

**Classification automatique de données du TENASOSP en RD Congo :
analyse exploratoire et expérimentale,**

**Automatic classification of TENASOSP data in the Democratic Republic
of Congo: exploratory and experimental,**

KABONGO Bandowe Pierre

Enseignant chercheur et Doctorant

Université Pédagogique de Kananga

République Démocratique du Congo

Date de soumission : 01/09/2025

Date d'acceptation : 22/10/2025

Pour citer cet article :

KABONGO-Bandowe, P., (2025). «Classification automatique de données du TENASOSP en RD Congo : analyse exploratoire et expérimentale», Revue Internationale du chercheur, « Volume 6 : Numéro 4 » pp : 12- 53

Résumé

Le recours aux nouvelles techniques de l'intelligence artificielle (IA) a été conseillé par plusieurs recherches (BADACHE, I. et al, 2024 ; Bates et al., 2020; Ilkka, 2018; Niemi, 2021; Swiecki et al., 2022; Zawacki-Richter et al., 2019; Zhai et al., 2021) appuyant ainsi l'idée selon laquelle, les récentes applications de l'IA améliorent efficacement le processus d'enseignement. Pour d'autres recherches (Akyuz, 2020), ces applications sont capables de fournir un enseignement personnalisé et recommander des ressources aux apprenants selon leurs niveaux de compétence, de leurs besoins et de leurs styles d'apprentissage. Ces applications sont capables d'analyser des données d'apprentissage des apprenants, comme leurs résultats aux différents tests, leurs interactions avec les ressources pédagogiques ainsi que les données de leurs comportements, afin de fournir aux enseignants des informations d'aide à la décision pour que ceux-ci adaptent leurs enseignements et leurs pratiques pédagogiques (Ouyang et al., 2023; Sousa et al., 2021). Dans d'autres recherches (Kuhail et al., 2023), les chatbots éducatifs sont encouragés car, ils utilisent des techniques d'IA pour répondre aux questions des apprenants en fonction de leurs besoins spécifiques. Dans le prolongement normal des recherches ci-haut citées, la présente recherche se base sur une analyse exploratoire et expérimentale du processus d'orientation scolaire dénommé TENASOSP en RD Congo à partir de données collectées dans des bases de données de la Division Provinciale de l'Education Nationale et Nouvelle Citoyenneté (EDUC-NT). Cette étude analyse toutes les étapes du processus de ce test. Elle conduira à l'implémentation des classificateurs automatiques d'orientation scolaire des candidats à ce test, lesquels classificateurs constituent des outils d'IA. Les résultats antérieurs issus d'un échantillon stratifié de 3000 candidats au TENASOSP révèlent des disparités selon les milieux urbains et ruraux, des tendances évolutives en matière de changement d'orientation, ainsi qu'un besoin croissant de réformes pour rendre ce processus plus transparent, participatif et adapté aux aspirations des élèves.

Mots clés : Classification automatique, IA, Analyse exploratoire, Analyse expérimentale, Orientation scolaire, TENASOSP, Changement d'orientation, Education,

Abstract

The use of new artificial intelligence (AI) techniques has been recommended by several studies (BADACHE, I. et al, 2024; Bates et al., 2020; Ilkka, 2018; Niemi, 2021; Swiecki et al., 2022; Zawacki-Richter et al., 2019; Zhai et al., 2021), supporting the idea that recent AI applications effectively improve the teaching process. According to other research (Akyuz, 2020), these applications are capable of providing personalized teaching and recommending resources to learners based on their skill levels, needs, and learning styles. These applications are capable of analyzing learners' learning data, such as their test results, their interactions with educational resources, and their behavioral data, in order to provide teachers with decision-making support information so that they can adapt their teaching and educational practices (Ouyang et al., 2023; Sousa et al., 2021). In other research (Kuhail et al., 2023), educational chatbots are encouraged because they use AI techniques to answer learners' questions based on their specific needs. As a natural extension of the above-mentioned research, this study is based on an exploratory and experimental analysis of the educational guidance process known as TENASOSP in the Democratic Republic of Congo, using data collected from databases belonging to the Provincial Division of National Education and New Citizenship (EDUC-NT). This study analyzes all stages of the test process. It will lead to the implementation of automatic school guidance classifiers for candidates taking this test, which are AI tools. Previous results from a stratified sample of 3,000 TENASOSP candidates reveal disparities between urban and rural areas, changing trends in terms of orientation, and a growing need for reforms to make this process more transparent, participatory, and adapted to students' aspirations.

Keywords : Automatic classification, Artificial intelligence, Exploratory analysis, Experimental analysis, Academic orientation, TENASOSP, Change of orientation, Education.

Introduction

L'orientation scolaire constitue un enjeu central dans le parcours éducatif des élèves dans le monde entier en général et ceux de la RD Congo en particulier, notamment à la fin du premier cycle du secondaire. Le test national d'orientation scolaire et professionnelle (TENASOSP en sigle) a été conçu pour guider les élèves finalistes de l'éducation de base vers des choix d'orientation scolaire conformes à leurs aptitudes, intérêts et résultats scolaires (Législateur Congolais, 2014). Toutefois, plusieurs défis persistent quant à son efficacité, sa mise en œuvre, et surtout son acceptabilité par les élèves et leurs familles. L'orientation scolaire et professionnelle est considérée comme une action collective, des interactions entre différents principaux acteurs locaux (élèves, parents d'élèves et agents de l'institution scolaire) dans les relations entre les usagers et les agents de l'institution scolaire d'une part, dans les relations entre ces agents dans le cadre du processus d'orientation, d'autre part (METRAL, J-F., 2018, pp.4-5) ; (LAZIB Ep., 2020, p.17) ; (Masson, 1997, p.131), tous, cités par METRAL. Toutefois, dans le contexte du TENASOSP, ces trois acteurs ne sont pas associés par le pouvoir organisateur de ce test. Dans un contexte où les inégalités géographiques, socioéconomiques et culturelles influencent largement les décisions d'orientation scolaire et professionnelle, il est impératif d'examiner les pratiques actuelles du test, voir s'il y a lieu d'y apporter un plus sur le plan organisationnel ou fonctionnel. Notre problématique s'articule au tour des questions suivantes : Pourquoi tous les lauréats orientés au TENASOSP ne respectent-ils pas les avis de l'orientation scolaire émis par les conseillers d'orientation ? Quelles sont les déterminants d'un choix d'une orientation scolaire dans le contexte du TENASOSP ? Quel algorithme d'apprentissage automatique faut-il implémenter qui puisse mieux prédire l'orientation scolaire d'un candidat ? Il semble que, tout le processus d'orientation scolaire, de la collecte de données au traitement et aux recommandations des orientations scolaires serait manuel. L'objectif principal de cette recherche est de construire un modèle de prédiction automatique de l'orientation scolaire et professionnelle en passant par l'analyse descriptive de données antérieures du TENASOSP. Les objectifs spécifiques de cet article sont les suivants : le premier, collecter et analyser les données antérieures du TENASOSP afin d'en découvrir les points forts à consolider et les points faibles à remédier ; le deuxième, identifier les déterminants qui prédisent mieux l'orientation scolaire de candidat

