

La régression linéaire et la prédiction de l'inflation monétaire: Application à l'intention de la Banque Centrale du Congo (RDC)

[Linear regression and prediction of monetary inflation: Application for the Central Bank of Congo (DRC)]

KABEYA TSHISEBA Cédric, Christ TSUNGU MIJIMBU, Glad LUFIMPU, and Christian MUABI KATENDA

Université Pédagogique Nationale (UPN), RD Congo

Copyright © 2023 ISSR Journals. This is an open access article distributed under the *Creative Commons Attribution License*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: With the advent of automatic learning methods and the exponential growth of computer power, several things are being facilitated, in particular the prediction of certain behaviors, made by banking establishments.

Although some prediction tools such as Excel are nowadays used by some of our establishments, we find that machine learning remains unused so far, even less with all its power, yet many advantages and opportunities present themselves to it. use.

We argue that elsewhere, experiments based on machine learning, in other words automatic learning, are more than topical, even more so in the banking sector. We are therefore going through this experience, to propose as an illustration and educational, for our local banking establishments, an activity of prediction of monetary inflation, using the power of artificial intelligence.

The prediction in question here will be made on the basis of a set of data collected from a few banks in the city-province of Kinshasa, and particular emphasis will be placed on general consumer price indices.

It is important to note that the prediction in question here goes as far as clearly specifying the causes of the inflation being analyzed, or predicted, of course on the basis of the different variations of the indicators.

The machine learning used here offers us several possibilities in terms of algorithms and models, but in the context of this work, we will only address a few, in particular Linear Regression, the Random Forest Regression algorithm or Radom Forest Regression, and the Regression model decision tree, will get we will get the best algorithm with respect to its score.

KEYWORDS: Machine Learning, Monetary Inflation and Artificial Intelligence.

RESUME: Avec l'avènement des méthodes d'apprentissage automatique et la croissance exponentielle de la puissance des ordinateurs, plusieurs choses se sont vues être facilitées, notamment la prédiction de certains comportements, faite par les établissements bancaires.

Bien que certains outils de prédictions comme Excel sont à ces jours utilisés par certains de nos établissements, nous constatons que l'apprentissage automatique reste jusqu'ici pas exploitée, moins encore avec toute sa puissance, pourtant des nombreux avantages et opportunités se présentent à son utilisation.

Nous avouons que sous d'autres cieux, les expériences basées sur le machine learning, autrement dit apprentissage automatique, sont plus que d'actualité, plus encore dans le secteur bancaire. Nous allons donc à travers cette expérience, proposer à titre d'illustration et pédagogique, pour nos établissements bancaires de la place, une activité de prédiction de l'inflation monétaire, usant de la puissance de l'intelligence artificielle.

La prédiction dont il est question ici sera faite sur base d'un jeu de données recueillies auprès de quelques banques de la ville province de Kinshasa, et un accent particulier sera mis sur les indices généraux de prix à la consommation.

Il importe de noter que la prédiction dont il est questions ici, va jusqu'à la précision de manière claire des causes de l'inflation encours d'analyse, ou prédite, bien sur base des différentes variations des indicateurs.

L'apprentissage automatique ici utilisée, nous offre plusieurs possibilités en terme d'algorithmes et modelés, mais dans le cadre de cet travail, nous n'allons aborder que quelques-uns, notamment la Régression Linéaire, l'algorithme de Forêt Aléatoire de Régression ou Radom Forest Régression, et l'arbre de décision modèle de Régression, desquels nous allons ressortir le meilleur algorithme au regard de son score.

MOTS-CLEFS: Apprentissage automatique, Inflation monétaire et Intelligence Artificielle.

1 INTRODUCTION

Une forte inflation peut provoquer le ralentissement de l'économie d'un pays d'où l'intérêt de mieux la contrôler. L'un des contrôles de l'inflation menés par la banque centrale du Congo (RDC) consiste à calculer et à prévoir l'inflation à l'aide d'indicateurs de l'indice des prix à la consommation (IPC), qui est effectivement l'un des indicateurs les plus couramment utilisés pour mesurer le taux d'inflation.

