International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Tax Administration in Niger : A Data Mining Approach to Outlier Detection

Moussa Khane*[1], Harouna Naroua[2], Chaibou Kadri[3], Yacouba Moumouni[4]

Submitted: 12/01/2025 **Revised:** 25/02/2025 **Accepted:** 08/03/2025

Abstract*:* Developing countries face significant challenges in accurately forecasting tax revenues due to disparate databases and the presence of outliers in collected taxes. These anomalies can lead to inconsistencies in revenue predictions, impacting economic planning and policy decisions. This study applies the CRoss Industry Standard Process for Data Mining (CRISP-DM) framework to support Niger's tax administration in detecting and addressing outliers. Boxplot analysis and extreme value detection algorithms were utilized to visualize outliers, while the Interquartile Range (IQR) Machine Learning (ML) algorithm was employed to remove them. The dataset covers the period from January 2019 to December 2022. The current analysis identified significant outliers in June 2020 and December 2021 for Value Added Tax (VAT) and in August 2021 for Processing Tax and Salary (ITS). The study found that with outliers, VAT, ITS, and Profit Tax (ISB) accounted for 61.2% of total tax revenues, whereas without outliers, their combined contribution increased to 64.8%, highlighting the importance of accurate anomaly detection for better fiscal planning.

Keywords: CRISP-DM, Data mining, Machine Learning (ML), Outlier detection, Tax administration

1. Introduction

Tax revenue plays a crucial role in the development and implementation of a government's economic and social agenda[1] . Tax pressure, measured as the tax-to-GDP ratio, reflects a country's efforts in mobilizing internal resources. On average, OECD countries have a tax pressure of 34.0%, with France reaching 45.15%, whereas in African nations, it averages around 16.0%. In developing countries (DC), tax pressure tends to be relatively low; for example, in Niger, it stands at just 9.6% [2] [3].
Niger, a vast sub-Saharan country, faces significant challenges in tax administration due to limited human and material resources. The tax administration lacks a unified database, and existing tax data is fragmented across disparate systems [4] . Although a web application exists, it has not yet been deployed nationwide, leaving many Tax administration's offices branches in the regions without real-time data access. In these branches, revenue centralization is managed through phone calls or social media, leading to inefficiencies and delays.
Additionally, the absence of data mining tools prevents the tax administration from leveraging analytics for informed decision-making. There is no automated system to detect anomalies, making it difficult to improve tax revenue forecasts. As a result, actual annual tax revenues consistently fall short of projections made by the Planning Directorate of the General Directorate of Taxes.
Currently, only 62% of Niger's budget is funded through national revenue, with the remaining portion reliant on external financing, according to the European Union[5]. Strengthening data management, expanding digital coverage, and implementing advanced analytics could significantly enhance tax collection efficiency and revenue forecasting in the country.
With the advent of artificial intelligence, tax administration is becoming increasingly intelligent[6]. Machine Learning (ML) introduces advanced analytical and statistical tools that enhance the effectiveness of tax control. By detecting anomalies, these technologies enable governments to improve monitoring and oversight [7].
Despite significant progress in modernizing tax administrations worldwide, Niger still lags behind. No studies have been conducted to develop an anomaly detection scheme within the Nigerien tax administration system. To address this gap, this study focuses on detecting anomalies in the unified tax database using the CRISP-DM methodology.
For anomaly detection, visualization techniques, such as boxplots and extraction methods like the interquartile range are employed. The primary objective of this study is to identify tax data anomalies and assess their impact. A comparative analysis is conducted between the original database, which contains anomalies, and a cleaned version from which anomalies have been removed. This approach aims to demonstrate the significance of anomaly detection in improving tax data reliability.
The remainder of this study is structured as follows: Section 2 presents the literature review, Section 3 outlines the methodology, Section 4 discusses the results, and Section 5 provides the conclusion.

