Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (51-57), Month: September - December 2023, Available at: www.noveltyjournals.com

Prediction of Monetary Inflation in DR Congo: Comparative Study of Some Models Based on Machine Learning

[1]Christ TSUNGU MIJIMBU, [2]Christian MUABI KATENDA,
[3]Professor KABEYA TSHISEBA Cédric

[1,2] Assistants at DRC Pedagogic National University / Faculty of Sciences / Department of Mathematics and Computing Science

[3]Associate Professor at DRC Pedagogic National University / Faculty of Sciences / Department of Mathematics and Computing Science

Associate Professor at the Catholic University of Congo / Faculty of Computer Sciences

DOI: https://doi.org/10.5281/zenodo.10082547

Published Date: 08-November-2023

Abstract: **With a view to better economic growth, the Democratic Republic of Congo has been committed for some time through its Central Bank, among other things, to the path of major monetary inflation forecasts.**

In this article, we want to contribute to the above approach, by embarking on a comparative study between some monetary inflation forecast models based on machine learning, namely multiple regression, random forest and the tree of decision, with a view to identifying the best of these models applicable to this type of forecast, and with this type of data.

Keywords: **monetary inflation, artificial intelligence, machine learning models.**

1. INTRODUCTION

Guaranteeing a strong economy for a country also means planning for major monetary inflations through its banking institutions, something which is being done by the DR Congo, and which we want to support in this article by proposing beyond the traditional models, a much better model based on artificial intelligence, precisely machine learning.

Machine Learning is a concept of artificial intelligence including within it several learning models, applicable depending on the type of analysis carried out, and the data constituting the dataset taken into account.

Most authors, both nationally and internationally, addressing the issue of monetary inflation concerning our country, choose traditional models, ignoring what the models currently used in other countries, based on current technologies such as machine learning. On the other hand, at a slightly broader level, we note that the few authors who address this question using said current methods, in particular those based on machine learning, choose their models without going through a comparison of scores, which can allow them to opt for the best.

We have therefore in this article, based on the consumer price index (CPI) and a database containing information between January and February 2012, proposed a predictive study, using these three models in a comparative manner. below, with the idea of opting for the model which will offer us the best score, with a view to a more relevant prediction of monetary inflation.

It is important to point out that the data used here is very old, but still served us by allowing us to carry out this comparative study, and that the results will always be the same when they are updated.

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (51-57), Month: September - December 2023, Available at: www.noveltyjournals.com

2. MONETARY INFLATION AND MACHINE LEARNING

2.1 Monetary inflation

Inflation corresponds to a continuous increase in the general price level within a given area. It corresponds to an increase in the average price of all goods and services. This upward movement in prices must also be sustainable, so when all prices increase on a certain date by 1% then remain stable for several months, we cannot speak of inflation. On the other hand, if they increase like this every month, we are indeed in the presence of inflation. (Jean – Yves CAPUL & al., [1])

2.1.1 Inflation forecasting models

Currently, there are different categories of inflation forecasting models: with forecasts observed in the past, with macroeconomic variables and with variables based on the price of financial assets. Three types of autoregressive models would make it possible to predict inflation. The first is the univariate (AR) model, but its results are too simplistic to describe all the influential factors of inflation. The second model is the vector autoregressive (VAR) model. It makes better predictions, but it has a degrees of freedom problem. Finally, there is the dynamic factors model (FAVAR). The advantage of this last model is that it allows a lot of information to be integrated.

2.1.2 Inflation prediction indicators

There are several indicators for predicting monetary inflation such as: The growth of the money supply, the gap between the quantities supplied and demanded of money, the money supply, external prices, the exchange rate, the production gap, etc.

(MARTIN NYONGOLO [2]) reveals that in DR Congo the most relevant inflation indicators are of both monetary and real origin. It is therefore shocks of monetary origin that affect the Congolese economy much more. This conclusion is validated by the results of the variance decomposition analysis which place public consumption, the exchange rate, and the money supply at the forefront as the most relevant indicators.

In this work we focused on the consumer price index which is the most influential indicator in predicting monetary inflation in the DRC.

2.2 Machine learning

Automatic learning or Machine Learning (ML) is a branch of artificial intelligence which consists of programming algorithms to learn automatically from data and past experiences or through interaction with the environment. What makes machine learning truly useful is the fact that the algorithm can “learn” and adapt its results based on new data without any a priori programming.