Sauf que, la plus part des Banques Centrales de nos pays Africains recourent encore à des prévisions utilisant des variables macroéconomiques comme prédicteurs présentant ainsi deux limitations sérieuses: d'une part, un nombre assez important de prédicteurs potentiellement informatifs, et d'autre part, la durée limitée des séries chronologiques disponibles, ce qui conduit à ce que l'on appelle « malédiction de la dimensionnalité ».

Les limitations ci-dessus peuvent potentiellement conduire à ce que l'on appelle le « problème de surparamétrisation », soit le résultat d'un petit nombre d'observations temporelles par rapport au nombre de variables explicatives.

Ces différents obstacles rencontrés par les techniques de prévision traditionnels encore utilisées par les établissements de la place hantent notre curiosité et nous incite à produire la réflexion dont il est question dans cet article.

Les approches pour résoudre le problème du surparamétrisation ont longtemps été une préoccupation du domaine de l'informatique connu sous le nom d'apprentissage automatique (ML). Au cours des dernières décennies, de nombreux modèles ML différents ont été créés : régression LASSO (Least Absolute Shrinkage and Selection Operator), Ridge regression, Principal Components Analysis, Decision Trees, Random Forest (RF), Boosting, Neural Networks, etc. Leur application a conduit à des avancées significatives et a permis de réaliser des résultats plus satisfaisants et fiables que les techniques classiques.

Les algorithmes d'apprentissage automatique effectuent la présélection des paramètres sur un sous-échantillon de données, mais pas uniquement sur la base de concepts théoriques, afin d'améliorer les résultats de prédiction. Cette présélection des variables explicatives peut être considérée comme une sorte d'hyperparamètre initial, dont dépend principalement la qualité de la prévision ce qui permet de contourner le problème de surparamétrisation (surapprentissage).

Dans cette expérience, nous proposons une prédiction à l'aide de l'indice des prix à la consommation (IPC) et la régressions linéaire, un algorithme du machine learning adopté ici au regard de son score, l'inflation au sein de notre Banque Centrale, à travers ses organes d'études et sur base de données mensuelles entre janvier 2012 et Février 2022.

2 NOTIONS ESSENTIELLES SUR L'INFLATION MONÉTAIRE

L'inflation correspond à une hausse continue du niveau général des prix à l'intérieur d'un espace donné. Elle correspond à une hausse du prix moyen de tous les biens et services. Il faut aussi que ce mouvement de hausse des prix soit durable, ainsi, lorsque tous les prix augmentent à une certaine date de 1% puis restent stables pendant plusieurs mois, on ne peut pas parler d'inflation. En revanche, s'ils augmentent ainsi tous les mois, on est bien en présence d'inflation.

2.1 MODÈLES DE PRÉVISION DE L'INFLATION

Actuellement, il existe différentes catégories de modèles de prévision de l'inflation: avec prévisions observées par le passé, avec variables macroéconomiques et avec variables basées sur le prix des actifs financiers. Trois types de modèles autorégressifs permettraient de prédire l'inflation. Le premier est le modèle univarié (AR), mais ses résultats sont trop simplistes pour décrire tous les facteurs influents de l'inflation. Le second modèle est le modèle de vecteur autorégressif (VAR). Il fait de meilleures prévisions, mais il a un problème de degrés de liberté. Enfin, il y a le modèle de facteurs dynamiques (FAVAR). L'avantage de ce dernier modèle est qu'il permet d'intégrer beaucoup d'informations.

2.2 LES INDICATEURS DE PRÉDICTION DE L'INFLATION

Il existe plusieurs indicateurs de prédiction de l'inflation monétaire tels que:

La croissance de la masse monétaire, l'écart entre les quantités offertes et demandées de monnaie, la masse monétaire, les prix extérieurs, le taux de change, l'écart de production, etc...

(MARTIN NYONGOLO [2]) révèle qu'en R.D.Congo les indicateurs d'inflation les plus pertinents sont tant d'origine monétaire que réelle. Ce sont donc les chocs d'origine monétaire qui affectent beaucoup plus l'économie congolaise. Cette conclusion est validée par les résultats de l'analyse de la décomposition de variance qui placent au premier plan la consommation publique, le taux de change, et la masse monétaire comme des indicateurs les plus pertinents.