au TENASOSP ; le troisième, identifier les facteurs qui favorisent ou entravent l'adhésion aux décisions d'orientation scolaire émanant des conseillers d'orientation scolaire afin d'éclairer les perspectives d'amélioration de ce processus ; le quatrième, comparer les modèles de Machine Learning afin d'en découvrir celui qui prédit mieux l'orientation scolaire et professionnelle dans le contexte du TENASOSP et enfin implémenter le modèle valide de Machine Learning. Pour atteindre ces objectifs, notre recherche suit les approches : exploratoire mixte, expérimentale et orientée données. Notre population cible est constituée des élèves de 1^{ère} secondaire orientés au TENASOSP pour les éditions 2021-2024. Les données de notre étude seront collectées à partir des bases de données de la Division Provinciale de l'Education Kasaï-Central 1 et cela selon un échantillonnage stratifié, en fonction de milieux scolaires (ruraux ou urbains) et des éditions de TENASOSP. Pour répondre à la problématique soulevée ci-haut, notre recherche est organisée en trois parties distinctes :

- Une revue de la littérature est présentée, synthétisant ainsi les recherches antérieures et récentes afin d'établir les spécificités et la valeur ajoutée de la présente étude;
- Une méthodologie de recherche, centrée sur les données antérieures du TENASOSP et sur la prédiction automatique de l'orientation scolaire et professionnelle des candidats, laquelle méthodologie part de la collecte des données à la production des résultats ;
- Des résultats exploratoires et expérimentaux suivis des analyses, des discussions et des limites d'application du modèle élaboré et proposé.

2. Revue de littérature

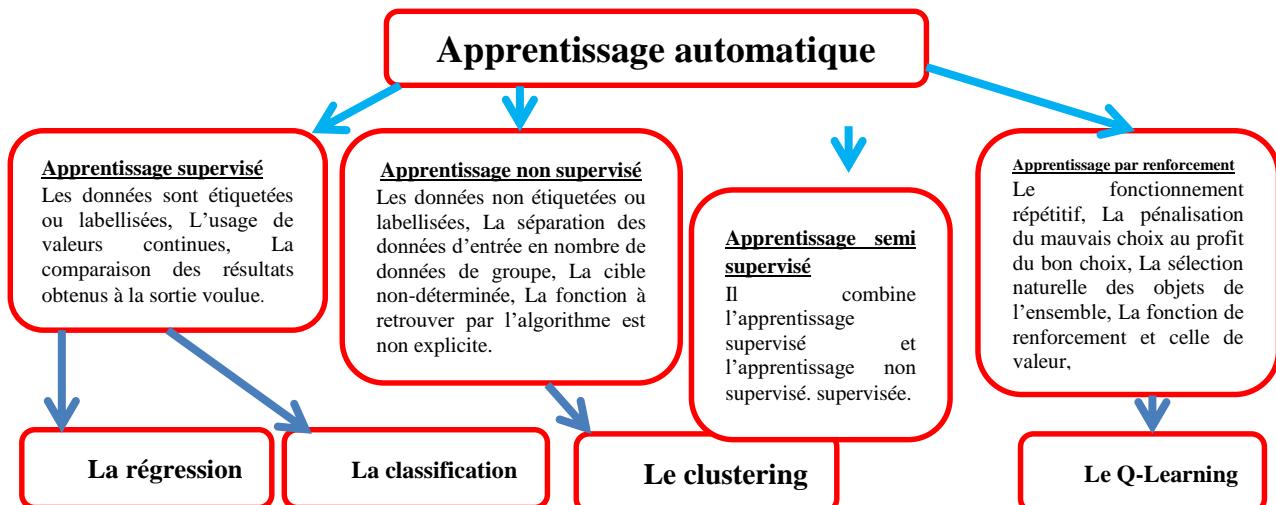
2.1 Revue de littérature théorique

La présente section donne un coup de torche sur des concepts, modèles, théories et cadres d'analyse élaborés par d'autres chercheurs ayant trait au processus d'orientation scolaire et professionnel en général et au TENASOSP comme étude de cas en particulier, en RD Congo. Ces concepts sont entre autre l'apprentissage automatique, la classification automatique, l'intelligence artificielle, l'analyse exploratoire, l'orientation scolaire, le TENASOSP, Changement d'orientation et Education, etc.

2.1.1 Apprentissage automatique

L'apprentissage automatique ou Machine Learning est un processus de mise en œuvre des programmes informatiques capables d'apprendre seuls à partir des données existantes (AZENCOTT, C-A., 2017, pp.10-11). L'apprentissage automatique est défini par d'autres chercheurs comme une partie de l'intelligence artificielle qui donne aux ordinateurs et aux appareils électroniques la capacité d'apprendre à partir des données afin de résoudre des problèmes complexes de la société sans être explicitement programmé en avance (MOKDAR, S., 2024, p.275; Joshi, 2020). Pour Dong, un apprentissage automatique consiste à la mise en œuvre d'algorithmes et de modèles statistiques qui permet aux systèmes informatiques grâce à l'exploitation des données et à des processus d'apprentissage itératifs de prédire les événements (Dong et al., 2024). Les algorithmes d'apprentissage automatique permettent de traiter les problèmes de classification, de régression et de clustering. Nous distinguons ci-dessous les quatre catégories d'apprentissage automatique.