1 Département de Mathématiques et Informatique, Faculté des Sciences et Techniques, Université Abdou Moumouni, Niamey, Niger
2 Département de Mathématiques et Informatique, Faculté des Sciences et Techniques, Université Abdou Moumouni, Niamey, Niger
3Département de Mathématiques et Informatique, Faculté des Sciences et Techniques, Université Abdou Moumouni, Niamey, Niger
4 Electrical and Electronics Engineering, Higher Colleges of Technology, Ras Al Khaimah, United Arab Emirates
** Corresponding Author Email: khane12002@yahoo.fr*

2. Literature review

A review of existing literature on data mining methodologies for outlier detection, particularly within the context of tax administration, is essential for informing research relevant to the Niger tax administration. Subsection 2.1 provides an overview of data mining concepts, while Subsection 2.2 examines the Cross-Industry Standard Process for Data Mining (CRISP-DM). Subsection 2.3 discusses related studies and applications of data mining in tax administration.

2.1. Data mining

Data mining is the process of extracting and discovering patterns from large datasets using methods that intersect machine learning, statistics, and database systems[8] [9][10][11]. It is also considered a field focused on pattern discovery in vast amounts of data through a combination of techniques from machine learning, statistics, database systems, and computer science. Cheng noted that the development of data mining was not uniform. Rather, it was problem oriented according to different fields[8].

Outlier detection is a crucial task of safety-critical environments because outliers indicate abnormal operating conditions from which performance significantly deteriorates[12]. Several machine learning techniques can be used for outlier detection, often involving the visualization and analysis of extreme values. Outlier analysis plays a crucial role in this process, as most algorithms generate a numerical value representing the deviation from the normal pattern. By examining this deviation, outlier analysis enhances the detection algorithm, ensuring more accurate identification of anomalies[13]. The approach used is that of the interquartile range (IQR). Indeed, the concept of IQR is calculated by using the difference between the third quartile and the first quartile, which is represented by Eq. (1) as

$$IQR = Q3 - Q1 \quad (1)$$

Where, Q1 is the First quartile ; Q3 stands for Third quartile ; and IQR represents the Interquartile range.

Likewise, Eq. (1) is utilized to construct box and whisker plots. In addition, Eq. (2) is used to determine the value of outlier (s) at the upper bound of the IQR, whereas Eq. (3) computes the value of outlier (s) at the lower bound of IQR.

$$Outliers_{upperbound} = Q3 + 1.5 \cdot IQR \quad (2)$$

$$Outliers_{lowerbound} = Q1 - 1.5 \cdot IQR \quad (3)$$

Where, 1.5 represents a Multiplier in the IQR.

These thresholds upper and/or lower bounds identify outliers that deviate by approximately 1.5 times the interquartile range (IQR) from the lower or upper quartiles. Accordingly, Fig 1. illustrates the lower and upper bounds used to detect outliers based on the IQR[14].

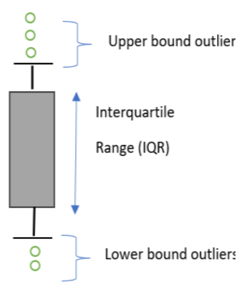

Fig 1. Interquartile range

2.2. Cross Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM is a freely available and widely used methodology in the field of data mining. It offers structured guidelines for the organized and transparent execution of projects by dividing all planned tasks into six interrelated phases. The latter phases are 1) Business Understanding – Defining project objectives and requirements from a business perspective ; 2) Data Understanding – Collecting and exploring data to identify quality issues and gain initial insights ; 3) Data Preparation – Cleaning, transforming, and organizing data for analysis ; 4) Modeling – Selecting and applying appropriate modeling techniques and tuning their parameters ; 5) Evaluation – Assessing the model's performance and ensuring it meets business goals ; and 6) Deployment – Implementing the model in a real-world setting for decision-making or process improvement. More details on the aforementioned phases can be found in [15][16][17][18][19]. Furthermore, CRISP-DM is the most widely used data mining methodology[20] and is broadly adopted across various industries[21]. It is also considered technology-neutral, industry-independent, and the de facto standard for data mining [15] .

The survey of the current literature has not revealed any physical links between the CRISP-DM phases and the data or database. Fig 2. shows the relationship between CRISP-DM phases and the database[16]. During the Modeling phase, data is processed and prepared to advance to the Evaluation phase. Similarly, data from the Evaluation phase is used to transition into the Deployment phase, where it is implemented and further populated.