There are several ways to automatically learn from data depending on the problems to be solved and the data available. Figure X gives a summary of the most common types of machine learning.

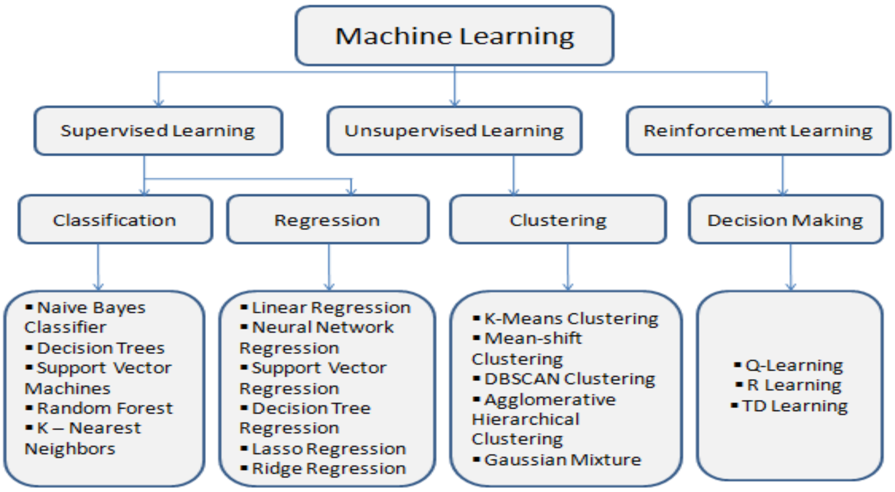

Figure n°1 The main classes of machine learning

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (51-57), Month: September - December 2023, Available at: www.noveltyjournals.com

As the figure above indicates, there are three main classes of machine learning, each of which is used according to the problem and need.

In this work, the objective is to predict monetary inflation based on time series and historical data. As the target to be predicted is already known in advance, supervised learning techniques make it possible to build models from learning or training examples whose behavior or response is known. These models can then be used in different applications, such as prediction or classification.

2.3 Comparative studies

In this part of this work we will implement on the basis of the same Dataset, under Python, three systems based on the three models mentioned above, then study the scores and the results produced by these models.

2.3.1 Some prerequisites for machine learning

2.3.1.1 Preparing the dataset

It is important to specify here that for the prediction of the general index, the data were extracted from the annual reports of the Central Bank of Congo from 2012 to 2022.

ANNEE	MOIS	PRODUIT ALI	BOISSONS A	ARTICLES HA	LOGEMENT E	MEUBLES, A	SANTE	TRANSPORTS	COMMUNIC	LOISIR ET CU	ENSEIGNEM	RESTAURANT	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

Feuil1 Feuil2

Figure n° 2. Raw dataset with MS Excel(Source: BCC Annual Report 2012-2022)

2.3.1.2 Importing libraries

A library is a set of functions and routines that can be easily reused. Python is an open source programming language that has many libraries.

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
```

Figure n°3 Importing libraries Source: ourselves

2.3.1.3 Data preparation (data cleaning)

Also called “data preprocessing”, this step is implemented before the implementation of a learning algorithm; because raw data is often noisy, unreliable and incomplete. To carry out this step we will start by importing the data:

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (51-57), Month: September - December 2023, Available at: www.noveltyjournals.com

- **Importing data**

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

Figure n° 4 Importing data into Jupyter Notebook Source ourselves

- **Variable selection**

The imported data being raw, they must undergo grooming, which implies that we must select the explanatory variables which influence the explained variable.

In order to better establish the correlation between variables, it is better to recover the explanatory variables, study their correlations and finally delete those which are not strongly correlated.

```
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 STUPIFIANTS',
         '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']
```

Figure n° 5 Selection of source variables: Ourselves

- **Correlation study between variables**

Since we have a target variable which is in this case, the variable "GENERAL_INDICE", we will check if the other variables manage to explain the target variable. The figure below will show the correlation matrix between the different variables.

