

Conceptual Model of the Decision-making System Based on Fuzzy Clustering and Nash Equilibrium: Application to the Financial Authorities of the DRC

**Assistant Jean Cibamba Kanyinda^a, Head of Works Bruce Mbombi Bakondolo^b,
Assistant Daniel Aluba Mulumba^c, Professor Richard Kitondua Lubanzadio^c**

^a Academy of Fine Arts, section: Graphic Art, Research Unit, Specialization: Visual Communication/ICT, Kinshasa -DRC;

^b Luozi Higher Pedagogical Institute, section: Science and Technology, Specialization: Management Information Systems, ISP-LUOZI, Kinshasa-DRC;

^c Faculty of Science and Technology, Department of Mathematics and Computer Science, National Pedagogical University (UPN), Kinshasa-DRC;

Abstract

This article proposes a conceptual and strategic overhaul of the decision-making system of the financial authorities of the Democratic Republic of Congo (DGI, DGDA, and DGRAD) within a context marked by the fragmentation of tax data, the inefficiency of traditional control mechanisms, information asymmetries, and difficulties in institutional coordination. Addressing the limitations of classical approaches to financial governance, this research develops a hybrid decision-making model integrating artificial intelligence and game theory techniques to improve the performance of public tax administrations.

Fuzzy C- Means (FCM) clustering algorithm and Nash equilibrium to design an intelligent system capable of segmenting taxpayers in an environment characterized by uncertainty, imprecision, and variability in tax behavior. Fuzzy clustering allows taxpayers to be classified according to multiple degrees of membership, thus offering a more realistic representation of tax risk profiles and compliance levels. In parallel, Nash equilibrium is used to model the strategic interactions between different tax authorities in order to optimize mechanisms for cooperation, information sharing, and collective decision-making.

The methodology relies on a conceptual model of the system, a mathematical formalization of the segmentation and coordination mechanisms, and an analytical simulation to evaluate the performance of the proposed model. The results show that a fuzzy risk-based approach significantly improves the identification of high-risk taxpayers, reduces tax harassment related to redundant audits, and promotes a better allocation of administrative resources. Furthermore, the integration of the Nash equilibrium allows for the transformation of situations of institutional competition into cooperative strategies that maximize social surplus, decision-making efficiency, and public revenue.

This research represents a scientific contribution at the intersection of artificial intelligence, intelligent decision-making systems, game theory, and digital financial governance. It also opens up important perspectives for modernizing tax administrations in emerging economies through the use of hybrid approaches based on AI and strategic optimization.

Keywords : *Fuzzy clustering ; Fuzzy C- Means ; Nash equilibrium; Game theory; Intelligent decision-making system; Financial authorities; Tax governance; Artificial intelligence; Tax risk analysis; Decision optimization; Big data analysis ; Democratic Republic of Congo.*

I. Introduction

In modern economies, the performance of tax administrations is a strategic lever for economic development, public governance, and the financing of government policies. The ability of revenue collection agencies to effectively mobilize public revenue directly influences fiscal stability, the quality of public services, and government investment capacity. However, in many developing countries, and particularly in the Democratic Republic of Congo (DRC), tax systems continue to face numerous structural challenges related to ineffective control mechanisms, economic informality, information asymmetry, and weak institutional coordination among revenue collection agencies.

In the DRC, despite considerable tax potential, the actual level of revenue collection remains relatively low compared to the country's economic capacity. This situation is largely due to the functional fragmentation of the main revenue collection agencies, namely the General Directorate of Taxes (DGI), the General Directorate of Customs and Excise (DGDA), and the General Directorate of Administrative, Judicial, State Property, and Equity Revenue (DGRAD). These institutions often operate in silos, with heterogeneous information systems and limited sharing of tax data. This lack of interconnectivity fosters redundant audits of certain taxpayers, tax harassment, jurisdictional conflicts, and the existence of tax havens.

Furthermore, the growth in the volume of administrative and financial data generated daily by economic activities opens new perspectives for the use of artificial intelligence technologies in tax governance systems. Traditional approaches to taxpayer segmentation and tax audits are now showing their limitations in the face of the complexity of economic behavior, data uncertainty, and the multiplicity of risk profiles. In this context, fuzzy clustering techniques appear as particularly well-suited tools for analyzing uncertain and imprecise environments, enabling the flexible classification of taxpayers according to varying degrees of membership in tax risk categories.

In parallel, the interactions between revenue authorities can be analyzed from a game theory perspective, particularly through the concept of Nash equilibrium. Indeed, each authority adopts decision-making strategies aimed at maximizing its own revenue collection objectives, sometimes at the expense of the overall coordination of the tax system. Integrating game theory into an intelligent decision-making system thus makes it possible to model the strategic behaviors of institutional actors in order to identify optimal cooperation mechanisms and reduce decision-making conflicts.

It is within this perspective that the present research is situated, the central problem of which is the following:

— How can the integration of fuzzy clustering and game theory enable the financial authorities of the Democratic Republic of Congo to transform heterogeneous tax data into intelligent, coordinated and optimized recovery strategies?

To address this issue, this study proposes a hybrid conceptual model combining the Fuzzy C- Means (FCM) algorithm and Nash Equilibrium within the framework of an intelligent decision-making system applied to the financial management of the DRC.

The proposed model is based on two fundamental mechanisms:

- Fuzzy C - Means clustering , used to segment taxpayers according to dynamic and imprecise risk profiles based on multiple degrees of membership ;
- The Nash equilibrium, used to model strategic interactions between financial authorities in order to identify cooperation strategies that maximize public revenue while minimizing institutional conflicts and administrative interference.

The main objective of this research is to contribute to the modernization of the Congolese tax decision-making system through the integration of artificial intelligence, data analysis, and strategic optimization approaches. This study lies at the intersection of intelligent decision-making systems, game theory, applied artificial intelligence, and the digital governance of public administrations.

2. Research Objectives

The main objective of this research is to design an intelligent conceptual decision-support model for the financial authorities of the Democratic Republic of Congo, integrating fuzzy clustering techniques and strategic coordination mechanisms from game theory. This model aims to improve the efficiency of tax governance, optimize controls, and enhance institutional coordination among financial administrations. Specifically, this study pursues the following objectives:

- design a centralized decision-making architecture enabling the integration and consolidation of tax data from different financial authorities through a Data Lake type environment in order to build a unified and consistent view of the taxpayer;
- develop an intelligent taxpayer segmentation mechanism based on the Fuzzy C- Means (FCM) algorithm, capable of identifying tax risk profiles from multiple degrees of membership in an environment characterized by data uncertainty and imprecision;
- develop a dynamic tax risk assessment system to detect taxpayers with a high probability of tax fraud, evasion or non-compliance;
- modeling the strategic interactions between the financial authorities (DGI, DGDA and DGRAD) using non-cooperative game theory in order to identify optimal decision-making strategies based on the Nash Equilibrium;
- optimize tax control and collection mechanisms by reducing administrative overlaps, institutional conflicts and tax harassment of taxpayers;
- improve coordination, information sharing and operational efficiency of financial authorities in the process of mobilizing public revenue;
- to propose a mathematical, analytical and technological framework that can serve as a basis for the future development of an intelligent decision-making system applied to digital fiscal governance in the Democratic Republic of Congo.

3. State of the Art and Related Works

3.1. Big Data and Artificial Intelligence in Public Administration

The work of Han, Kamber, and Pei (2012) on data mining forms the technical basis of this study. The use of distributed architectures such as Apache Spark , documented by Zaharia et al. (2016), makes it possible

to process the massive volumes of tax transactions (VAT, customs) that traditional systems can no longer handle in the DRC.