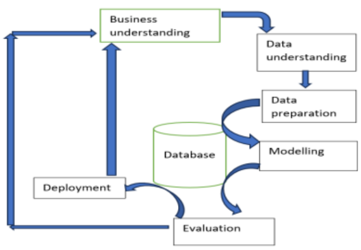

Fig 2. Data mining phases

2.3. Related work on Data Mining in Tax Administration

Tax administrations have a huge volume of taxpayer data. These data constitute a source of information especially for the audit department. Data mining makes it easier for tax administrations to manage these volumes of data. Indeed, data mining enables tax administrations not only to better analyze and predict data, but also to more effectively detect abnormal behavior [22]. The use of data science to which data mining belongs within tax administrations is not new. This mathematical and technical framework for interpreting reality of which statistics has historically been a major component is a key driver in the transformation of the modern state [23]. In the context of tax administration, artificial intelligence could be used to develop a multitude of functions such as data exploration commonly called data mining. Smart taxation has become an important development trend for tax collection and management in the future [7] . But [24] and [25] estimated that the key in combating and rectifying tax risk behaviors is accurate tax risk detection. Despite the studies concerning data mining in tax administration, it seems judicious to carry out work on the behavior of taxes and duties in relation to tax revenues in the event of anomalies and in the absence of these anomalies. This work, once carried out, will contribute to the decision support system of tax administrations.

Existing studies on outlier detection using data mining, along with related research in tax administration, provide a valuable foundation for conducting investigations in this domain — particularly within the context of the Niger tax administration. To the best of our knowledge, there is currently no published research specifically addressing the application of data mining techniques for outlier detection in the Nigerien tax administration.

3. Methodology

The anomaly detection system for the Nigerien tax administration was developed using the CRISP-DM framework. To better align the methodology with the specific context of tax administration, certain phases of the original framework were renamed. For instance, the "Business Understanding" phase was adapted to "Tax Administration Understanding," and "Data Understanding" was revised to "Understanding Tax Data.

3.1. Tax Administration Structure and Functioning

The Niger tax administration was established on October 13, 1983. Niger, a sub-Saharan country, has an area of 1,267,000 km². The administration faces a significant staff shortage, as retiring employees are not being replaced. The tax administration is fragmented into many departments like the Headquarter. The Taxpayer departments are subdivised in term of the type of tax to those based on the type of taxpayer (medium or large). There is also the Audit department, Compliance and enquiry department, Legislation department. The Tax administration comprises also office branches in the regions and subregions. Besides all this these there is some support department like the Information Technology department, Human resource department or Planing and forecasting department. Tax administration plays a crucial role in enabling the government to implement its budgetary policy while ensuring the collection of tax revenues. However, the tax administration's database is not an unique or let's an integrated database. The database comes from multiple sources, increasing the risk of errors during centralization and retrieval.

3.2. Understanding Tax Data

The research data consists of monthly tax records spanning from January 2019 to December 2022. This data was obtained from the Statistics Division, which is responsible for centralizing tax data within the administration. It includes all taxes actively managed by the Niger tax administration, which oversees a total of 61 different taxes. Tax collection may occur daily, weekly, or monthly, depending on the specific tax type.

Niger's tax system operates on a declarative basis. Some taxes, such as the communication tax, have been discontinued due to political decisions, while others particularly those related to the mining sector are inconsistently reported in the database. These taxes were excluded from the dataset used in this study. Table 1, structured in a spreadsheet format, summarizes the data collection process.

Table 1. Typical data collection file

Period	Tax revenue	ISB	…	VAT	ITS	…
January 2019	22335454577	2344455500		4444888211	9087777777	
…………						
December 2022	32335454597	2147455501		1344788210	5087737710	

designed to reflect a typical and realistic data collection process, for the months of January 2019 and december 2022, were synthetically generated solely for the purposes of this research. Consequently, these data points do not represent any actual office, business, or corporate entity, and any resemblance to real-world data is purely coincidental.

3.3. Preparation of Tax Data

According to Vandenbergh et al. (2022), the principal sources of tax revenue in Niger are the Value Added Tax (VAT), the Individual Tax on Salaries (ITS), and the Business License Tax (ISB). These taxes play a critical role in national revenue generation and form the core focus of this study. The dataset employed for the analysis spans a four-year period, from January 2019 to December 2022, providing a comprehensive temporal scope for evaluating tax trends and anomalies. The key input variables include monthly revenue figures for VAT, ITS, and ISB, along with total tax revenue. To improve interpretability and ensure consistency in the analysis, all monetary values were transformed into logarithmic form. This transformation aids in managing skewed distributions and stabilizing variance, thereby enhancing the robustness of statistical modeling techniques. The selected variables and preprocessing methods are aligned with best practices in fiscal data analysis and anomaly detection.