```
Entrée [5]: df[variables].corr(method='spearman')
```

Out[5]:

	ANNEE	PROD_AL_ BOIS _NON_ALCOL	BOISSONS ALCOLISEES, TABACS ET STUPIFIANTS	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
ANNEE	1.000000	0.628941	0.362834	0.502016	0.474675	0.600814	0.569287	0.500007	0.560201	0.551843
PROD_AL_ BOIS _NON_ALCOL	0.628941	1.000000	0.787327	0.886798	0.839783	0.887664	0.876720	0.817454	0.765896	0.823051
BOISSONS ALCOLISEES, TABACS ET STUPIFIANTS	0.362834	0.787327	1.000000	0.786629	0.696862	0.763608	0.745865	0.793009	0.728703	0.776447
ARTICLES HABILLEMENT ET CHAUSSURES	0.502016	0.886798	0.786629	1.000000	0.781144	0.825381	0.782417	0.738261	0.654459	0.747589
LOGEMENT EAU, GAZ, ELECTRICITE ET AUTRES COMBUSTIBLE	0.474675	0.839783	0.696862	0.781144	1.000000	0.838256	0.778110	0.754560	0.701070	0.816433

Figure n° 6 Study of the correlation between source variables: ourselves

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering

Vol. 10, Issue 3, pp: (51-57), Month: September - December 2023, Available at: www.noveltyjournals.com

Using Spearman's method to study the correlation between variables we found that the month and year variables are not correlated therefore they must be deleted to keep only those which influence the prediction of the target variable.

Regarding our dataset, we did not have any missing data or outliers, thanks in particular to data augmentation (increase in data quality, in French). But only, there are some attributes that are not relevant for learning and categorical variables that we must transform into numeric (encoding), a format conducive to the development of machine learning models.

Therefore, we will impute the year and month variables because they are not relevant for learning (Absence of variance).

```
Entrée [5]: df[variables].corr(method='spearman')
```

Out[5]:

	ANNEE	PROD_AL_BOIS _NON_ALCOL	BOISSONS ALCOLISEES, TABACS ET STUPIFIANTS	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
ANNEE	1.000000	0.628941	0.362834	0.502016	0.474675	0.600814	0.569287	0.500007	0.560201	0.551843
PROD_AL_BOIS _NON_ALCOL	0.628941	1.000000	0.787327	0.886798	0.839783	0.887664	0.876720	0.817454	0.765896	0.823051
BOISSONS ALCOLISEES, TABACS ET STUPIFIANTS	0.362834	0.787327	1.000000	0.786629	0.696862	0.763608	0.745865	0.793009	0.728703	0.776447
ARTICLES HABILLEMENT ET CHAUSSURES	0.502016	0.886798	0.786629	1.000000	0.781144	0.825381	0.782417	0.738261	0.654459	0.747589
LOGEMENT EAU, GAZ, ELECTRICITE ET AUTRES COMBUSTIBLE	0.474675	0.839783	0.696862	0.781144	1.000000	0.838256	0.778110	0.754560	0.701070	0.816433

Figure n° 7 Recovery of correlated variables source: Ourselves

2.3.1.4 Creation of the training game and the test

We have subdivided our game into two parts: 70% for training and 30% for testing. To do this, the target variable was imputed to better test the performance of our future predictions.

We defined four variables; x_train, y_train, x_test and y_test. These four variables will be useful during training and testing of the different techniques that we have chosen.

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

Figure n° 8 Creation of the Train and Test set (Source: Ourselves)

2.3.1.5 Comparison itself

- **From the precision point of view**

```
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 %
```

Figure n° 9 comparison of the different source models: Ourselves

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (51-57), Month: September - December 2023, Available at: www.noveltyjournals.com

In the figure above we see that the linear regression has an accuracy of 100%, next comes the decision tree with an accuracy of 96.78%, the random drill last with 94.61%.

- **from the error rate point of view**