Dans ce travail nous nous sommes focalisés sur l'indice de prix à la consommation qui est l'indicateur le plus influent dans la prédiction de l'inflation monétaire en RDC.

3 LES MÉTHODES CLASSIQUES DE PRÉDICTION ET LEURS LIMITES

Avec l'économétrie, la méthode consiste à utiliser des données passées pour établir des relations statistiques qui à leur tour peuvent être utilisées pour prévoir des futurs possibles.

La plupart des méthodes économétriques espèrent faciliter l'interprétation et l'inférence statistique, car les méthodes économétriques conventionnelles reposent sur des hypothèses et des relations linéaires, comme la fameuse hypothèse de linéarité. "Vous voulez tracer du passé vers l'avenir.

Cette approche fonctionne bien dans certains cas, mais s'avère également être une limitation: la plupart des modèles économétriques conventionnels ne peuvent pas gérer les mégadonnées ou les relations compliquées. Si vous avez de nombreux prédicteurs ou une relation compliquée, les méthodes économétriques échoueront. Et c'est une limitation sérieuse à l'ère du big data et dans de nombreux cas importants.

4 PRÉDICTION À L'AIDE DES MÉTHODES D'APPRENTISSAGE AUTOMATIQUE

La banque centrale du Congo utilise des méthodes économétriques classiques dont les limites ont été énumérées dans les précédents points. Pour y remédier nous proposons une solution se basant sur les techniques d'apprentissage automatique.

4.1 APPRENTISSAGE AUTOMATIQUE

L'apprentissage automatique ou Machine Learning (ML) est une branche de l'intelligence artificielle qui consiste à programmer des algorithmes permettant d'apprendre automatiquement à partir des données et d'expériences passées ou par interaction avec l'environnement. Ce qui rend l'apprentissage machine vraiment utile est le fait que l'algorithme peut "apprendre" et adapter ses résultats en fonction de nouvelles données sans aucune programmation à priori.

Il existe plusieurs façons d'apprendre automatiquement à partir des données dépendamment des problèmes à résoudre et des données disponibles. La figure X donne un sommaire des types d'apprentissage automatique les plus connus.

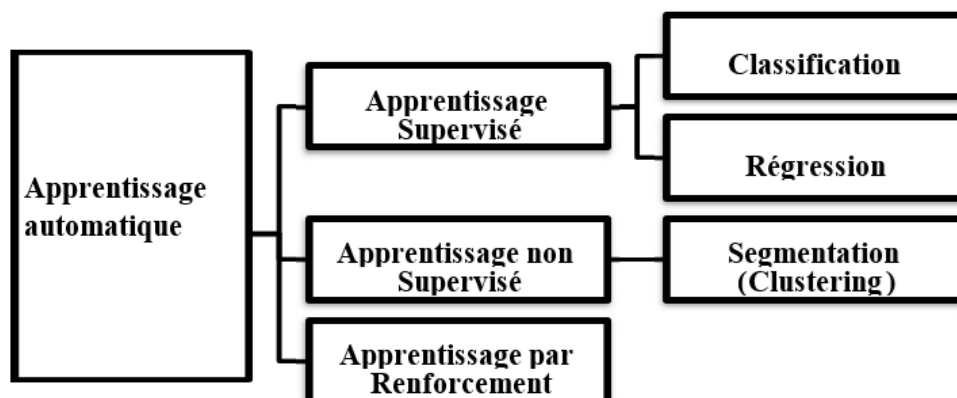


Fig. 1. Les grandes classes d'apprentissage automatique

4.2 APPLICATION DU MODÈLE

Il existe de nombreux outils permettant d'effectuer de l'apprentissage statistique. Dans ce travail, nous avons utilisé le langage python avec l'IDE Anaconda, Excel pour la création de notre dataset.

Pour mieux apprendre, le modèle a besoin de données avec lesquelles nous allons passer tous les tests possibles pour mettre en évidence les différents comportements et de bien capter les phénomènes sous-jacents de façon plus technique.