3.4. Modeling technique

To achieve the objectives of this study, a specific modeling technique along with a set of algorithms were proposed and utilized accordingly. Hence, the proposed modeling technique was the multiple linear regression and the algorithms were the 1) Boxplot and 2) Analysis of extreme values. The outliers are visualized using boxplot and the analysis of extreme values allows the removal of outliers. Equation (4) presents the multiple linear regression model utilized to establish a relation between Tax revenue and other taxes such as ITS, ISB, VAT, etc

So $Yi = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta p X p i + \varepsilon_i$ (4)

Where, β_0 is a constant ; X represents explanatory variable and each explanatory variable has its own β coefficient from 1 through p ; i is a time series variable ; and ε represents an error term.

It is also important to note that certain assumptions were made based on the hypothesis that the coefficients, β, for ISB, ITS, and VAT would be positive.

3.5. Assessment and deployment of the model

This study evaluates and deploys a tax forecasting model for the Nigerien tax administration, comparing datasets with and without outliers using the Ordinary Least Squares (OLS) method. Accurate outlier handling improves forecasting precision, with key taxes assessed at 1%, 5%, and 10% significance levels. The model is implemented in R Studio using Shiny for an interactive interface, and deployed with stakeholder agreement under strict data security protocols.

3.5.1. Model Assessment

The proposed model is evaluated through a detailed, step-by-step process to ensure alignment with the tax administration's objectives. The assessment focuses on a comparative analysis between two datasets: one containing outliers and the other with outliers removed. The OLS regression algorithm is applied to both datasets to evaluate model accuracy in forecasting tax revenue. This comparison is essential for understanding the impact of data quality on predictive performance. Statistical significance levels of

1%, 5%, and 10% are used to assess the influence of key tax variables — such as VAT, ITS, and ISB — on overall revenue generation. The results indicate that the presence of outliers negatively affects forecasting accuracy and leads to less reliable decision-making. The model that handles outliers effectively demonstrates improved performance, making it more suitable for use in fiscal planning. Therefore, careful outlier management is crucial in developing robust forecasting tools for tax administration.

3.5.2. Model Deployment

The model is deployed using R-Studio, utilizing the Shiny package — an open-source tool for building interactive web applications. Input variables are imported from a Microsoft Excel database and analyzed through a user-friendly interface that includes widgets such as text boxes, radio buttons, and drop-down menus. This interface allows stakeholders to interact with the model efficiently and visualize outputs in real time. The deployment process must be carried out in alignment with business objectives and requires formal approval from the tax administration. To ensure transparency and knowledge transfer, an experience document should be prepared, compiling reports and insights from all project contributors. Security is a critical component of the deployment phase. A comprehensive security framework must be implemented, including user authentication, access control, information classification, identification protocols, and audit trails. These measures are essential to safeguard sensitive fiscal data and ensure compliance with institutional policies and data governance standards.

4. Results and discussion

Recall: An outlier is an extreme value that significantly deviates from the overall distribution of a variable. The detection of outliers using the extreme values approach involves the following steps: first, determine the first quartile (Q1) using the formula QUARTILE(cell range, 1) and the third quartile (Q3) using QUARTILE(cell range, 3) for the variable of interest. Next, calculate the interquartile range (IQR), which represents the middle 50% spread of the data, as the difference between Q3 and Q1. Finally, the lower and upper bounds for identifying outliers are computed using standard formulas, as presented in Equations (2) and (3).

The visualization of outlier through the boxplot algorithm, Figure 3 shows three (03) outliers represented by a small circle which are three extreme values that deviate from the rest of data. Outliers have a significant impact on the performance and accuracy evaluation for the linear regression model. Outliers have a negative impact on tax revenue forecasting; the negative impact is evaluated during the execution of ordinary least square algorithm of the two dataset. These outliers can be caused by data entry error, tax calculation error, the declarative system of Niger tax administration. The ITS has one (01) outlier which is represented by a small circle at its upper bound and the VAT has (02) outliers. But the Tax revenue and ISB does not have outlier. The periods and the values or amounts corresponding to these outliers are going to be determined to allowed the top management to instruct the audit department to check the reasons of occurrence of the outliers.