3.2. Fuzzy Logic and Uncertainty Management.

The originality of this model lies in the application of Zadeh 's (1965) "Fuzzy Sets." In developing economies, the boundary between the formal and informal sectors is porous. Xu and Wunsch (2005) demonstrated that Fuzzy C- Means (FCM) outperforms rigid algorithms (KMeans) when it comes to modeling ambiguous human behaviors, such as partial tax compliance.

3.3. Game Theory and Fiscal Governance.

The coordination dilemma between the DGI, the DGDA, and the DGRAD is a classic example of game theory. Drawing on Osborne (2004), this article proposes moving beyond the "Prisoner's Dilemma" (where each agency maximizes its own gain at the expense of the state) to reach a socially optimal Nash equilibrium through information sharing.

Table 1: Positioning of Research (DRC)

Author / Works	Approach	Limitations addressed by this model
Current Systems (DRC)	Simple declaration / Random audit	Information asymmetry, corruption, harassment.
Han et al. (2012)	Classic Data Mining	Integration of the "Fuzzy" dimension for informality.
This Model	Hybrid: FCM + Nash	Inter-agency coordination and granular segmentation.

4 Methodological Synthesis

The transition from theoretical work to practice occurs via the following pipeline:

Data Collection (Hadoop) → Data Cleaning (Hive) → Data Segmentation (FCM) → Data Optimization (Nash) → Data Decision (Ministerial Dashboard/Power BI). This approach directly addresses the recommendations of the **Central Bank of Congo (2025)** regarding the digitalization of public finances.

Table 2: Summary of Scientific Positioning

Domain	Key reference	Contribution of the article
data mining	Han et al. (2012)	application to the DRC 's fiscal data lake
fuzzy logic	zadeh (1965)	modeling informality as a "degree of belonging"
game theory	nash / osborne (2004)	conflict resolution dgi-dgda - dgrad

5 Contribution and Justification for the Choice of the FCM-Nash Hybrid Model.

The hybrid model presented offers a robust scientific solution to the fiscal coordination challenges encountered in the DRC.

- Fuzzy C- Means Segmentation : The implementation of the FCM algorithm captures the complexity of the informal sector and precisely targets "Tax Potential." Each taxpayer is assigned degrees of membership in different segments, accurately reflecting the reality of their tax-paying behavior.
- Nash equilibrium: This formal framework reduces jurisdictional conflicts between the various agencies (DGI, DGDA, DGRAD). Revenue optimization does not result from maximum aggression by each actor, but from a balanced distribution of control efforts to minimize administrative interference.
- Flexibility of Mixed Strategies: Unlike traditional approaches, the use of mixed strategies makes it possible to capture the uncertainty and decision-making flexibility inherent in financial management.
- The choice of this model in the governance of public revenue is justified as follows:
- Regarding the choice of "Fuzzy" : in the DRC economy, nothing is black or white. "Fuzzy" allows us to model the reality of the informal sector, which classical K- means ignores.
- On Nash Equilibrium: Let's clarify that we are not changing the laws, but we are changing the administration's strategy to make it rational and coordinated. - On technology: PySpark - hive takes into account the scalability and interoperability of data: the system is ready to process millions of customs and tax transactions.

6. Data Structure

For such a system in the DRC, it is difficult to find a ready-made public dataset containing real data from the DGI (General Directorate of Taxes) or the DGDA (General Directorate of Customs and Excise) due to tax confidentiality reasons. However, to build our proof of concept (PoC), we used a synthetic dataset and an open-source data collection structure that mimics the behavior of revenue authorities.

Here are the essential variables that our dataset must contain to make Fuzzy Clustering and the Nash Model work:

1. Dataset Structure (Features per Taxpayer)

For the PySpark algorithm to be able to segment, each row must represent a taxpayer (or customs entity) with the following columns:

Category	Variable(feature)	Description
Identify	Taxpayer ID	Unique key (Tax number)
Tax	Amount Declared	Revenue reported for the financial year
Tax	Amount Paid	Amount actually received by the Public Treasury
Behavior	Payment_Delay	Actual amount received by the public treasury
Historical	Number of Rectifications	Number of adjustments made during an audit
Sector	Activity_Code	Sector of activity (mining, telecom , commerce...)
location	Province_DP	Provincial Directorate of Affiliation

2. Data sources for the DRC

Since we do not have direct access to the Isys -Régies or Logirad servers , contextual data is retrieved:

- EITI-DRC (Extractive Industries Transparency Initiative) reports: They publish detailed Excel files on payments made by each mining company to the DGI, DGDA, and DGRAD. This is the richest source for the mining sector. [18]
- DRC Open Data Portal: Sometimes, statistics by sector and by province are available, although often aggregated. [17]
- Statistics from the Central Bank of Congo (BCC): Useful for correlating clusters with real economic activity (sectoral GDP). [15]
- Kaggle data on taxRix

3. Generating a Test Dataset (PySpark Code)

To test our Nash model immediately, we generated simulated data reflecting the three profiles (Partner, Undecided, Refractory) with 1,000,000 taxpayers.

4. Fuzzy Segmentation of the Taxpayer (Input)

Instead of rigid categories, each taxpayer is defined by a membership vector of three typical profiles, calculated by your PySpark engine :

- The Partner (C₁): Compliant, pays on time, minor errors (often large companies).
- The Indecisive (C₂): Late payments, sometimes incomplete declarations, sensitive to penalties.
- The Refractory (C₃): High-risk profile, willful omissions, informal or opaque sectors.

5. Advantages for Public Authorities in the DRC

- Resource optimization: We no longer control everyone in the same way. We concentrate inspectors (a scarce resource) on profiles where the gain in Nash equilibrium is highest.
- Reducing corruption: By automating risk scoring via PySpark , we limit the discretionary power of agents, which is a pillar of governance change.
- Revenue predictability: The model allows us to simulate how an increase in penalties (change in state strategy) will impact taxpayer behavior.

7. Fuzzy SEGMENTATION MODEL (FCM)

The model processes the data via the FCM algorithm to minimize the intra-cluster distance while allowing multiple membership. [06]

Either $i = 1, 2, \dots, n$: the taxpayers, $K = 1, 2, \dots, K$: the segments, We note : $\sum_{i=1}^K u_{ik} = 1 \quad \forall i$

7.1. Objective Function.

Segmentation can be modeled by the objective function of the fuzzy C- means . The FCM algorithm minimizes the following objective function: [05]

$$J_m = \sum_{i=1}^n \sum_{k=1}^K u_{ik}^m \| X_i - C_k \|^2$$

Or :

- $X_i \rightarrow$ Vecteur des caractéristiques du Contribuables i
- $u_{ij} \rightarrow$ Degré d'appartenance du contribuable i au cluster k .
- m et le parametre de flou ($m > 1$), plus m est élevé pus les clusters sont flous .
- $C_K \rightarrow$ Centre du cluster k
- $\| X_i - C_k \|^2 =$ distance euclidienne entre le contribuable et le centre du cluster .

The constraint is that the sum of memberships for an individual must equal 1:

$\sum_{j=1}^c u_{ij} = 1$ $\forall i$ that is to say each taxpayer defines 100% of their membership in three clusters.