```
df_compare.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x26b11cbf288>
```

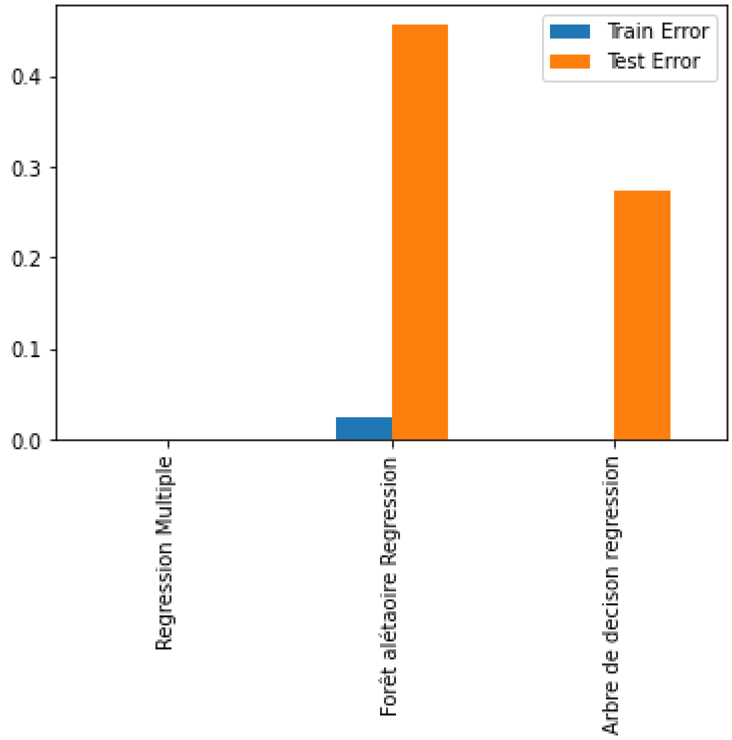

Figure n° 10 visualization of the error rate of source models: Ourselves

- **From a predictive point of view**

```
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]])
```

Figure n° 11 Loading new source data: Ourselves

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (51-57), Month: September - December 2023, Available at: www.noveltyjournals.com

The figure above allows us to better visualize the error rate that our algorithms can have. We note that the multiple regression has a precision of 100%, that is to say no error in the learning nor in the test while the decision tree is precise in the learning but presents a significant rate of error in the test for its part the random drill has a low percentage of error in learning and a fairly high error rate in the test.

3. DISCUSSION

The above analyzes allowed us to better visualize the error rate that our algorithms may have. We note that the multiple regression has a precision of 100%, that is to say no error in the learning nor in the test while the decision tree is precise in the learning but presents a significant rate of error in the test for its part the random drill has a low percentage of error in learning and a fairly high error rate in the test.

All these clarifications allowed us to conclude that multiple linear regression is best suited for inflation prediction compared to random forest and decision tree algorithms.

4. CONCLUSION

In this article, we have contributed to the approach of the Democratic Republic of Congo regarding the forecasting of monetary inflation, by proposing a comparative study between some models based on machine learning, well beyond traditional models, namely the Multiple regression, random forest and decision tree, with a view to identifying the best of these models applicable to monetary inflation forecasts.

We have therefore achieved our double objective assigned at the very beginning of this article, namely that of presenting the advantages brought by the three models based on machine learning compared to traditional models, also that of comparing these models studied here from the point of view of view of perdition, also of the prediction score.

It is important to note that after our comparative study of these three machine learning models, that multiple linear regression was declared the best suited for the prediction of inflation, compared to the two other models presented here in this article.

REFERENCES

[1] Alber, Maximilian, Sebastian Lapuschkin, Philipp Seegerer, Miriam Hägele, Kristof T. Schütt, Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller, Sven Dähne, and Pieter-Jan Kindermans. "iNNvestigate neural networks!." J. Mach. Learn. Res. 20, no. 93 (2019): 1-8.

[2] Jean – Yves CAPUL and Olivier GARNIER, Economic and Social Sciences Dictionary, Hatier, April 1999, p;

[3] Adebayo, Julius, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. "Sanity checks for saliency maps." arXiv preprint arXiv:1810.03292 (2018).

[4] Martin Nyongolo. IMPACT OF INFLATION ON TAX REVENUES IN THE DRC FROM 2005 TO 2014. CAHIERS DU CERUKI, NEW SERIES, CERUKI, 2015, pp.275-285. Ffhal 01235985;

[5] Alain, Guillaume, and Yoshua Bengio. "Understanding intermediate layers using linear classifier probes." arXiv preprint arXiv:1610.01644 (2016).

[6] Stock, J.H. and Watson, M.W. (2011). Dynamic factor models. In: M.P. Clements and D.F. Hendry, eds. The Oxford Handbook of Economic Forecasting. Oxford University Press;

[7] Aamodt, Agnar, and Enric Plaza. "Case-based reasoning: Foundational issues, methodological variations, and system approaches." AI communications 7.1 (1994): 39-59.

[8] Alberto, Túlio C, Johannes V Lochter, and Tiago A Almeida. "Tubespam: comment spam filtering on YouTube." In Machine Learning and Applications (Icmla), Ieee 14th International Conference on, 138–43. IEEE. (2015).