Figure N°1 : Types d'apprentissage automatique



Source : Auteur

Dans le contexte actuel de notre étude sur le TENASOSP, l'apprentissage supervisé peut être utilisé pour prédire le profil cognitif d'un candidat en fonction de ses réponses à ce test et aux autres évaluations déjà présentées dans le passé. Il est généralement adapté par le fait que, dans la plupart des cas, les profils d'orientation scolaires et professionnelle (Pédagogie,

technique, scientifique, etc.) des élèves sont déjà connus dans les données historiques et ensuite, le modèle apprend à prédire le bon profil à partir de données étiquetées, comme les résultats aux tests, les avis d'orientation des candidats et des conseillers d'orientation, les performances scolaires, etc. Chaque candidat ayant présenté ce test, représente une observation étiquetée car, son profil final est connu, par exemple la pédagogie, la scientifique, la littéraire, la commerciale et gestion, ou toute autre option technique). En entraînant un modèle supervisé sur ces données historiques, il devient possible de détecter automatiquement les profils futurs des candidats, d'offrir une orientation personnalisée sans nécessiter une correction manuelle, de réduire les erreurs et d'améliorer la rapidité et la précision du processus d'orientation. Dans ce cas, les algorithmes de l'apprentissage supervisé les plus couramment utilisés sont entre autres l'arbre de décision, le SVM, la Forêt aléatoire, le k-NN, le réseau de neurones et deux nouvelles méthodes de classification supervisée développées en collaboration entre biologistes, médecins et mathématiciens pour une utilisation en métabolomique : la méthode PD-CR (Primal Dual for Classification with Rejection) et un autoencodeur supervisé (CHARDIN D., 2023, p4).

2.1.1.1 Modèles d'apprentissage automatique

Le modèle d'apprentissage consiste à construire des raisonnements sous forme d'algorithmes en vue de rendre possible des prédictions en utilisant des variables dites indépendantes pour déterminer l'une des variables cibles. Une prédition correspond donc à des fonctions mathématiques sur des variables prédictives d'une observation X. Dans ce domaine, il existe plusieurs algorithmes qui servent de créer des modèles Machine Learning. Dans notre recherche, nous pouvons détailler quelques-uns de type apprentissage supervisé utilisés pour construire notre modèle.

a. La régression linéaire

La régression est une méthode statistique qui modélise la relation entre une variable dépendante contenue ou cible (prédictive) et une ou plusieurs variables indépendantes(s) ou prédicteurs (explicatives) en ajustant une droite aux données observées. Dans le contexte du TENASOSP, la régression peut être utilisée pour prédire le score d'aptitude scientifique en fonction des notes réalisées en dessin scientifique ou en mathématique.

b. Arbre de décision

Nous appelons arbre de décision, un algorithme d'apprentissage supervisé utilisé pour résoudre des problèmes de classification ou de régression (KANDUKI, M., et al., 2023, pp.3). Il fonctionne en divisant récursivement l'espace des données en sous-ensembles, à l'aide des conditions logiques sur les variables d'entrée jusqu'à la prédiction. Chaque nœud de cet arbre teste respectivement une condition sur une variable d'entrée et chaque subalterne de ce nœud correspond à une réponse possible à cette condition. Les feuilles de l'arbre correspondent à une étiquette. Pour prédire l'étiquette d'une observation, l'algorithme suit les réponses aux tests depuis la racine de l'arbre, et il retourne l'étiquette de la feuille à laquelle il arrive. L'arbre de décision est capable de prédire des valeurs numériques tout comme des valeurs catégorielles. Dans le contexte du TENASOSP, l'arbre de décision peut être utilisé pour classifier les candidats selon leurs profils scolaires ou leurs filières recommandées.

c. Random forest

Le Random forest (forêt aléatoire) est une méthode d'apprentissage statistique très populaire, introduite en 2001 par BREIMAN (MOKDAR, S., 2024, p275 ; Breiman, 2001). Cette est très répandue pour le traitement des tâches de prédiction. Son principe réside dans la construction d'un vaste ensemble d'arbres de décision qui ne sont pas corrélés, suivis de la prise de la moyenne de leurs prédictions. Sa prédiction repose sur les prédictions de chaque arbre de décision individuel (Madhumita, et al., 2022). Le Random Forest (RF) est basé sur une combinaison de plusieurs arbres de décision. Le RF est un remède pour les arbres de décision en ce sens que ces derniers présentent un sur-ajustement des données d'apprentissage.

d. Machines à vecteurs de support

Les machines à vecteurs de support (Support Vector Machines en anglais (SVM)), est un algorithme d'apprentissage automatique supervisé utilisé pour des tâches de classification ou de régression. Cet algorithme est focalisé sur un algorithme linéaire proposé par Vladimir Vapnik et Aleksandr Lerner en 1963, et il cherche à trouver l'hyperplan optimal séparant ainsi les classes de la plus grande marge possible entre les points de données des différentes classes. Le principe fondamental de SVM consiste à intégrer à l'estimation de contrôle de la complexité, c'est-à-dire le nombre de paramètres qui est associé au nombre de vecteurs supports (KANDUKI, M., et al., 2023, p.3).

e. Naïve Bayes

Le Naïve Bayes est un algorithme de classification automatique supervisé basé sur le théorème de Bayes. Cet algorithme suppose que, toutes les variables d'entrée (X) sont indépendantes les unes des autres, ce qui justifie le mot naïve (KANDUKI, M., et al., 2023, p.4). Il prédit la classe de données la plus probable d'un exemple donné tout en calculant les probabilités conditionnelles des classes en fonction des caractéristiques. Cet algorithme est particulièrement utile pour la résolution de problèmes de classification textuelle.

f. Réseau de neurones

Le réseau de neurones artificiels est un algorithme inspiré du fonctionnement du cerveau humain (KANDUKI, M., et al., 2023, p.4). Il est composé de couches de neurones ou unités interconnectées entre-eux, capables d'apprendre des représentations complexes à partir des données d'entrée. Chaque neurone applique une fonction mathématique sur ses variables d'entrée pondérées puis transmet le résultat aux neurones suivants ainsi de suite jusqu'à obtenir une prédiction.

