19
3. Le modèle PROPHET	20
4. Critères d'évaluation des modèles de prévision.....	20
V. Classification.....	21
1. Variables utilisées pour la typologie.....	22
2. Algorithmes.....	22
a. Clustering K-means.....	22
b. Classification ascendante hiérarchique (CAH)	23
3. Critères de choix.....	24
a. Indice de Silhouette	24
b. Indice de Davies–Bouldin	24
c. Indice de Calinski–Harabasz	25
4. Choix du meilleur modèle	25
5. La méthode de HOLT-WINTERS	26
6. Construction de la plateforme de visualisation	26
Chapitre 3 : Analyses des Résultats et Discussions.....	27

I.	Stationnarité	27
II.	Architecture générale de l'application	28
1.	Exploration	29
2.	Corrélations	30
3.	État Base.....	31
4.	Choix	31
5.	Prévisions	32
6.	Classification	34
7.	Alertes	35
8.	Traitement.....	36
9.	Scraping.....	37
III.	ANALYSE DU MARCHÉ DE CONSOMMATION DE LA COTE D'IVOIRE	38
1.	La structure du marché de consommation.....	38
2.	Implications économiques.....	39
IV.	Perspectives d'évolution de l'application	40
1.	Extension aux autres produits.....	40
2.	Ajout de filtres géographiques	40
	Conclusion et recommandations	41
	Bibliographie.....	x
	Annexes	xiii
	Tables des matières.....	xv