7.2. Key Performance Indicators (KPIs)

The system relies on indicators that allow management to monitor the financial health of the State [07]:

Key indicator	Usefulness for management
- Assignment completion rate	- Compare actual revenues to forecasts
- Tax gap	- Identify the difference between tax due and tax collected
- Cost of collection	- Evaluate how much the collar costs
- Anomaly score per taxpayer	
- Detect suspicious statements using filtering algorithms	

7.3. Mathematical Steps of the Algorithm

Step 1 : Data preparation and normalization.

Before the decision, each tax authority has a vague segmentation of taxpayers.

We have a data matrix $X = \{X_1, X_2, \dots, X_N\}$

Where each taxpayer is represented by: $X_i = \{X_{i1}, X_{i2}, \dots, X_{ip}\} \quad \forall i \in k$

With :

- N = nombre des contriuanables
- p = variables explicatives
- $i = 1 \dots, N$: the taxpayer
- $k = 1, \dots, k$: nombre des segments (cluster)

X_i Contains: amount paid , payment frequency, arrears, litigation, compliance, sector of activity, location, risk score.

The variables must be normalized because they do not have the same scale:

min-max normalization

$X'_{ik} = \frac{X_{ik} - \min(X_{ik})}{\max(X_k) - \min(X_k)}$ Or : X_{ik} = Valeur brute, X'_{ik} = normalized value between 0 and 1. Very important to prevent one variable from dominating the other variables.

Step 2 : The Degree of Membership Matrix (U)

Clustering (FCM) produces a matrix U of dimension $N \times C$ (N taxpayers, C clusters).

Contrary to binary logic, each taxpayer i has a hybrid profile.

$U = [u_{ik}]$, where $0 \leq u_{ik} \leq 1$ and $\sum u_{ik} = 1$

• Initializing the fuzzy center

- Choose k segments ($k=3$)
- $C = \{c_1, c_2, c_3\}$
- Randomly initialize the matrix U: $U^{(0)} = [u_{ik}^{(0)}]_i$ avec $u_{ik} \in [0, 1]$ et $\sum_{k=1}^K u_{ik}^{(0)} = 1$

$U = \mathbb{R}^{N \times C}$:

- Each line $\rightarrow$ represents one taxpayer
- Each column is $\rightarrow$ a cluster

Table 3: Example of a Membership Degree Matrix

Taxpayer (ID)	Cluster 1 (Compliant)	Cluster 2 (Area Grey)	Cluster 3 (Top Risk)	Dominant Profile
CTB-001	0.85	0.10	0.05	Compliant (Safe)
CTB-002	0.20	0.65	0.15	Ambiguous (To be monitored)
CTB-003	0.02	0.08	0.90	Probable Fraud (Audit)

The Transition from FCM to Nash Equilibrium The transition between AI (FCM) and Game Theory (Nash) is achieved by transforming the degree of membership into a risk parameter θ (theta) which defines the "Payoff" (gain) of the financial management.

The risk-weighted average of taxpayer i is calculated as follows:

Step 3 : Update (Cluster center calculation)

At iteration t : $V_k^{(t)} = \frac{\sum_{i=1}^n (u_{ik}^{(t)})^m x_i}{\sum_{i=1}^n (u_{ik}^{(t)})^m}$ where $m \geq 1$ = coefficient flou (souvent = 2)

Taxpayers with a strong cluster affiliation have a greater influence on its center.

Step 4 : Calculating distances

For each taxpayer i and cluster j , we calculate: $d_{ij} = \|X_i - V_j\|$

Often with Euclidean distance: $d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - v_k)^2}$

Step 5 : Updating membership degree

The matrix is updated by: $u_{ij}^{(t+1)} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}}$ If a taxpayer is close to a cluster, their degree of belonging

to that cluster increases.

Step 6 Convergence test / FCM stopping criterion

We compare the successive matrices : $\|U_k^{(t+1)} - U_k^{(t)}\| \leq \epsilon_1$

- If true, the algorithm will stop
- If not, we return to step 3.

8. Modeling Strategic Interaction (Nash Equilibrium)

The modeling of the strategic interaction between the three financial authorities (DGI, DGDA, DGRAD) is based on a non-cooperative game where the gains (payoffs) are directly fueled by the results of fuzzy clustering .

Here are the constituent elements of the 3D payoff matrix as defined in our work:

8.1. The Transition from FCM to Nash Equilibrium

The transition between AI (FCM) and Game Theory (Nash) occurs through the transformation of the **degree** of membership into a risk parameter θ (theta) which defines the " Payoff " (gain) of the financial management system. For each management system r , a strategic index of fiscal potential is calculated [24]:

$$\theta_r = \sum_{k=1}^n \sum_{k=1}^K u_{ik} \alpha_{rk} v_i$$

Or :

- θ_r = potentiel global exploitable
- u_{ik} = degré d' appartenance de contribuable i au segment k
- α_{rk} = importance stratégique du segment k pour la région r.
- v_i = valeur fiscale /économique du contribuable.

8.2. The Three-Dimensional Strategic Space

To simplify and make the model robust: we define for each management system r , a three-dimensional strategy S_r representing the allocation of its resources:

$S_r = (C_r, q_r, d_r)$ Or :

- C_r = effort de contrôle fiscal/douanier
- q_r = effort de recouvrement
- d_r = effort de digitalisation intelligence décisionnelle

Given the budget constraint:

$$C_r + q_r + d_r \leq \beta_r \quad C_r + q_r + d_r \geq 0 \quad \text{where : } \beta_r = \text{budget/total capacity of the control room } r$$

8.3. Non-cooperative interactions between agencies.

The agencies interact:

- Overlap of actions
- Pressure on the same targets
- Budget competition
- Imperfect coordination

We model: $I_r(S_r, S_{-r}) = \sum_{j \neq r} \beta_{rj} (C_r C_j + q_r q_j + d_r d_j)$

Or : β_{rj} = intensité de l' interdépendance entre région r et j .

8.4. Payoff Function (Utility)

payoff matrix for a management company r is calculated according to the following utility function:

$$U_r(S_r, S_{-r}) = \theta_r (\lambda_1 C_r + \lambda_2 q_r + \lambda_3 d_r) - \frac{1}{2} (a_r C_r^2 + b_r q_r^2 + g_r d_r^2) + \sum_{j \neq r} \beta_{rj} (C_r C_j + q_r q_j + d_r d_j)$$

Under duress:

$$C_r + q_r + d_r \leq \beta_r \quad C_r + q_r + d_r \geq 0$$

By temporarily neglecting the active budget constraint, the first-order conditions Y give:

- ✓ For C_r : $\frac{\partial U_r}{\partial C_r} = \theta_r \lambda_1 - a_r C_r + \sum_{j \neq r} \beta_{rj} C_j = 0$, SO : $C_r^* = \frac{\theta_r \lambda_1 + \sum_{j \neq r} \beta_{rj} C_j}{a_r}$
- ✓ For q_r : $\frac{\partial U_r}{\partial q_r} = \theta_r \lambda_2 - b_r q_r + \sum_{j \neq r} \beta_{rj} q_j = 0$, SO : $q_r^* = \frac{\theta_r \lambda_2 + \sum_{j \neq r} \beta_{rj} q_j}{b_r}$
- ✓ For d_r : $\frac{\partial U_r}{\partial d_r} = \theta_r \lambda_3 - g_r d_r + \sum_{j \neq r} \beta_{rj} d_j = 0$, SO : $d_r^* = \frac{\theta_r \lambda_3 + \sum_{j \neq r} \beta_{rj} d_j}{g_r}$

Global version:

$$U_r(e_r, e_{-r}) = \theta_r e_r - \frac{\gamma_r}{2} e_r^2 + \sum_{j \neq r} \beta_{rj} e_r C_j$$

e : overall effort with: $\theta_r = \sum_{i=1}^n \sum_{k=1}^K u_{ik} \alpha_{rk} v_i$ and $e_r^* = \frac{\theta_r + \sum_{j \neq r} \beta_{rj} e_j}{\gamma_r}$

This formula incorporates three major components:

- Expected gain: The segmented tax potential multiplied by the effort provided.
- Internal cost: The disutility linked to the intensity of the effort (modeled in a quadratic way).
- Strategic interference (β_{rj}): The effect of synergy or overlap with the decisions of other governments j .