Dans le contexte du TENASOSP, les couches d'entrée sont des scores, les réponses au test, etc. Les couches cachées interviennent dans le traitement et la transformation et les couches de sortie sont des classes prédictives ou scores.

2.1.2 Intelligence artificielle

L'intelligence artificielle (IA) rassemble les techniques qui permettent aux ordinateurs et appareils électroniques de simuler et de reproduire l'intelligence humaine. L'IA a été inventée par Marvin Minsky et son collègue du MIT John McCarthy, dans une démarche clairement pluridisciplinaire, vers les années 1956 (Benoît GEORGES, B., 2019, p.4). Le concept IA apparaît pour la toute première fois à lors d'un colloque scientifique organisé à l'été 1956, à l'université de Dartmouth, dans le New Hampshire (États Unis). La notion d'IA est apparue dans le langage courant ces dernières années, mais nous pouvons considérer qu'elle existe depuis que l'ordinateur tourne des algorithmes qui ne sont que des reproductions du raisonnement humain (MOKDAR, S., 2024, pp.273-274). L'IA englobe plusieurs sous-domaines permettant aux appareils électroniques de prédire et d'interpréter le langage tels que l'apprentissage automatique et le traitement du langage naturel (Djellaba, et al., 2023).

Son objectif est de reproduire l'intelligence afin de résoudre des problèmes complexes de la société, en passant par la modélisation de l'intelligence humaine en tant que phénomène, comme cela pourrait être fait dans plusieurs domaines de la vie courante. L'apprentissage automatique (Machine Learning) est un sous-domaine de l'IA permettant aux systèmes d'apprendre et de s'améliorer automatiquement à partir de l'expérience sans un programme classique (LAHRACHE R. & BEKKAOUI A., 2024). C'est un domaine de recherche visant à reproduire par des systèmes artificiels les diverses capacités cognitives de l'être humain par la création d'algorithmes capable de réaliser certaines tâches humaines intelligentes (ZOUHRI, A., 2019, p.5)

2.1.3 Analyse exploratoire de données

L'analyse exploratoire des données (Exploratory Data Analysis, en anglais ou EDA), terme proposé par le statisticien J.W. Tukey, à l'Université de Princeton et des Laboratoires AT&T Bell (WANIEZ, P., 1991). L'EDA insiste sur l'inadaptation des hypothèses sous-jacentes à la statistique classique, hypothèses souvent trop fortes au regard de la complexité des univers analysés. Elle cherche à prendre en compte les anomalies ou les cas extrêmes, trop souvent considérés comme aberrants, car s'ajustant mal aux lois et principes statistiques. L'EDA est adaptée à l'analyse exploratoire participant ainsi à l'amélioration de l'interprétabilité des résultats obtenus par la sélection des variables importantes pour chaque classe. De plus, à la manière des techniques de coclustering, elle organise les moyennes de représentation de chaque classe par blocs qui permet de mettre en évidence les variables communes, celles qui discriminent, et aussi celles qui sont inutiles pour définir le clustering (BARBARO, F., 2022, p.6). Dans le contexte du TENASOSP, ces données sont entre autre l'âge, le sexe, la sous division, l'école, les résultats au test psychotechnique, contexte familial et institutionnel de provenance qui participent dans la prédiction de l'orientation scolaire.

2.1.4 Orientation scolaire

L'orientation scolaire et professionnelle est un processus de réflexion permettant à un élève assisté par les autres acteurs d'orientation, de choisir une orientation entre les différentes filières possibles et existants dans un système éducatif pour construire son avenir scolaire et professionnel (KAVUGHO, F., 2024).

L'orientation scolaire est un processus qui se trouve à l'intersection de l'action des systèmes éducatifs, des dispositions, des attributs sociaux et des culturels des élèves (DORAY, P., 2020, p.2). Elle est pour un élève, la détermination de la meilleure voie scolaire ou professionnelle, en particulier à la fin d'un cycle, après inventaire de ses aptitudes et de ses motivations (LAZIB Ep., 2020, p.16). L'objectif principal de l'orientation scolaire est d'accompagner les élèves dans le choix cohérent avec leurs centres d'intérêt, leurs compétences et leur personnalité. L'orientation est à la fois la détermination des points cardinaux, capacité à se situer dans le temps et dans l'espace (WEIXLER, F., 2021, p.4). L'orientation désigne aussi bien les modalités de production et de reproduction de la division sociale et technique du travail et l'action de donner une direction déterminée à sa vie. « Scolairement, c'est conseiller un enfant sur le métier qu'il peut choisir ». L'orientation scolaire vise à améliorer la qualité des systèmes éducatifs et celle des rapports entre éducation et vie professionnelle (LAZIB Ep., 2020, p.4).

2.1.5 TENASOSP

Le TENASOSP (Test National de Sélection et d'Orientation Scolaire et Professionnelle) est organisé au passage entre (6 années du cycle primaire et deux années du cycle secondaire général) aux quatre années des humanités. Ce test a été organisé au cours de l'année scolaire 2021-2022, huit ans après que celui-ci ait été institué comme le prescrit l'article 193, alinéa 2 de la loi-cadre de l'enseignement national de 2014 (Législateur Congolais, 2014). Les objectifs du TENASOSP sont multiples : orienter les élèves vers les humanités en tenant compte de leurs compétences et aptitudes, leur proposer des choix conformes à leurs profils cognitifs (mentaux) et psychologiques (comportementaux), etc. Ce test national est créé par l'arrêté Ministériel N°MINEPST/CABMIN/005/2022 du 13/01/2022 (Arrêté ministériel N°MINEPST/CABMIN/005/2022 du 13/01/2022), il totalisé quatre éditions et mérite une évaluation pour des éventuelles améliorations et outils d'accompagnement.

2.1.6 Changement d'orientation

Le changement d'orientation scolaire est une tentative de réajustement du parcours scolaire et professionnel de l'élève face aux écarts entre ses capacités, ses intérêts et les exigences de la filière engagée (BOUSSOUKAYA, F., 2010).