C'est l'étape la plus importante en apprentissage automatique, car celui-ci dépend fortement des données. Plus vous avez des données de qualité mieux seront vos prédictions.

Pour la prédiction de l'indice général, les données ont été extraites dans les rapports annuels de la Banque centrale du Congo de 2012 à 2022.

Variations mensuelles de l'indice des prix moyens de détail calculées par la B.C.C. (par groupe d'articles)														
ANNEE	MOIS	PRODUIT ALI	BOISSONS A	ARTICLES HA	LOGEMENT	MEUBLES, A	SANTE	TRANSPORT	COMMUNIC	LOISIR ET CU	ENSEIGNEM	RESTAURAN	AUTRES BIEN	INDICE GENERAL
2012	JANVIER	4,83	4,18	4,91	3,44	4,93	6,14	2,24	2,06	2,56	1,11	4,60	2,97	4,25
	FEVRIER	0,13	0,08	0,11	0,09	0,05	0,16	0,09	0,40	0,21	0,00	0,02	0,04	0,11
	MARS	0,18	0,08	0,21	0,36	0,08	0,12	0,95	0,08	0,08	0,00	0,14	0,09	0,24
	AVRIL	0,12	0,05	0,04	0,10	0,04	0,04	0,84	0,00	0,02	-0,01	0,02	0,04	0,15
	MAI	0,04	0,02	0,02	0,01	0,00	0,04	0,01	0,00	0,01	0,00	0,01	0,03	0,03
	JUIN	0,13	0,07	0,09	0,06	0,02	0,15	0,05	0,01	0,04	0,03	0,05	0,09	0,10
	JUILLET	0,12	0,03	0,08	0,08	0,06	0,09	0,05	0,00	0,04	0,02	1,00	0,08	0,10
	AOUT	0,19	0,05	0,14	0,19	0,07	0,11	0,09	0,00	0,04	0,17	0,10	0,14	0,17
	SEPTEMBRE	0,11	0,03	0,09	0,11	0,04	0,07	0,06	0,00	0,05	0,10	0,06	0,08	0,10
	OCTOBRE	0,07	0,01	0,01	0,03	0,01	0,03	-0,10	0,00	0,02	0,01	0,05	0,03	0,04
	NOVEMBRE	0,13	0,67	0,10	0,08	0,05	0,08	0,08	0,01	0,04	0,08	0,11	0,10	0,12
	DECEMBRE	0,11	0,07	0,11	0,07	0,05	0,07	0,06	0,00	0,06	0,07	0,09	0,09	0,11
	JANVIER	0,15	0,10	0,07	0,13	0,06	0,08	0,11	0,08	0,05	0,03	0,13	0,07	0,13
	FEVRIER	0,15	0,36	0,10	0,13	0,07	0,08	0,09	0,05	0,09	0,08	0,15	0,09	0,13
	MARS	0,12	0,08	0,09	0,11	0,05	0,07	0,07	0,03	0,06	0,06	0,13	0,09	0,10

Fig. 2. Dataset brut avec MS Excel

Source: Rapport annuel de la BCC de 2012-2022

4.2.1 IMPORTATION DES BIBLIOTHÈQUES

Une bibliothèque est un ensemble de fonctions et de routines que l'on peut réutiliser facilement. Python est un langage de programmation open source qui possède de nombreuses librairies.

```
Entrée [1]: import numpy as np
import pandas as pd

from sklearn.impute import SimpleImputer
from sklearn import linear_model
from sklearn import tree
from sklearn.ensemble import RandomForestRegressor

from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import r2_score

import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
```

Fig. 3. Importation des bibliothèques

Source: nous-même

4.2.2 PRÉPARATION DES DONNÉES (NETTOYAGE DES DONNÉES)

Aussi appelée « prétraitement des données », cette étape est mise en œuvre avant l'implémentation d'un algorithme d'apprentissage; car les données brutes sont souvent bruitées, peu fiables et incomplètes. Pour réaliser cette étape nous allons commencer par importer les données:

4.2.2.1 IMPORTATION DES DONNÉES

```
Entrée [2]: df = pd.read_excel('../dataset_indice_prix_moyens.xlsx', header=0)
```

Fig. 4. Importation des données dans jupyter Notebook

Source : Nous-mêmes

4.2.2.2 SÉLECTION DES VARIABLES

Les données importées étant brutes, elles doivent subir un toilettage ceci implique que nous devons sélectionner les variables explicatives qui influent sur la variable expliquée.