8.5. Structure of the Discrete Payoff Matrix

To illustrate the Nash equilibrium, the article defines three levels of finite strategies per rule:

- A (Aggressive): Intensive recovery and systematic targeting.
- M (Moderate): Awareness and support for taxpayers.
- P (Prudent): Passive approach favoring digital interconnection.

5.1. Game Definition

In our context:

- The players are the financial authorities
- Each agency chooses a strategy for managing its resources
- This strategy depends on the taxpayer segments resulting from fuzzy clustering
- The agencies pursue their own objectives (revenue, collection, compliance, efficiency)
- Their decision is interdependent.

The game is defined by $G = [N, S, U]$:

- N : The players (DGI, DGDA, DGRAD).
- S : The strategic space (levels of control, recovery, digital).
- U : The utility function (Payoff), where the gain is a function of the segmented fiscal potential less control costs and interferences.

The goal of the model is to find a Nash equilibrium, that is, a stable state where no governing body has an incentive to change its strategy on its own [05]. It is a model of a non-cooperative strategic game driven by fuzzy indicators.

5.2. Pay Off Matrix

generic Π_r payoff is defined as : = Expected Gain – Cost +/- Interference (synergy)

In our model, the game being non-cooperative, the synergy is additive, unlike the classic economic model.

• Parameters derived from fuzzy clustering

Let's assume four segments (clusters)

- $C = \text{for/confome}$, $C_2 = \text{à potentiel}$, $C_3 = \text{à risque}$

For each taxpayer , i the potential tax liability is calculated for each tax authority : r

$$\theta_r = \sum_{i=1}^n \sum_{k=1}^K u_{ik} \alpha_{rk} v_i, u_{ik} \in [0,1], \sum_{k=1}^K u_{ik} = 0$$

Or :

- θ_r = potentiel global exploitable
- u_{ik} = degré d'appartenance de contribuable i au segment k

- α_{rk} = importance strategique du segment k pour la regie r.
- v_i = valeur fiscale du contribuable i

• Discreet payoff function

For each management team r ,

We define: $\Pi_r(x_r, x_{-r}) = \theta_r x_r - \frac{\gamma_r}{2} x_r^2 + \sum_{j \neq r} \beta_{rj} x_r, C_j$

Or :

- θ_r = potentiel fiscal flou
- γ_r = coût interne d' intensité
- β_{rj} = interference strategique.

FCM provides the proportion of taxpayers:

- The Partner (P_1): Compliant, pays on time, minor errors (often large companies).
- The Indecisive One (P_2): Late payments, sometimes incomplete declarations, sensitive to penalties.
- The Refractory (P_3): High-risk profile, deliberate omissions, informal or opaque sectors.

Then $P_1 + P_2 + P_3 = 1$, these proportions come into play in the gain function.

$$P_r = \frac{1}{N} \sum_{i=1}^N u_{ir} , \quad P_r = \text{poids global des contriuable } i .$$

Ce sont ces variables qui alimentent le jeu strategique

5.3. NASH model stage

1. Definition of Nash equilibrium

A strategy profile: $s^* = (s_1^*, s_2^*, s_3^*)$ is a suitable Nash equilibrium if: $U_r(s_r^*, s_{-r}^*) \geq U_r(s_r, s_{-r}^*)$ for all $s_r \in S_r \forall r \in N[06]$

Interpretation: no agency can single-handedly improve its performance by unilaterally modifying: its level of control, its collection efforts, its digital investment.

Given the decisions already made by others.

1) Detailed step of the Nash model

Step 1: Initializing strategies.

Initialize for each control unit: $e_r^{(0)} > 0$ ou dans sa version vectorielle $(C_r^{(0)}, q_r^{(0)}, d_r^{(0)})$

Step 2: Update by BR (Best Response Match)

We define : $BR_i(x_{-i}) = \operatorname{argmax}_{x_i \in \Delta(S_i)} U_i(x_i, x_{-i})$, the profile x^* is a Nash equilibrium if: $x_i^* \in$

$BR_i(x_{-i}^*), \forall i \in N$ For the simple version: $e_r^{(t+1)} = \max(0, \min(Br, \theta_r - \sum_{j \neq k} \beta_{ij} e_r^{(t)})$

We apply the projection onto the interval $[0, Br]$

Step 3: Nash model stopping criterion

Stop if: $\|e_k^{(t+1)} - e_k^{(t)}\| \leq \epsilon_2$ where $e_k^{(t)} = (e_1^{(t)}, e_2^{(t)}, e_3^{(t)})$

Intensity-based strategy encoding

To move from qualitative to quantitative, we take: $A = 3, M = 2, P = 1$

Note: $X_r = \{3, 2, 1\}$ depending on the strategy chosen

5.4. Complete algorithm of the proposed model

- 1) Taxpayer data
- 2) Standardization
- 3) Initializing the FCM
- 4) FCM Algorithm
- 5) FCM execution
 - Calculating the centers
 - Calculation of degrees of membership
 - Distance calculation
- 6) FCM Result
- 7) Tax functions
- 8) Game settings:
 - $\theta_r = \sum_{i=1}^n \sum_{k=1}^K u_{ik} \alpha_{rk} v_i$ aleatoirement ($\theta_r = 12, \theta_1 = 10, \theta_1 = 10$)
- 9) Payoff function :
 - $\Pi_r(x_r, x_{-r}) = \theta_r x_r, -\frac{\gamma_r}{2} x_r^2 + \sum_{j \neq r} \beta_{rj} x_r, C_j$
- 10) 3D payoff matrix :
 - $E = 2, F = 3, P = 1,$
 - Matrice $U = [U_1(x_1, x_2, x_3), U_2(x_1, x_2, x_3), U_3(x_1, x_2, x_3)]$
- 11) Slice display
- 12) Nash equilibrium
 - Definition of Nash equilibrium
 - $BR_{DGI}, BR_{DGDA}, BR_{DGRAD}$
 - Overall balance
- 13) Display.

5.5. FCM Username Code

Text: →entry :

- $X =$ matrice de contribuables
- $C =$ nombre de cluster
- $m =$ coefficient de flou ($m = 2$)
- $\epsilon =$ seuil de convergence $maxIter =$ nombre maximal d' iteration
- →exit :
- $U =$ matrice finale de degré d' appartenance
- $V =$ Centre finaux des clusters (cluster dominant de chaque contribuable)
- → **Algorithme:**
- 1) normaliser les données X

2) *initialiser aleatoirement la matrice U telle que :*

$$u_{ij} \in [0,1], \sum_{j=1}^c u_{ij} = 1$$

3) *pour $t = 1$ à $MaxIter$ faire:*

a) Calculer les centres de cluster : $V_j = [\sum_{i=1}^n (u_{ij}^m)]$

b) Calculer les distances d_{ij} entre X_i et V_j

c) Mettre à jour U : $u_{ij} = [1 / (d_{ij} / d_{ik})^{(2/(m-1))}]$

d) Si $\|U_k^{(t+1)} - U_k^{(t)}\| \geq \epsilon$ alors arreter

4) Assign each taxpayer to the dominant cluster: $\text{cluster}(i) = \text{Argmax}_j(u_{ij})$

5) Interpreter les cluster $C_1 \rightarrow \text{performant}$, $C_2 \rightarrow \text{à potentiel}$, $C_3 \rightarrow \text{à risque}$

fuzzy c- means algorithm forms the core of intelligent taxpayer segmentation in this model. It allows taxpayers to be grouped according to their level of contributions, regularity, compliance, risk, and digitalization, while recognizing that a single taxpayer may simultaneously exhibit characteristics belonging to several tax profiles.