Cette tentative peut se faire juste après l'orientation, motivée par plusieurs aspects. Il peut s'agir des parents d'élèves comme ils ne sont pas associées au processus, soit, ce sont les paires de l'élève, soit, c'est l'école d'accueil lors de l'inscription. Le changement d'orientation est une étape où l'élève procède aux tris à partir de ses propres motivations intrinsèques (subjectives) et extrinsèques (objectives). Il les confronte à la réalité de son milieu socio-économique et dépendamment de la superstructure composant le système éducatif adjacent. A partir des différents obstacles connus ou points forts classifiés, l'élève qui est décideur principal de son orientation négocie son choix qui peut se solder soit un réajustement (changement de la filière), soit par une protection de sa décision finale s'il le peut (NAOUIFI S., 2023, p.5).

2.1.7 Education

Le terme éducation prête à plusieurs significations, malgré sa proximité évidente de la vie contemporaine, en fonction des contextes et des centres d'intérêt. Ce terme désigne, de manière générale, une action que l'adulte exerce sur l'enfant, en vue d'orienter ou d'infléchir son comportement. Il peut s'appliquer à une action formelle et académique, ou à une action informelle individuelle ou sociale. Très souvent, nous parlons d'éducation familiale, sociale, académique, etc., pour signifier que l'éducation intègre différents cadres, dimensions et acteurs (LAZIB Ep., 2020, p.51). L'éducation apparaît nécessairement comme une action exercée par différents acteurs, visant la formation totale de l'homme par le développement de ses potentialités et leurs valorisations par le groupe social auquel il participe (LAZIB Ep., 2020, p.52). Le système éducatif Congolais est subdivisé en deux structures, à savoir l'enseignement formel et l'éducation non formelle. L'enseignement formel est dispensé sous forme d'enseignement classique et d'enseignement spécial. L'éducation non formelle est donnée sous forme d'activités assurées dans des établissements spéciaux et dans des centres de formation. L'enseignement national de type classique est organisé en quatre niveaux. A savoir l'enseignement maternel, l'enseignement primaire, l'enseignement secondaire et l'enseignement supérieur et universitaire (Législateur Congolais, LOI-CADRE 2014, pp.23-24). L'actuel test est organisé entre l'enseignement primaire et l'enseignement secondaire.

2.2. Revue de littérature empirique

Ce point va évoquer les enquêtes, les expériences, les études de cas, les observations, les statistiques des auteurs qui se sont focalisés sur des recherches ayant un trait commun avec notre étude afin de comparer les constats empiriques pour voir s'il existe des tendances, des contradictions ou des lacunes conceptuelles, méthodologiques, techniques et de situer notre recherche dans une perspective nouvelle. Pour ce qui concerne notre recherche, cette revue empirique va compléter la revue théorique en analysant des travaux antérieurs sur l'utilisation des algorithmes d'apprentissage automatique en orientation scolaire et professionnelle, en formulant des hypothèses de recherche en fonction des études présentées. C'est ainsi que, la revue de littérature empirique a révélé de manière sélective ce qui suit : En France, les recherches de (DANVERS, F., 2022), qui sont revenues sur des données issues du système scolaire français, notamment sur les taux du respect des avis d'orientation scolaire à la fin de la 3^{ème} secondaire, ont montré que 75 à 85 % des décisions d'orientation sont conformes aux demandes des familles des apprenants, mais 15 à 25 % des décisions finales sont imposées par le conseillers d'orientation de classe ou contredisent les choix exprimés. Ces recherches ont révélé un taux non négligeable d'orientation subie, souvent plus élevé chez les élèves en difficulté d'apprentissage scolaire ou issus de milieux modestes. Pour le cas de la réorientation (lycéens), environ 1/5 élève change d'orientation scolaire au lycée, ce qui souligne une orientation anticipée ou prématuée. Danvers a relevé un nombre insuffisant de conseillers d'orientation dans les milieux scolaires en révélant qu'il y a 1 psychologue de l'Éducation nationale pour un total de 1 500 élèves en moyenne en milieux scolaires. Pour conclure, Danvers invite tous les acteurs de l'éducation nationale française à repenser l'orientation non pas comme un outil de sélection, mais comme l'art de vivre. En Afrique, par contre au Burkina Faso, (MOUMOULA I., et al., 2005), ont mobilisé des données qualitatives renforcées par des statistiques contextuelles pour éclairer l'orientation scolaire au Burkina. Après l'analyse de ces données collectées, les auteurs ont trouvé que, s'il y a changement de filière après orientation scolaire, cela est dû par plusieurs déterminants. Entre autre : la pression parentale qui prime sur autonomie de l'élève. Environ 12 % des parents camerounais déclarent ne pas savoir quoi conseiller, 48 % cherchent plutôt à rencontrer un conseiller. Ces deux auteurs ont souligné que, la pression parentale est d'autant plus forte que le milieu social.

Selon leurs recherches complétées par d'autres données, une forte autocensure, plus marquée chez les milieux modestes, influence le choix d'orientation. Ils ont montré que l'échec scolaire détermine fortement la trajectoire : les élèves mal notés voient leur accès aux voies générales ou professionnelles limité, amplifiant les inégalités. Les statistiques sur le système d'orientation Burkinabais de ces chercheurs montrent que, toutes les écoles de Burkina Faso comptent environ 100 conseillers d'orientation. Ce qui prouve qu'il y a un déficit de collaboration entre établissements et marché du travail. Leur recherche prouve en suite qu'il y a : la complexité du choix (l'élève est tiraillé entre aspiration personnelle), la pression familiale et contraintes institutionnelles, l'inégalité structurelles (pression sociale et capacité d'accès à l'orientation varient selon le milieu), le déficit institutionnel (très peu de conseillers pour accompagner efficacement tous les élèves). Pour conclure, les auteurs ont associé des données qualitatives (entretiens, analyse de dispositifs) et des statistiques simples (proportions, effectifs, ..) permettant d'objectiver la pression familiale, le déséquilibre dans l'accès aux services d'orientation et la rareté des professionnels formés. En revanche, (LAZIB Ep., 2020), a fait un constat amer selon lequel, le processus de l'orientation scolaire au Nigeria n'associe pas toujours tous les acteurs locaux de l'orientation scolaire qui sont les parents d'élèves, les enseignants et parfois les élèves eux-mêmes qui sont totalement oubliés dans ce processus. Pour remédier à ces difficultés, l'auteur a proposé un système d'E-orientation basé sur l'analyse d'opinion (OPinOR). Ce système a pour objectif d'améliorer l'orientation des élèves du secondaire vers les domaines d'études universitaires ou professionnelles. Il enrichit les critères d'orientation par l'intégration de l'opinion des intervenants dans ce processus d'une part et les aptitudes des élèves représentés par leurs intelligences multiples de Gardner d'autre part. Pour réaliser ce système, l'auteur a conçu une ressource lexicale de la langue française basée à la fois sur SentiWordNet et Wolf. Cette ressource est utilisée pour l'extraction de l'opinion à partir de textes. Le système OPinOR propose de recueillir, périodiquement, les opinions des intervenants à travers des enquêtes menées avec des questionnaires en ligne. Par contre, en RD Congo, le travail précédent de (KAVUGHO, f., 2024) a été réalisé sur le rendement scolaire des candidats ayant été orientés au TENASOSP et ayant accepté les avis de l'orienteur. Dans ce travail, le chercheur avait une préoccupation majeure, celle de comparer le rendement scolaire entre les candidats ayant respecté les avis d'orientation émis après TENASOSP et ceux ne les ayant pas respectés afin