Afin de mieux établir la corrélation entre variables, il est mieux de récupérer les variables explicatives, étudier leurs corrélations et enfin supprimer celles qui ne sont pas fortement corrélées.

```
Entrée [4]: variables = list(df.columns[:-1]) #ici nous recuperons toutes Les variables sauf la variable explicative
```

```
variables
```

```
Out[4]: ['ANNEE',
        'MOIS',
        'PROD_AL_ BOIS_NON_ALCOL',
        'BOISSONS ALCOLISEES, TABACS ET STUFIANTS',
        'ARTICLES HABILLEMENT ET CHAUSSURES',
        'LOGEMENT EAU, GAZ, ELECTRICITE ET AUTRES COMBUSTIBLE',
        'MEUBLES, ARTICLE DE MENAGE ET ENTRETIEN COURANT DU FOYER',
        'SANTE',
        'TRANSPORTS',
        'COMMUNICATIONS',
        'LOISIR ET CULTURE',
        'ENSEIGNEMENT',
        'RESTAURANTS ET HOTELS',
        'AUTRES BIENS ET SERVICES']
```

Fig. 5. Sélection des variables

Source : Nous-mêmes

4.2.2.3 ETUDE DE CORRÉLATION ENTRE VARIABLES

Puisque nous avons une variable cible qui est dans ce cas, la variable « INDICE_GENERAL », nous allons vérifier si les autres variables parviennent à expliquer la variable cible. La figure ci-dessous va reprendre la matrice corrélation entre les différentes variables.