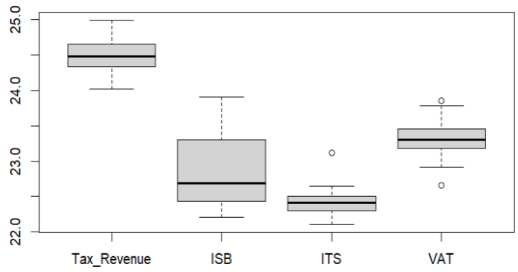

Fig.3. Graphical visualization of the initial database

Table 2. Quartile of taxes

Quartile	Tax Revenue	ISB	ITS	VAT
1st quartile	24.34	22.44	22.30	23.19
3rd quartile	24.66	23.30	22.50	23.45

Table 3. Distribution of outliers

Period	Taxes	Monthly amount
June 2020	VAT	22.66
August 2021	ITS	23.11
December 2021	VAT	23.86

To remove the outlier, two taxes (VAT and ITS) draw the attention. Table 2 shows quartile representation for Tax revenue, ITS, VAT, ISB. These quartiles, i.e., the first and third quartile allow the calculation of the extreme values by the application of Eq. (1), Eq. (2), and Eq. (3), as seen in section 2.1

Proceeding first by manual calculation of outliers :

4.1. Outliers calculation for ITS

In the case of ITS, Q1= 22.30 and Q3=22.50, while applying Eq. (1) the IQR value is 0.2. Eq. (2) gives an upper bound of 22.8 and Eq. (3) gives a lower bound of 22. Any value(s) outside these bound (lower or upper) is considered as outlier. In the dataset, only the value 23.11 exceeds the expected range and is therefore considered an upper-bound outlier.

4.2. Outliers calculation for VAT

In the case of VAT, Q1= 23.19 and Q3=23.45, while applying Eq. (1) the IQR value is 0.26. Eq. (2) gives an upper bound of 23.84 and Eq. (3) gives a lower bound of 22.8. Any value(s) outside these bound (lower upper) is considered as outlier. In the dataset, two (02) values which are 22.66 and 23.86 fall outside the expected range and are identified as outliers for VAT. The value 22.6 is under the lower bound (22.8) and 23.86 is higher than the upper bound (23.84).

4.3. Outliers detection for ITS and VAT

Table 3 shows the monthly amount of outliers for the taxes and the periods where the outliers occur. The values of outliers in Table 3 were produced when running the extreme value algorithm in R-

Studio. Rstudio gives the value of outliers and their corresponding period for Niger Tax Administration. The value 22.66 in log form corresponds to a VAT outlier recorded in June 2020, while 23.86 in log form represents a VAT outlier recorded in December 2021. Additionally, the value 23.11 in log form corresponds to an ITS outlier recorded in August 2021. Due to the presence of outlier in Niger tax administration dataset, ITS represents 21.15% of the variations of the tax revenue where in the dataset without outlier ITS represents 28.03%. Again, in the dataset with outlier the VAT represents 16.07% of the variations of outliers where in the dataset without outliers it represents 18.3% of the variation of tax revenue. Without outliers ITS and VAT increase their influence to the tax revenue. Thus, outlier decrease the influence of ITS and VAT on tax revenue.

It is important to note that all amounts presented on the x-axis in Figure 4 were expressed in logarithmic form to improve data normalization, reduce skewness, and enhance the accuracy of statistical analysis. Figure 4 presents the visualization of the database after removing outliers and applying the same boxplot algorithm. The Ordinary Least Squares (OLS) analysis indicates that, in the dataset containing outliers, 61.2% of the variation in tax revenue is explained by ISB, ITS, and VAT. In contrast, the dataset without outliers shows a higher explanatory power of 64.8%, highlighting improved model accuracy and significance.

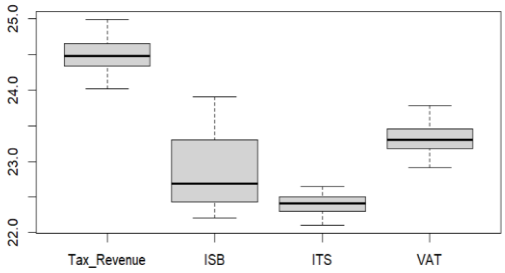

Fig.4. Database visualization without outliers (database after removing outliers)

4.4. Proposed work

Section 2.2 proposes the relationship or link between phases and the dataset (database) while other works show the link between phases only.