Thus, each taxpayer is assigned degrees of membership in different clusters reflecting the uncertain and evolving nature of tax behavior.

The FCM results then feed into the decision-making model based on a detailed mapping of taxpayer segments and strategic priorities for financial authorities in the DRC.

Note that PySpark MLLib directly contains K_means but not the native FCM in its classic version.

Thus, FCM will be implemented under Apache Spark via custom distributed logic, based on Operations Data Frames/RDD for the iterative calculation of distance centers and degrees of membership. [12]

5.6. Numerical example of the application of the FCM and Nash equilibrium:

Application of fuzzy clustering (FCM) and Nash equilibrium for taxpayer segmentation and performance management of revenue collection agencies in the DRC. We will construct:

- A dataset of 10 taxpayers;
- The tax variables taken into account;
- Fuzzy segmentation by FCM into 3 clusters;
- The matrix of degrees of membership;
- The interpretation of the 3 clusters;
- Calculation of tax potentials by agency (DGI, DGDA, DGRAD);
- Parameter calibration;
- payoff matrices ;
- The search for Nash equilibrium;
- Decision-making interpretation.

1) Fictitious dataset: 10 taxpayers

We assume 10 taxpayers (companies / economic operators) observed according to 3 strategic tax variables:

- Annual revenue (in millions of CDF)
- Import/export volume (in millions CDF)
- Administrative fees and taxes (in millions of CDF)

Table 4: Raw taxpayer data

Taxpayer	CA(X ₁) : import	CA(X ₁) : export	Relevance (X ₂)
C1	120	80	15
C2	150	95	20
C3	200	130	18
C4	250	170	25
C5	300	220	28
C6	380	260	35
C7	450	310	40
C8	520	370	45
C9	500	420	52
C10	700	500	60

2) Objective of fuzzy segmentation

We want to group taxpayers into 3 vaguely defined segments:

- Cluster 1 (Low tax potential)
- Cluster 2 (Medium Fiscal Potential)
- Cluster 3 (Strong tax potential)

Unlike traditional clustering, a taxpayer can partially belong to several clusters.

Fuzzy C- Means (FCM) Parameters

We choose:

- Number of clusters: $C = 3$
- Fuzzy exponent: $m = 2$
- Number of taxpayers: $n = 10$
- Number of variables: $p = 3$

4) Initial (hypothetical) cluster centers

For example, let's assume the final centers are obtained by FCM:

$$V_1 = (150, 100, 18) \quad V_3 = (600, 430, 52)$$

Interpretation:

- V_1 = low-potential taxpayers
- V_2 = taxpayers of intermediate potential
- V_3 = high-potential taxpayers

5) FCM Degrees of Membership Formula

$$\text{For each taxpayer } i \text{ and cluster } k: u_{ik} = \frac{1}{\sum_{j=1}^3 \left(\frac{d_{ik}}{d_{ij}} \right)^{\frac{2}{m-2}}}$$

$$\text{With : } d_{ik} = |x_i - v_k| = \text{Euclidean distance } m = 2 \text{ Therefore: } u_{ik} = \frac{1}{\sum_{j=1}^3 \left(\frac{d_{ik}}{d_{ij}} \right)^2}$$

6) Detailed example of FCM calculation for C_4

Taxpayer: $C_4 = (250, 170, 25)$

- Distance to cluster 1: $d_{41} = \sqrt{(250 - 300)^2 + (170 - 100)^2 + (25 - 18)^2} \approx 122.27$
- Distance to cluster 2: $d_{42} = \sqrt{(250 - 300)^2 + (170 - 100)^2 + (25 - 18)^2} \approx 122.27$
- Distance to cluster 3: $d_{43} = \sqrt{(250 - 300)^2 + (170 - 100)^2 + (25 - 18)^2} \approx 436.84$
- Degree of membership in Cluster 1:

$$u_{41} = \frac{1}{\left(\frac{122.27}{122.27}\right)^2 + \left(\frac{122.27}{122.27}\right)^2 + \left(\frac{122.27}{436.84}\right)^2} \approx 0.481$$
- Degree of membership in Cluster 2:

Same symmetry: $u_{41} = \frac{1}{\left(\frac{122.27}{122.27}\right)^2 + \left(\frac{122.27}{122.27}\right)^2 + \left(\frac{122.27}{436.84}\right)^2} \approx 0.481$

Degree of membership in Cluster 3: $u_{41} = \frac{1}{\left(\frac{436.84}{122.27}\right)^2 + \left(\frac{436.84}{122.27}\right)^2 + \left(\frac{1436.847}{436.84}\right)^2} \approx 0.038$

Verification: $0.481 + 0.481 + 0.038 = 1$ ✓ **Correct.**

7) Simplified final matrix of membership degrees

To simplify the study, we can present a coherent final matrix (from the FCM after convergence).

Table 5: Fuzzy Membership Matrix

Taxpayer	Cluster 1	Cluster 2	Cluster 3
C1	0.88	0.10	0.02
C2	0.80	0.17	0.03
C3	0.62	0.30	0.08
C4	0.48	0.48	0.04
C5	0.25	0.65	0.10
C6	0.10	0.78	0.12
C7	0.05	0.70	0.25
C8	0.03	0.40	0.57
C9	0.02	0.20	0.78
C10	0.01	0.08	0.91

1) Interpretation of clusters

- Cluster 1: Low potential, Taxpayers mainly: C1, C2, C3
- Cluster 2: Average potential, Taxpayers predominantly: C4, C5, C6, C7
- Cluster 3: High potential, Taxpayers predominantly: C8, C9, C10

9) Construction of tax potential by management

We will associate each variable with a control:

- DGI → mainly linked to CA ()
- DGDA → primarily related to Import/Export ()

- DGRAD → mainly related to Royalties ()

10) Calculation of aggregate fuzzy potentials per cluster

We calculate a tax potential weighted by degrees of membership .[23]

10.1 DGI Potential per Cluster

$$p_k^{(DGI)} = \sum_{i=0}^{10} u_{ik} X_{1k}$$

Cluster 1 Approximate calculation:

- C1: 0.88 x 120 = 105.6	- C6: 0.10 x 380 = 38
- C2: 0.80 x 150 = 120	- C7: 0.05 x 450 = 22.5
- C3: 0.62 x 200 = 124	- C8: 0.03 x 520 = 15.6
- C4: 0.48 x 250 = 120	- C9: 0.02 x 600 = 12
- C5: 0.25 x 300 = 75	- C10: 0.01 x 700 = 7

$$p_1^{(DGI)} \approx 639,7$$

Cluster 2: Same procedure: $p_1^{(DGDA)} \approx 1561,6$

Cluster 3 : same logic : $p_1^{(DGRAD)} \approx 1868,7$

10.2 DGDA Potential per Cluster

$$p_k^{(DGDA)} = \sum_{i=0}^{10} u_{ik} X_{2i} \text{ Approximate results: } p_1^{(DGDA)} \approx 432,8 \quad p_3^{(DGDA)} \approx 1504,4$$