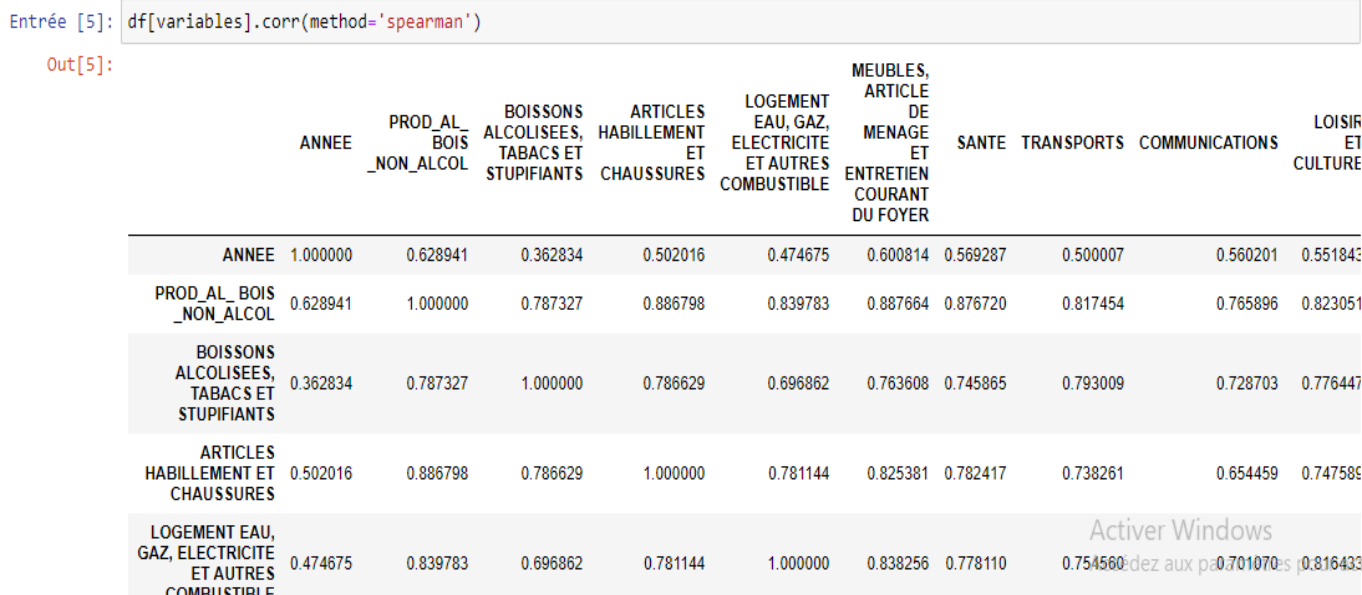


Fig. 6. Etude de la corrélation entre variables

Source : Nous-mêmes

En utilisant la méthode de Spearman pour étudier la corrélation entre variables nous avons constaté que les variables mois et année ne sont pas corrélées par conséquent elles doivent être supprimées pour ne garder que celles qui influent sur la prédiction de la variable cible.

En ce qui concerne notre jeu de données, nous n’avons pas eu des données manquantes, ni des valeurs aberrantes, grâce notamment à la data augmentation (augmentation de la qualité des données, en français). Mais seulement, il y a quelques attributs non-pertinents pour l’apprentissage et des variables catégorielles que nous devons transformer en numériques (encodage), format propice au développement des modèles de machine learning.

De ce fait, nous allons imputer les variables année et mois parce qu’elles ne sont pas pertinentes pour l’apprentissage (Absence de variance).

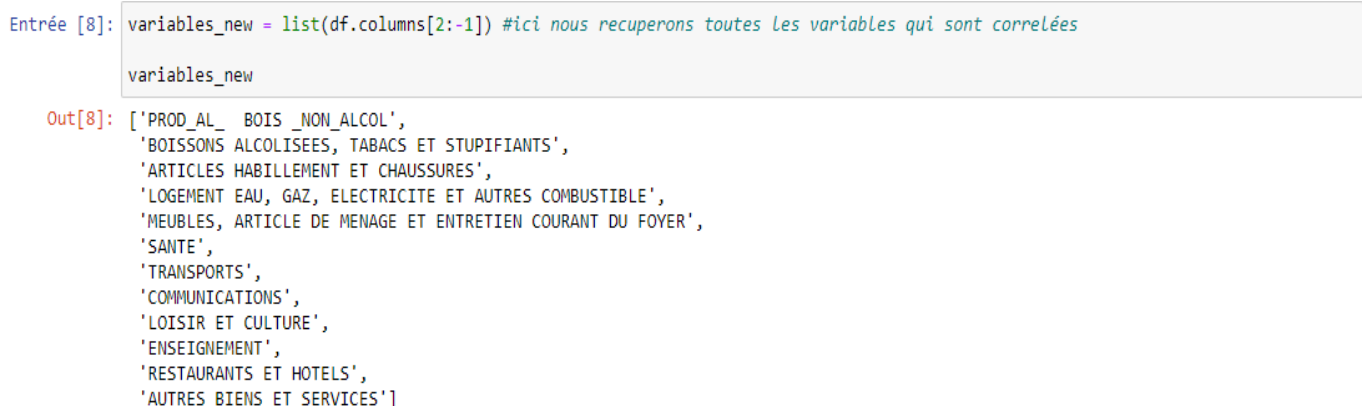


Fig. 7. Récupération des variables corrélées

Source : Nous-mêmes

4.2.3 CRÉATION DU JEU D'ENTRAÎNEMENT ET DU TEST

Nous avons subdivisé notre jeu en deux parties: 70% pour l'entraînement et 30% pour le test. Pour ce faire, la variable cible a été imputée pour mieux expérimenter la performance de nos futures prédictions.

Nous avons défini quatre variables; `x_train`, `y_train`, `x_test` et `y_test`. Ces quatre variables seront utiles lors de l'entraînement et le test des différentes techniques que nous aurions retenues.

```
Entrée [15]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.3, random_state = 0)
```

Fig. 8. Création du Train et Test set

Source : Nous-mêmes

4.3 CHOIX D'ALGORITHMES, MODÉLISATION ET RÉSULTATS EXPÉRIMENTAUX

L'exploration de données nous a permis de les sélectionner, analyser, nettoyer et normaliser. Etant dans un état utilisable, dans cette section nous allons passer à l'étape de la modélisation, qui comportera deux phases.