de savoir si le respect des avis d'orientation émis après le TENASOSP peut influencer positivement le rendement scolaire. Après l'analyse de données de résultats scolaires basés sur un échantillon aléatoire de 254 enquêtés de différentes écoles de la sous division Butembo1, en R D Congo, l'auteur a montré que le respect total des avis d'orientations émis après TENASOSP influence positivement le rendement scolaire. Malgré que les objectifs fixés dans différentes recherches ci-haut citées soient atteints, les solutions proposées peuvent être améliorées, complétées et contextualisées. C'est ce qui ouvre pour nous une brèche aujourd'hui pour prévoir un modèle de collecte des données de l'orientation scolaire et professionnelle qui associe les avis de tous les acteurs locaux de l'orientation scolaire qui sont entre autre l'élève, les parents, les conseillers d'orientation scolaire de l'école de provenance de l'élève et autres facteurs socioéconomiques et infrastructurels pouvant influencer positivement les compétences et l'orientation scolaire chez l'apprenant.

3. Méthodes de recherche et matériels de traitement de données

3.1 Type d'étude : Notre recherche suit des approches : exploratoire mixte, expérimentale et orientée données.

3.2 Population cible : les élèves de 1^{ère} secondaire orientés au TENASOSP des éditions 2021-2024,

3.3 Echantillonnage

Les données de notre étude ont été collectées à partir des bases de données de la Division Provinciale de l'Education Kasai-Central 1 et cela selon un échantillonnage stratifié, en fonction de milieux scolaires (six petites sous-divisions rurales ou deux grandes sous-divisions urbaines). Cette méthode assure une représentation équilibrée des différents groupes (urbains et ruraux) pour permettre une comparaison rigoureuse à l'aide de l'ANOVA. Nos entretiens et interviews ont eu lieu entre 8 conseillers d'orientation de ces 8 sous-divisions.

Notre échantillon se présente de la manière suivante :

Strate 1 : Les élèves (375) des écoles rurales de trois éditions (2021 à 2024),

Strate 2 : Les élèves (375) des écoles urbaines de trois éditions (2021 à 2024).

Tableau N°1. Echantillon de la population

SOUS-DIVISION		SOUS-DIVISION		TOT
URBAINE		RURALE		
F	G	F	G	FG
454	1197	350	999	3000

Source : Auteur.

3.4 Méthodes de collecte de données :

Le choix de méthodes et matériels à utiliser dans notre étude a tenu compte de la nature de notre recherche. Nous avons retenu la méthode quantitative et qualitative des données du type secondaire qui n'ont pas été obtenues à partir d'un questionnaire. Au-delà de cette collecte, d'autres données ont été recueillies de manière raisonnée lors des entretiens directs avec les conseillers d'orientation, les chefs d'établissement et les parents d'élèves et les OPS de la division provinciale de l'éducation nationale. L'analyse documentaire des rapports annuels du ministère de l'éducation nationale, des palmarès scolaires et de publication des résultats du TENASOSP des éditions passées. Une étude de cas comparatif entre les milieux urbains et les milieux ruraux fait partie de nos techniques d'analyse de données. Au-delà des méthodes mixte et expérimentale, nous avons retenu l'approche orientée données sur une étude de cas afin vérifier la validité de la solution informatique proposée.

3.5 Outils d'analyse de données

Le choix des outils d'analyse dépend de la méthodologie utilisée car, les méthodes sont accompagnées par les techniques et aussi les outils. Pour ce qui concerne notre étude, nous avons utilisé les outils ci-après de la statistique descriptive :