The application of Eq. (5) gives the proposed linear regression model equation for the database without outlier as

$$Tax_{Revenuei} = -2.16 + 0.25ISB_i + 0.71ITS_i + 0.21VAT_i + \varepsilon_i \quad (5)$$

The intercept (or constant term), which represents the expected mean value of the response variable when all predictor variables are set to zero, is negative with a value of –2.16. This indicates that the regression line crosses the y-axis below zero, suggesting a baseline tax revenue deficit in the absence of the specified tax inputs. The regression coefficients for the tax variables are all positive: ISB = 0.25, ITS = 0.71, and VAT = 0.21. These values imply a positive relationship between each tax type and overall tax revenue — i.e., as ISB, ITS, and VAT increase, the total tax revenue also increases proportionally. Among them, ITS has the strongest influence on revenue.

Figure 5 presents the flowchart of the proposed framework. It outlines the entire process, from the input of tax data to the dataset analysis, illustrating the structured steps required to ensure the model's effectiveness within the Nigerien tax administration.

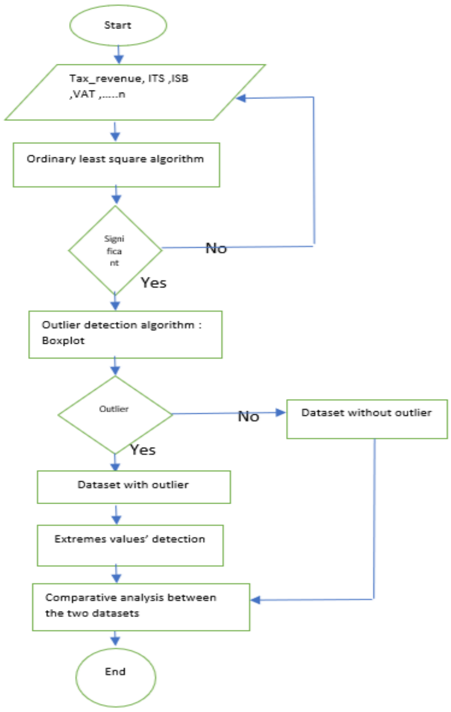

Fig. 5. Flowchart of the work

5. Conclusion

This study successfully consolidated disparate databases into a unified dataset, enabling a more comprehensive analysis of tax data in Niger. An anomaly detection system was developed to serve as a decision-support tool for the Nigerien tax administration, improving tax revenue forecasting and financial planning. Given the government's reliance on external aid to balance its budget, this system provides a data-driven approach to strengthening fiscal management.

The research employed the CRISP-DM methodology alongside linear regression, boxplot analysis, and the interquartile range (IQR) method to detect and remove anomalies. The results demonstrate that eliminating outliers enhances the accuracy and efficiency of tax data analysis.

A key finding of this study is the significant impact of accurate data processing on revenue predictions. In the cleaned dataset, key tax variables are 1. Value Added Tax (VAT), 2. Individual Tax on Salary (ITS), and 3. Business Income Tax (ISB) accounted for 64.8% of the variance in tax revenue. In contrast, in the dataset with anomalies, these variables explained only 61.2% of the variance. This underscores the importance of anomaly detection in ensuring reliable fiscal data.

By integrating advanced data analytics into tax administration, Niger can improve revenue collection, reduce reliance on external funding, and enhance overall financial stability. Further research could explore additional machine learning techniques to refine anomaly detection and optimize tax revenue forecasting.

References

[1] A. T. Olufemi, O. Jayeola, A. S. Oladele, and A. O. Naimot, 'Tax revenue and economic growth in Nigeria', Scholedge International Journal of Management & Development, vol. 5, no. 7, pp. 72–85, 2018.