10.3 DGRAD Potential per Cluster

$$p_k^{(DGRAD)} = \sum_{i=0}^{10} u_{ik} X_{3i} \text{ Approximate results: } p_1^{(DGRAD)} \approx 84,6 \quad p_3^{(DGRAD)} \approx 258,2$$

11) Translation of clusters into strategic levels

We use:

- F = Cluster 1 = weak
- M = Cluster 2 = average
- E = Cluster 3 = high

$$SO : F \rightarrow 3 \quad M \rightarrow 2 \quad E \rightarrow 1$$

12) Calibrating parameters from the FCM

To make the model more realistic, the global potentials are normalized .[23]

To simplify, we define:

- DGI stronger:
- DGDA intermediary:
- DGRAD more moderate:

☞ This calibration is consistent with: $p_3^{(DGI)} > p_3^{(DGDA)} > p_3^{(DGRAD)}$

13) Adapted payoff functions (derived from FCM)

We are using the following model:

$$[U_1 = 12x_1 - x_1^2 - x_1x_2 + x_1x_3], [U_2 = 10x_2 - x_2^2 - x_2x_1 + x_2x_3], [U_3 = 10x_3 - x_3^2 - x_1x_3 + x_2x_3]$$

With : $x_1 \rightarrow \text{strategie DGI}$ $x_2 \rightarrow \text{strategie DGDA}$ $x_3 \rightarrow \text{strategie DGRAD}$

Complete payoff matrices (3 tranches)

- Tranche 1: DGRAD = F (3)

DGI / DGDA	F	M	E
F	(13,11,11)	(13,20,12)	(15,27,23)
M	(24,12,12)	(26,22,13)	(28,30,14)
E	(33,13,13)	(36,24,14)	(39,33,15)



Python code

```
import numpy as np
import pandas as pd
from itertools import product

# =====
# 1. CONTRIBUTOR DATA (Module 1)
# =====
data = {
    "taxpayers ": ["C1", "C2", "C3", "C4", "C5", "C6", "C7", "C8", "C9", "C10"],
    "AC": [120, 150, 200, 250, 300, 380, 450, 520, 600, 700],
    "ImportExport ": [80, 95, 130, 170, 220, 260, 310, 390, 420, 500],
    "Fee ": [15, 20, 18, 25, 28, 35, 40, 45, 52, 60]
}
df = pd.DataFrame (data)
# =====
#2. MIN-MAX NORMALIZATION (Module 2)
# =====
features = ["CA", " ImportExport ", " Royalty "]
X = df[ features]. values.astype (float)
X_min , X_max = X.min ( axis=0), X.max (axis=0)
X_norm = (X - X_min ) / ( X_max - X_min )
# =====
# 3. FCM INITIALIZATION (Module 3)
# =====
def initialize_membership ( n_samples , n_clusters ):
    U = np.random.rand ( n_samples , n_clusters )
    return U / np.sum (U, axis=1, keepdims =True)
```

3.1 Calculation of centers (Submodule)

```
def compute_centers (X, U, m):
    Um = U**m
    return ( Um.T @ X) / np.sum ( Um.T , axis=1, keepdims =True)
```

3.2 Updating Memberships (Sub-module)

```
def update_membership (X, centers, m):
    n_samples , n_clusters = X.shape [ 0], centers.shape [0]
    U_new = np.zeros ( ( n_samples , n_clusters ))
    for i in range( n_samples ):
        dist = np.linalg.norm (X[ i ] - centers, axis=1) + 1e-8
        for k in range( n_clusters ):
            U_new [ i , k] = 1.0 / np.sum (( dist [k] / dist ) ** (2 / (m - 1)))
    return U_new
```

=====

4. FCM ALGORITHM (Module 4)

=====

```
def fuzzy_c_means (X, n_clusters =3, m=2, max_iter =100, tol =1e-5):
    U = initialize_membership ( X.shape [ 0], n_clusters )
    for i in range( max_iter ):
        U_old = U.copy ()
        centers = compute_centers (X, U, m)
    U = update_membership (X, centers, m)
    if np.linalg.norm (U - U_old ) < tol : break
    return centers, U
```

=====

5. FCM EXECUTION (Module 5)

=====

```
centers , U = fuzzy_c_means ( X_norm )
```

=====

6. FCM RESULTS (Module 6)

=====

```
for k in range( U.shape [1]):
    df [ f'cluster_{k+1}'] = U[:, k]
df [ " cluster_Dominant "] = np.argmax (U, axis=1) + 1
```

=====

#7. TAX POTENTIAL (Module 7)

=====

```
def compute_fiscal_potentials ( df , U):
    potentials = { "DGI": [], "DGDA": [], "DGRAD": []}
    for k in range( U.shape [1]):
        mu_k = U[ :, k]
        potentials[ "DGI"].append( np.sum ( mu_k * df ["CA"].values))
        potentials[ "DGDA"].append( np.sum ( mu_k * df [" ImportExport "].values))
```

```
potentials[ "DGRAD"].append( np.sum ( mu_k * df [" Royalty "].values))
return potentials
potentials = compute_fiscal_potentials ( df , U)
# =====
#8. GAME SETTINGS (Module 8)
# =====
theta1, theta2, theta3 = 12, 10, 10
# =====
#9. PAYOFF FUNCTIONS (Module 9)
# =====
def U1(x1, x2, x3): return theta1 * x1 - (x1**2) + x1 * x2 + x1 * x3
def U2(x1, x2, x3): return theta2 * x2 - (x2**2) + x2 * x1 + x2 * x3
def U3(x1, x2, x3): return theta3 * x3 - (x3**2) + x1 * x3 + x2 * x3
# =====
#10. PAYOFF 3D MATRIX (Module 10)
# =====
strategies = [1, 2, 3] # F, M, E
labels = {1: "F", 2: "M", 3: "E"}
payoff_matrix = {s: ( U1( *s), U2(*s), U3(*s)) for s in product(strategies, repeat=3)}
# =====
# 11. DISPLAYING THE SLICES (Module 11)
# =====
def print_slice ( x3_val ):
    print ( f"\n==== SLICE DGRAD = {labels[x3_val]} ====")
    for x1 in strategies:
        row = [ payoff_matrix [(x1, x2, x3_val)] for x2 in strategies]
        print( f"DGI {labels[x1]}: {row}")
    for s in strategies: print_slice (s)
# =====
#12. NASH EQUILIBRIUM (Module 12)
# =====
def is_nash ( p, pm ) :
    x1 , x2, x3 = p
    if pm[p][0] < max(U1(s, x2, x3) for s in strategies): return False
    if pm[p][1] < max(U2(x1, s, x3) for s in strategies): return False
    if pm[p][2] < max(U3(x1, x2, s) for s in strategies): return False
    return True
nash_equilibria = [p for p in payoff_matrix if is_nash ( p, payoff_matrix ) ]
# =====
#13. INTERPRETATION OF EQUILIBRIA (Module 13)
# =====
print ( "\n==== FINAL EQUILIBRIUM ANALYSIS ====")
```

for eq in nash_equilibria :

pay = payoff_matrix [eq]

print(f"Nash -> DGI:{labels[eq [0]]}, DGDA:{labels[eq [1]]}, DGRAD:{labels[eq [2]]} | Payoffs: {pay}")

Final result

=====

===== SLICE DGRAD = F =====

DGI F: [(13, 11, 11), (14, 20, 12), (15, 27, 13)]

DGI M: [(24, 12, 12), (26, 22, 13), (28, 30, 14)]

DGI E: [(33, 13, 13), (36, 24, 14), (39, 33, 15)]

===== DGRAD SLICE = M =====

DGI F: [(14, 12, 20), (15, 22, 22), (16, 30, 24)]

DGI M: [(26, 13, 22), (28, 24, 24), (30, 33, 26)]

DGI E: [(36, 14, 24), (39, 26, 26), (42, 36, 28)]

===== DGRAD SLICES = E =====

DGI F: [(15, 13, 27), (16, 24, 30), (17, 33, 33)]

DGI M: [(28, 14, 30), (30, 26, 33), (32, 36, 36)]

DGI E: [(39, 15, 33), (42, 28, 36), (45, 39, 39)]

===== FINAL EQUILIBRIUM ANALYSIS =====

Nash -> DGI:E, DGDA:E, DGRAD:F | Payoffs : (15, 9, 0.5)

❖ Best answer

Step 1: The best answer for the DGI

In all brackets, DGI prefers E because its gains are maximized in line E.