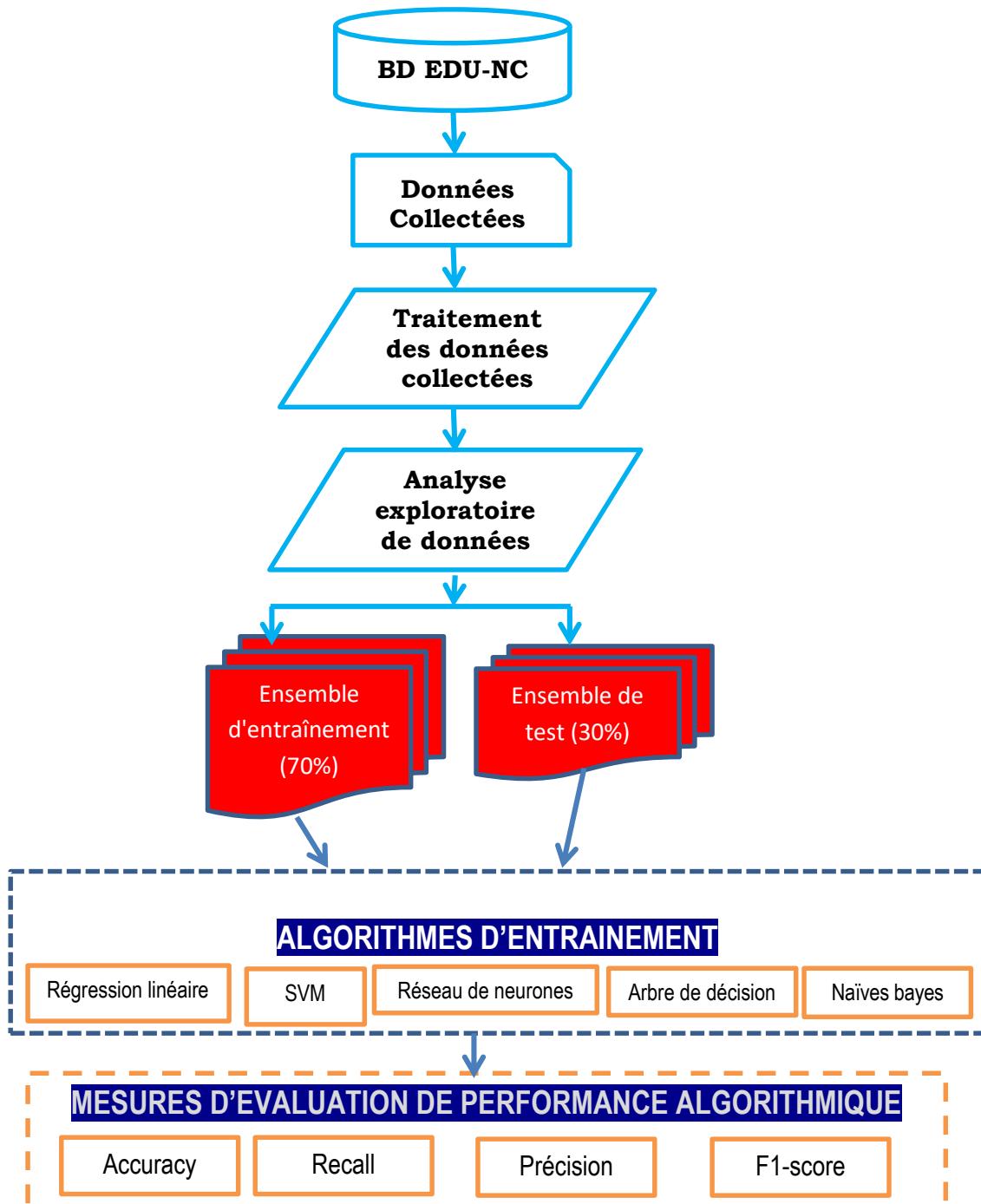
- Moyenne, médiane, mode des scores obtenus au TENASOSP,
- Fréquence des options choisies, par exemple, la pédagogie, la scientifique, la commerciale, etc,
- Taux de changement d'orientation par année ou par milieu (urbain et/ou rural),
- Comparaison des taux de respect/changement d'orientation entre les éditions,

- Test du Chi² d'indépendance, voir s'il existe une relation significative entre deux variables catégorielles. C'est-à-dire, voir si le milieu (urbain et/ou rural) peut-il influencer l'option choisie ou si le genre garçon/fille peut influencer le respect de l'orientation scolaire,
- Test t de Student ou Analyse de la Variance (ANOVA) pour comparer les moyennes entre les élèves qui respectent l'orientation s'ils ont un score moyen plus élevé que ceux qui la refusent. Ensuite, voire si le score moyen au TENASOSP peut varier selon l'année scolaire,
- Analyse de tendance pour observer l'évolution d'un indicateur sur plusieurs années, de 2021 à 2024. Comme par exemple, le taux de changement d'orientation par année, la Moyenne des scores au TENASOSP par année,
- Analyse Factorielle des Correspondances (AFC) pour explorer les relations entre plusieurs variables catégorielles. Par exemple, visualiser le lien entre options d'orientation, milieu, sexe et respect des avis,
- Excel et Orange Data Mining comme tableur et logiciel de traitement et de présentation des données sont intervenus pour l'analyse de nos données. Pour l'implémentation de notre solution et son expérimentation, nous avons retenu le langage de programmation Python, scikit-learn qui demeure le cœur du modèle prédictif (arbre de décision). La bibliothèque pandas et ses objets numpy, matplotlib, graphviz pour la gestion et la visualisation des données. Notre base de données est créée en MySQL, connectée au Flask/FastAPI pour le stockage des orientations scolaires. L'interface utilisateur utilisable par les élèves et enseignants est créée à partir de Flutter/Kivy.

3.6 Critères de sélection des individus

Les élèves ayant passé le TENASOSP au cours des années scolaires 2022 à 2024, dont les données sont disponibles dans des bases de données de l'EDUC-NC Kasaï-central 1, c'est-à-dire, les données démographiques des candidats, les scores aux différentes épreuves du TENASOSP, orientations attribuées, informations personnelles de base (âge, sexe, école, sous-divisions, etc.).

Figure N° 02. Le schéma suivant structure clairement les étapes de notre méthodologie suivie



Source : Auteur

La représentation ci-haut reprend la suite logique des étapes de notre méthodologie suivie ou étapes à franchir pour atteindre les objectifs de notre recherche. Nous sommes partis de la collecte des données historisées et secondaires présentes dans des bases de données de la Province Educationnelle Kasaï-Central 1. Après l'extraction des données des bases de données, c'est le prétraitement suivi du traitement des données qui conduit au nettoyage des données qui est le processus d'identification des parties incorrectes, incomplètes, inexactes, non pertinentes ou manquantes des données, puis leur modification, leur remplacement ou leur suppression en fonction de la nécessité, à l'harmonisation de leurs natures en vue d'une bonne analyse préalable. Dans ce contexte, l'analyse exploratoire des données (AED) nous a permis de répertorier toutes les valeurs manquantes qui existent dans notre base des données qui n'est rien d'autre que les datasets, constitués lors de l'extraction de données. Après avoir apprêté toutes les données, nous sommes passés à la phase d'entraînement des données qui est une étape de la modélisation proprement dite. Pour y arriver, nous avons procédé à une comparaison des algorithmes d'apprentissage supervisé afin d'en déterminer celui qui convient le mieux dans le contexte de l'orientation scolaire au TENASOSP, en RD Congo. Cette comparaison s'est basée sur cinq algorithmes qui ont été expérimentés pour prédire automatiquement l'orientation scolaire et professionnelle d'un candidat à la fin de son cycle de l'éducation de base (EB). Il s'agit notamment de Régression Linéaire (RL), Support Vector Machine (SVM), Réseau de Neurones (RN), Arbre de décision (AD) Naïves de bayes (NB). Comme mesure d'évaluation de performance algorithmique, nous avons utilisé comme mesures d'évaluation, Accuracy (qui est un % candidats correctement classés), Recall (qui est un % des candidats de type filière quelconque bien identifiés), Précision (qui est un % des candidats que le modèle dit type filière qui le sont réellement) et enfin F1-score (qui est un équilibrage entre Recall et F1-score).