- La phase descriptive: qui consiste à entraîner le modèle
- La phase prédictive: qui consiste à prédire la cible partant des features entraînés.

4.3.1 CHOIX D'ALGORITHME

Avant d'aborder ces deux phases le choix de (s) l'algorithme (s) à utiliser pour la création des modèles revêt un caractère obligatoire.

Partant des données en notre possession nous constatons que notre cible est connue et continue donc il s'agit de l'apprentissage supervisé de régression. D'où nous avons retenu principalement trois algorithmes:

- La Régression Linéaire,
- L'algorithme de Forêt Aléatoire de Régression ou Radom Forest Régression, en anglais,
- L'arbre de décision modèle de Régression.

L'implémentation successive de ces trois algorithmes nous permettra de choisir le meilleur pour notre cas en se basant sur la précision et la durée d'apprentissage.

4.3.2 MODÉLISATION

Après avoir créé nos modèles dans cette partie nous allons comparer la précision et le taux d'erreur de chacun d'eux, ensuite en fonction des nouvelles variables nous allons évaluer la précision de chaque modèle.

4.3.2.1 COMPARAISON DES MODÈLES

4.3.2.1.1 EN FONCTION DE LA PRÉCISION

```
Entrée [26]: print("La précision du modèle :\n Regression multiple est de ", np.round(precision_reg * 100, 2), "%",
          "\n Forêt aléatoire Regression est de ", np.round(score * 100, 2), "%",
          "\n Arbre de decision Regression est de", np.round(precision_arb * 100, 2), "%"
          )
```

```
La précision du modèle :
Regression multiple est de 100.0 %
Forêt aléatoire Regression est de 94.61 %
Arbre de decision Regression est de 96.78 %
```

Fig. 9. Comparaison des différents modèles

Source : Nous-mêmes

Dans la figure ci-dessus nous constatons que la régression linéaire a une précision de 100%, ensuite vient l'arbre de décision avec une précision 96,78%, le forêt aléatoire en dernier avec 94,61%.

4.3.2.1.2 EN FONCTION DU TAUX D'ERREUR

```
Entrée [28]: df_compare.plot(kind='bar')
```

```
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x26b11cbf288>
```