$BR_{DGI} = E$

Stage 2 : Best answer for the DGDA:

- DGDA often prefers E
- In certain extreme configurations, M becomes equivalent to the best responses

But on the dominant stable profiles: $BR_{DGDA} = E$

Step 3 : DGRAD's best answer

We compare DGRAD based on candidate profiles

Case (E , E):

- $U_3(E,E,F) = 0.5$, $U_3(E,E,M) = 9$, $U_3(E,E,E) = 15$, therefore $BR_{DGRAD} = F$

• Research into Global Nash Equilibrium

By analyzing the best response for each control room across all configurations (DGRAD case = F, M, or E), the stable profile emerges:

Balance profile: ($s_{\{DGI\}}=E, s_{\{DGDA\}}=E, s_{\{DGRAD\}}=M$)

The only profile that simultaneously checks the best answers is (E, E, M)

i.e. $\therefore DGI = \text{high DGRAD} = \text{high DGDA} = \text{low}$

Interpretation of equilibrium

This balance means that:

- DGI & DGDA: Balance requires an aggressive strategy. Given the nature of the flows (domestic taxes and customs), inaction by one agency would unduly benefit the other at the expense of the Treasury.
- DGRAD: The best response is Moderate. Simultaneous aggressiveness from the three tax authorities creates "tax harassment" which reduces marginal yield and increases litigation costs.

In summary: the balance obtained reveals an asymmetrical strategic structure in which the DGI and the DGDA capture most of the high tax potential, while the DGRAD optimizes its performance for a moderate or low strategy, in order to limit the effects of administrative overlap and intra-state tax competition.

Why this script is essential for this model:

- Input vector: The VectorAssembler function groups all the variables X_n into a single mathematical structure usable by the algorithm.
- Data scale: Without the MinMaxScaler, a \$1000 variation in revenue (X_1) would have more weight than a 50% variation in regularity (X_2). Normalization puts everyone on an equal footing (between 0 and 1).
- Scalability: Although we have 10 taxpayers here, this Spark code can process 10 million lines of tax returns or tax returns on a server cluster.

To complete this model, visualization is essential: it allows the translation of fuzzy degrees of membership into a readable risk map for decision-makers in financial management. Here is the Python code (using matplotlib and seaborn) to integrate after the Spark processing to visualize the segmentation and the payoff matrix.

9. DISCUSSION AND INTERPRETATION OF THE RESULTS

The experimentation carried out on a massive dataset (1,000,000 rows) and matrix simulations reveal major lessons for the governance of financial management in the DRC.

9.1. Fuzzy Segmentation Analysis: From Mass to Precision.

clustering (FCM) has made it possible to move beyond the binary view of the taxpayer. The mapping obtained via PySpark shows that nearly 22% of taxpayers fall into a "grey area" (a degree of belonging shared between the compliant segment and the at-risk segment).

Interpretation: This means that traditional control methods, based on rigid thresholds, either generate lost revenue (false negatives) or unnecessary harassment (false positives). The FCM model allows advertising agencies to tailor their response: a full audit for high-risk targets ($NIR > 0.8$) and a simple digital reminder for moderate profiles.

9.2. Nash Equilibrium: Non-Cooperative Stability

Applying game theory to the DGI-DGDA-DGRAD interaction shows that the equilibrium (A, A, M) is the only stable profile (Nash-Stable).

- The necessary aggressiveness (DGI/DGDA): Given that these two agencies often share the same tax base (turnover and import-export flows), any unilateral reduction in effort by one is perceived as an opportunity for fraud by the taxpayer, reducing the overall payoff.
- Strategic selectivity (DGRAD): The model demonstrates that if the DGRAD also adopts an aggressive strategy (A, A, A), marginal utility collapses. The cost of collecting administrative taxes becomes greater than the gain due to the financial strain placed on companies already burdened by the DGI and DGDA.

9.3. Validation of the Hypothesis.

The research hypothesis is confirmed. The coupling of artificial intelligence (segmentation) and economic modeling (Nash) makes it possible to transform a chaotic mass of heterogeneous data into a coherent decision-making structure. The system no longer simply collects data; it anticipates and optimizes.

6. Managerial and policy recommendations

In light of these results, three recommendations are formulated for decision-makers in the DRC:

- Standardization of the Data Repository: The implementation of a Unique Tax Identifier (IFU) is imperative so that the PySpark script can cross-reference the NIRs (Levels of Irregularity) between the three agencies in real time.
- Transition to Intelligent Recovery: Replace calendar-based control waves with alerts based on risk scores from the FCM model.
- Coordination by the Global Payoff : Establish a consultation platform where the Nash equilibrium serves as a basis for defining annual control quotas, thus avoiding administrative overlap on the same economic targets.

In short:

- The systematic interconnection of the DGI-DGDA-DGRAD databases.
- Automating the calculation of the "Fuzzy Risk Score" to prioritize tax audits.

10. CONCLUSION

This scientific article has demonstrated the effectiveness of a hybrid decision-making system for performance management. This hybrid FCM-Nash model offers a scientific solution to the problem of tax coordination in the DRC. Fuzzy segmentation allows for the precise targeting of "Tax Potential," while the Nash equilibrium provides a strategic roadmap: revenue optimization does not require maximum aggression from all actors, but rather a balanced distribution of control efforts to minimize administrative interference.

enhanced star data warehouse architecture, implemented in Hive , organized into staging , integration, and analysis zones. This architecture enables the consolidation of data from the French tax authorities (DGI), the customs administration (DGDA), and the tax administration (DGRAD), the execution of fuzzy taxpayer segmentation in PySpark , and then the modeling of a non-cooperative three-player game whose

solution is obtained through a Nash equilibrium. The results are ultimately materialized in analytical data marts and presented in Power BI as decision-making dashboards.

Future prospects include the integration of deep learning for real-time prediction of tax fraud.