[2] 'statistiques-recettes-publiques-afrique-niger.pdf', 2024, [Online]. Available:

https://www.google.com/search?q=statistiques-recettes-publiques-afrique-niger.pdf
[3] F. Marro-Dauzat, ‘La pression fiscale dans l’Union européenne’, 2023,[Online]. Available: https://www.touteleurope.eu/economie-et-social/la-pression-fiscale-dans-l-union-europeenne/
[4] L. Lin, ‘Application of big data model in financial taxation management’, Scientific Programming, vol. 2021, no. 1, p. 7001456, 2021.
[5] ‘L’Union Européenne et le Niger | EEAS’, 2023, [Online]. Available: https://www.eeas.europa.eu/niger/lunion-europeenne-et-le-niger_fr?s=113
[6] F. Y. Mpofu, ‘Prospects, challenges, and implications of deploying artificial intelligence in Tax Administration in developing countries’, Studia Universitatis Babes Bolyai-Negotia, vol. 69, no. 3, pp. 39–78, 2024.
[7] I. El Yamlahi, N. B. Amine, and H. El Ghazlani, ‘La contribution de l’intelligence artificielle au developpement de la Direction Generale des Impots au Maroc’, Revue Economie & Kapital, vol. 1, no. 21, 2022.
[8] J. Cheng, ‘Data-mining research in education’, arXiv preprint arXiv:1703.10117, 2017.
[9] M. Koklu, K. Sabancı, and M. F. Unlersen, ‘Classification of Heuristic Information by Using Machine Learning Algorithms’, International Journal of Intelligent Systems and Applications in Engineering, vol. 4, no. Special Issue-1, pp. 252–254, 2016.
[10] J. Han, J. Pei, and H. Tong, Data mining: concepts and techniques. Morgan kaufmann, 2022.
[11] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, ‘Data mining techniques and applications–A decade review from 2000 to 2011’, Expert systems with applications, vol. 39, no. 12, pp. 11303–11311, 2012.
[12] S. Moro, R. Laureano, and P. Cortez, ‘Using data mining for bank direct marketing: An application of the crisp-dm methodology’, 2011.
[13] G. Mariscal, O. Marban, and C. Fernandez, ‘A survey of data mining and knowledge discovery process models and methodologies’, The knowledge engineering review, vol. 25, no. 2, pp. 137–166, 2010.
[14] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, ‘An outliers detection and elimination framework in classification task of data mining’, Decision Analytics Journal, vol. 6, p. 100164, 2023.
[15] C. Shearer, ‘The CRISP-DM model: the new blueprint for data mining’, Journal of data warehousing, vol. 5, no. 4, pp. 13–22, 2000.
[16] R. Wirth and J. Hipp, ‘CRISP-DM: Towards a standard process model for data mining’, in Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, Manchester, 2000, pp. 29–39.
[17] I. Wowczko, ‘A case study of evaluating job readiness with data mining tools and CRISP-DM methodology’, International Journal for Infonomics, vol. 8, no. 3, pp. 1066–1070, 2015.
[18] T. Darmawan, A. S. Birawa, E. Eryanto, and T. Mauritsius, ‘Credit classification using CRISP-DM method on Bank ABC customers’, International Journal of Emerging Trends in Engineering Research, vol. 8, no. 6, 2020.
[19] S. Studer *et al.*, ‘Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology’, Machine learning and knowledge extraction, vol. 3, no. 2, pp. 392–413, 2021.
[20] S. Sharma, K.-M. Osei-Bryson, and G. M. Kasper, ‘Evaluation of an integrated Knowledge Discovery and Data Mining process model’, Expert Systems with Applications, vol. 39, no. 13, pp. 11335–11348, 2012.
[21] G. C. Onwubolu and D. Davendra, Differential evolution : A handbook for global permutation-based combinatorial optimization, vol. 175. Springer, 2008.
[22] D. Micci-Barreca and S. Ramachandran, ‘Improving tax administration with data mining’, White paper. Elite Analytics LLC, 2004.
[23] C. L. Roth, ‘La science des données numériques au service du contrôle fiscal français - Réflexions sur l’Algocratie’.
[24] Q. Zheng, Y. Xu, H. Liu, B. Shi, J. Wang, and B. Dong, ‘A survey of tax risk detection using data mining techniques’, Engineering, vol. 34, pp. 43–59, 2024.
[25] M. G. Allingham and A. Sandmo, ‘Income tax evasion : A theoretical analysis’, Journal of public economics, vol. 1, no. 3–4, pp. 323–338, 1972.
[26] P. Vandenberghe, J. L. Carvajal, M. Kabaka, and T. Pichot, ‘Rapport d’Évaluation de la Performance’, 2022, [Online]. Available: https://www.tadat.org/assets/files/Niger_Replublic_TADAT_%202022.pdf