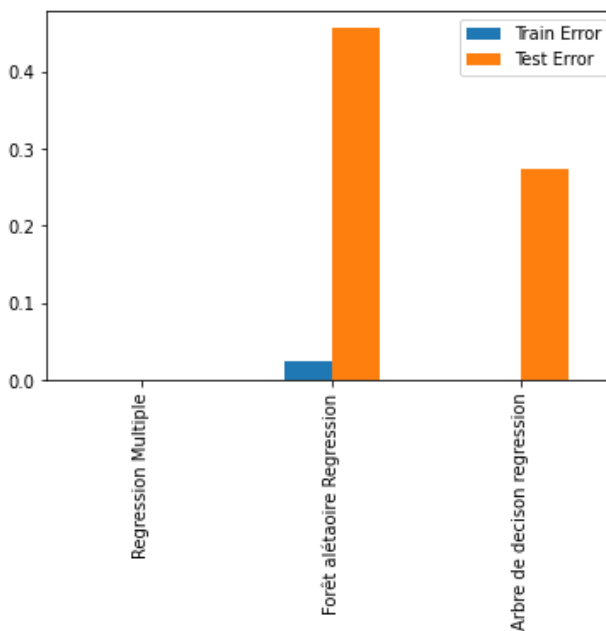


Fig. 10. visualisation du taux d'erreur des modèles

Source : Nous-mêmes

4.3.2.1.3 EN FONCTION DE LA PRÉDICTION DES NOUVELLES DONNÉES

```
Entrée [29]: INPUT = [[2012, 'JANVIER', 4.83, 4.18, 4.91, 3.44, 4.93, 6.14, 2.24, 2.06, 2.56, 1.11, 4.60, 2.97],  
                    [2022, 'FEVRIER', 0.45, 0.12, 0.14, 0.58, 0.25, 0.21, 0.33, 1.32, 0.25, 0.09, 0.29, 0.29]]
```

```
INPUT  
  
df_new = pd.DataFrame(data = INPUT, columns=variables)  
  
#Normalisation des données  
new_X = df_new[variables_new]  
  
new_X = imputer_model.transform(new_X)  
  
new_X
```

```
Out[29]: array([[4.83, 4.18, 4.91, 3.44, 4.93, 6.14, 2.24, 2.06, 2.56, 1.11, 4.6 ,  
                2.97],  
               [0.45, 0.12, 0.14, 0.58, 0.25, 0.21, 0.33, 1.32, 0.25, 0.09, 0.29,  
                0.29]])
```

Fig. 11. Chargement des nouvelles données

Source : Nous-mêmes

La figure ci-dessus nous permet de mieux visualiser le taux d'erreur que peuvent avoir nos algorithmes. Nous constatons que la régression multiple a une précision de 100%, c'est-à-dire pas d'erreur dans l'apprentissage ni dans le test alors que l'arbre de décision est précis dans l'apprentissage mais présente un taux conséquent d'erreur au test de son côté le forêt aléatoire a un pourcentage faible d'erreur dans l'apprentissage et un taux d'erreur assez élevé au test.

Tous ces éclaircissements nous permettent de conclure que la régression linéaire multiple est le mieux adapté pour la prédiction de l'inflation comparativement aux algorithmes forêt aléatoire et arbre de décision.

5 CONCLUSION

Il a été question dans cet article de prédire l'inflation monétaire par les techniques d'intelligence artificielle et nous pensons avoir atteint l'objectif assigné, soit celui de mener une expérience de prédiction à l'aide du machine learning et de ses outils, de l'inflation monétaire dans notre pays, déterminant ainsi sur base de l'indice des prix à la consommation, le comportement futur de notre système monétaire ainsi que les probables causes de sa perturbation.

Afin d'atteindre l'objectif ci-dessus assigné, plusieurs algorithmes ont été testés, mais au regard des résultats obtenus de ces tests, avons opté pour la régression linéaire qui nous a donné une précision de 100%.

Il est important de noter ici à titre conclusive, que cette réflexion est faite dans un contexte pédagogique pour les établissements bancaires de la place, particulièrement pour la Banque Centrale du Congo (RDC), qui une fois s'en approprie, a plus de chance de mieux pouvoir prédire de manière plus efficace, toute inflation monétaire susceptible de menacer le pays, d'en comprendre les causes et mieux de l'éviter.

REFERENCES

- [1] Jean – Yves CAPUL et Olivier GARNIER, Dictionnaire économique et des sciences sociale, Hatier, avril 1999, p.
- [2] Martin Nyongolo. INCIDENCE DE L'INFLATION SUR LES RECETTES FISCALES EN RDC DE 2005 A 2014. CAHIERS DU CERUKI, NOUVELLE SERIE, CERUKI, 2015, pp.275-285. Ffhal 01235985.
- [3] Stock, J. H. et Watson, M. W. (2011). Modèles à facteurs dynamiques. Dans: M.P. Clements et D.F. Hendry, éd. Le manuel d'Oxford de prévision économique. Presses de l'Université d'Oxford.
- [4] Hu, Junyan; Niu, Hanlin; Carrasco, Joaquin; Lennox, Barry; Arvin, Farshad (2020). «Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning». IEEE Transactions on Vehicular Technology. 69 (12): 14413–14423. doi: 10.1109/tvt.2020.3034800. ISSN 0018-9545. S2CID 228989788. Archived from the original on 2023-04-16. Retrieved 2023-04-16.
- [5] Yoosefzadeh-Najafabadi, Mohsen; Hugh, Earl; Tulpan, Dan; Sulik, John; Eskandari, Milad (2021). «Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean?». Front. Plant Sci. 11: 624273. doi: 10.3389/fpls.2020.624273. PMC 7835636. PMID 33510761.
- [6] Jump up to: a b c Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2.
- [7] Machine learning and pattern recognition «can be viewed as two facets of the same field».