Bibliographical References

- [1] ANOOP, KR, EMBAR, VR, DHAMAL, S., & BHAT, S. (2011). Pattern clustering using cooperative game theory, centenary conference, electrical engineering, Indian institute of science, BANGALORE.
- [02] Central Bank of Congo (BCC). (2025). Central Bank Report. <http://www.bcc.cd>
- [03] BARACHE, F. (2007). On the theory of evolutionary games and its applications in economics, Master's thesis in Applied Mathematics, A. Mira University of Bejaia, Algeria.
- [04] BATISTAKIS, Y., VAZIRGIANNIS, M., & HALKIDI, M. (2001). Clustering validity checking methods , *SIGMOD Record* .
- [05] BEZDEK, JC (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms* . Plenum Press .
- [06] BULO, S.R. (2009). A game-theoretic framework Similarity-based data clustering , Doctoral Thesis , Ca' Foscari University of Venice, Italy.
- [07] BULO, S.R., & PELILLO, M. (2020). Game-Theoretic Clustering: A Review.
- [08] CHUN, Y., & HOKARI, T. (2007). On the coincidence of the shapley value and the nucleolus in queuing problems , *Seoul Journal of Economics* .
- [09] DAVIES, DL, & BOULDIN, DW (1979). A cluster separation measurement , *IEEE Transactions on Pattern Analysis and Machine Intelligence* , pages 224-227.
- [10] DESKECO. (2025). The contribution of the DGI to public revenues as of March 3, 2025, 11:17, in Finances. <http://www.Deskeco.com>
- [11] DEVEZE, B., & FOUQUIN, M. (2004). Data mining C4.5 -DBSCAN, Course, EPITA School of Computer Engineering, France.
- [12] DUDA, RO, HART, PE, & STORK, DG (2001). *Pattern classification* , Wiley, New York.
- [13] DUNN, J. C. (1974). Well separated clusters and optimal fuzzy-partitions, *Journal of Cybernetics* , pages 95-104.
- [14] INTERNATIONAL MONETARY FUND (IMF). (2024). Democratic Republic of the Congo: Sixth Review under the Extended Credit Facility. Washington, DC
- [15] FORESTIER, G. (2010). Knowledge and collaborative clustering of complex multi-source objects, Doctoral Thesis , University of Strasbourg, France.
- [16] GIBBONS, R. (1992). *Game Theory for Applied Economists* . Princeton University Press .
- [17] GÖTHE-LUNDGREN, M., VÄRBRAND, P., & ENGEVALL, S. (1998). The traveling salesman game: An application of cost allocation in a gas and oil company, Department of Mathematics, Linköping Institute of Technology, Sweden.
- [19] HAN, EH, KARYPIS, G., & KUMAR, V. (1999). CHAMELEON: A hierarchical clustering algorithm using dynamic modeling, *IEEE COMPUTER Society* , pages 68-75.
- [20] HAN, J., & KAMBER, M. (2006). *Data mining: Concepts and techniques* , Management Systems (The Morgan Kaufmann Series in Data Management Systems), MORGAN KAUFFMAN, ISBN 1-55860-901-6.
- [21] HAN, J., KAMBER, M., & PEI, J. (2012). *Data Mining: Concepts and Techniques* . Morgan Kaufmann.

- [22] Extractive Industries Transparency Initiative (EITI-DRC).
(2020). EITI-DRC Flexible Report 2018, 2019 and 1st Half 2020. <https://www.itierdc.net>
- [24] LUO, Y., et al. (2023). Cooperative Game Theory for Big Data Clustering Optimization. *IEEE Transactions on Cybernetics* .
- [25] MAYER-SCHÖNBERGER, V., & CUKIER, K. (2013). *Big Data: A Revolution That Will Transform How We Live* .
- [26] DRC MINISTRY OF FINANCE. (2023). Revenue Mobilization Report.
- [27] Ministry of Finance, DRC. (2024). Strategic Plan for the Modernization of Financial Administrations (2024-2028). Kinshasa.
- [28] DRC MINISTRY OF FINANCE. (2025). Report on the modernization of information systems of financial authorities
- [29] [29] NASH, J. (1951). Non-Cooperative Games. *Annals of Mathematics* .
- [30] NSIKU, J., & MOKONDJI, P. (2022). Digitalisation of financial management and maximisation of public revenue in the DRC: Case of the DGI. *Congolese Journal of Economics* .
- [31] OSBORNE, M. J. (2004). *An Introduction to Game Theory* . Oxford University Press .
- [32] PEFA Secretariat. (2016). Report on the Evaluation of Public Financial Management Performance in the Democratic Republic of Congo according to the PEFA 2016 methodology.
- [33] RAI, P., & SINGH, S. (2010). A survey of clustering techniques, *International Journal of Computer Applications* (0975 - 8887) Volume 7- No.12, October 2010.
- [34] ROKACH, L., & MAIMON, O. (2002). Clustering methods, Department of Industrial Engineering, Tel- Aviv University .
- [35]) . Application of game theory for the definition, development and implementation of a clustering algorithm , Master's thesis, University of Béjaia , Algeria.
- [36] SCHMEIDLER, D. (1969). The nucleolus of a characteristic function game, *SIAM J. Appl. Math* , pages 1163 1170.
- [37] THUSOO, A. et al. (2009). Hive: A Warehousing Solution Over Hadoop . Clustering and Data Mining .
- [38] VARIAN, HR (2010). *Intermediate Microeconomics* .
- [39] XU, R., & WUNSCH, D. (2005). Survey of Clustering Algorithms.
- [40] ZADEH, LA (1965). "Fuzzy Sets". *Information and Control* .
- [41] ZAHARIA, M. et al. (2016). Apache Spark: A Unified Engine for Big Data Processing.
- [54] DUCANGE, P., FAZZOLARI, M., & MARCELLONI, F. (2020). An overview of recent distributed algorithms for learning fuzzy models in Big Data classification. *Journal of Big Data*, 7, 19. <https://doi.org/10.1186/s40537-020-00298-6>
- [55] ELKANO, M., SANZ, J. A., BARRENECHEA, E., BUSTINCE, H., & GALAR, M. (2020). CFM-BD: A distributed rule induction algorithm for building compact fuzzy models in Big Data classification problems. *IEEE Transactions on Fuzzy Systems*, 28(1), 163-177. <https://doi.org/10.1109/TFUZZ.2019.2900856>
- [56] ROY, P., & KHAN, M. (2021). Digitizing taxation and premature formalization in developing countries. *Development and Change*, 52(4), 855-877. <https://eprints.soas.ac.uk/35296>
- [57] AKRAM, A. W., & ALAMGIR, Z. (2022). Distributed fuzzy clustering algorithm for mixed-mode data in Apache SPARK. *Journal of Big Data*, 9(1). <https://doi.org/10.1186/s40537-022-00671-7>

- [58] ADVANI, A., ELMING, W., & SHAW, J. (2023). The dynamic effects of tax audits. *Review of Economics and Statistics*, 105(3), 545-561. https://doi.org/10.1162/rest_a_01101
- [59] HEBOUS, S., JIA, Z., LØYLAND, K., THORESEN, T. O., & ØVRUM, A. (2023). Do audits improve future tax compliance in the absence of penalties? Evidence from random audits in Norway. *Journal of Economic Behavior & Organization*, 207, 305-326. <https://doi.org/10.1016/j.jebo.2023.01.001>
- [60] BAH, M. (2024). Tax revenue mobilization and institutional quality in sub-Saharan Africa: An empirical investigation. *African Development Review*, 36(2), 201-221. <https://doi.org/10.1111/1467-8268.12752>
- [61] CHRISTIANSEN, T. G. (2024). Dynamic effects of tax audits and the role of intentions. *Journal of Public Economics*, 234, 105121. <https://doi.org/10.1016/j.jpubeco.2024.105121>
- [62] FIŠAR, M., KRČÁL, O., ŠPALEK, J., STANĚK, R., & TREMEWAN, J. (2026). A competitive audit selection mechanism with incomplete information. *International Tax and Public Finance*, 33(4), 1317-1334. <https://doi.org/10.1007/s10797-025-09937-1>