3.7 Procédure de collecte des données

Dans le cadre d'une recherche scientifique, des données peuvent être classées en deux catégories notamment, les données primaires et les données secondaires. Les données primaires ou brutes sont celles qui sont issues des résultats d'une enquête données. Tandis que les données secondaires sont extraites à partir des systèmes d'information existants dans les entreprises.

Pour ce qui concerne notre recherche, nous avons utilisé les supports de masse comme flash disk, disque dur externe, et à partir de leur base de données et application de stockage dénommée Diplôme .cd.

Tableau N°2. Variables numériques, catégorielles et leurs codages

1. Variables personnelles

N°	Variable	Type	Degré d'influence de l'orientation scolaire	Entrée	Sortie
1	Nom candidat	Texte	Aucun	Oui	
2	Post nom	Texte	Aucun	Oui	
3	Prénom	Texte	Aucun	Oui	
4	Sexe	Catégorielle	Influence certains choix	Oui +	
5	Âge	Numérique	Peut indiquer (retard ou avance scolaire)	Oui +	
6	Année scolaire	Numérique	Peut indiquer (favorable ou défavorable)	Oui	
7	Classe	Numérique	Vérifier la cohérence d'âge avec la classe (redoublements ou avance)	Oui	
8	Langue parlée à la maison	Catégorielle	Peut influencer la compréhension en classe	Oui +	
9	Province d'origine	Catégorielle	Influence potentielle des opportunités professionnelles	Oui	

2. Variables scolaires

N°	Variable	Type	Utilité pour l'orientation		
1	Moyenne en	Numérique	Orientation vers sciences,	Oui +	

	Mathématiques		techniques		
2	Moyenne en Français	Numérique	Indice de compréhension et expression	Oui +	
3	Moyenne en Sciences	Numérique	Pertinent pour les profils scientifiques	Oui +	
4	Moyenne en Techniques manuelles	Numérique	Utile pour filières techniques/professionnelles		
5	Moyenne en Éducation artistique	Numérique	Pertinent pour filières artistiques	Oui +	
6	Présence aux cours (%)	Numérique	Mesure de l'assiduité	Oui +	
7	Redoublements antérieurs	Numérique	Peut indiquer des difficultés persistantes	Oui +	
8	Pré-test et avis d'orientation des enseignants	Numérique	Score global, orientation initiale proposée	Oui +	

3. Variables liées au contexte scolaire et familial

N°	Variable	Type	Description		
1	École de provenance	Catégorielle	Qualité pédagogique, localisation	Oui	
2	Gestion de l'école	Catégorielle	Officielle ou conventionnée	Oui +	
3	Code école	Numérique	Identifie de manière unique une école	Oui	
4	Situation géographie école	Catégorielle	Situe le milieu d'implantation de l'école	Oui +	
5	Pool d'inspection	Catégorielle	Zone géographique de l'école	Oui	
6	Code pool	Numérique	Identifie de manière	Oui	

			unique un pool		
7	Situation géographie pool	Catégorielle	Situe le milieu d'implantation du pool	Oui	
8	Province éducationnelle	Catégorielle	Qualité pédagogique, localisation	Oui	
9	Code province éducationnelle	Numérique	Identifie de manière unique la province éducationnelle	Oui	
10	Niveau d'études du père	Catégorielle	Niveau scolaire le plus élevé atteint par le père	Oui +	
11	Niveau d'études de la mère	Catégorielle	Idem pour la mère	Oui +	
12	Profession du père	Catégorielle	Peut orienter ou influencer la perception des métiers	Oui +	
13	Profession de la mère	Catégorielle	Idem	Oui +	
14	Revenu mensuel du père	Numérique	Souvent corrélé avec les opportunités d'études	Oui +	
15	Revenu mensuel de la mère	Numérique	Souvent corrélé avec les opportunités d'études	Oui +	
16	Nombre de frères/sœurs scolarisés	Numérique	Indice de ressources familiales	Oui +	
17	Avis d'orientation des parents	Texte	Appréciations utiles (si disponibles de façon standardisée)	Oui +	

4. Variables psychosociales et comportementales

N°	Variable	Type	Utilité pour l'orientation	
1	Motivation scolaire (autoévaluée)	Échelle (1-5)	Élève motivé peut réussir dans différentes filières	Oui +
2	Activités extrascolaires	Catégorielle	Sport, musique, artisanat... révèle des centres d'intérêt	Oui +

3	Projet de carrière déclaré	Texte	Peut orienter les propositions d'orientation (devenir qui)	Oui +	
4	Avis d'orientation des enseignants	Texte	Appréciations utiles (si disponibles de façon standardisée)	Oui +	
5	Résultat au test psychotechnique	Numérique	Mesure d'aptitudes cognitives ou pratiques	Oui +	

5. Variables cibles (à prédire)

N°	Variable	Type	Description		
1	Cycle d'orientation	Catégorielle	Indique si c'est le Général, Professionnel ou Spécial	Non	Oui
2	Option d'études	Catégorielle	Indique la filière selon le cycle trouvé	Non	Oui

Source : Auteur

Il est à indiquer que, dans le tableau 2 ci-haut, chaque variable de la colonne « Entrée » ayant le signe + (Oui +) participent dans l'entraînement du modèle d'apprentissage. Les autres (Oui) non. Ces dernières restent dans la base de données pour le référencement ;

Tableau N°03. Variables codées d'entrée (Features)

1. Variables personnelles		
N°	Variable d'entrée (Feature)	codage
1	Sexe	M=0, F=1
2	Âge	Inférieur à 15=0, supérieur à 15=1
3	Langue parlée à la maison	F=0, L=1, S=2, T=3
2. Variables scolaires		
1	Moyenne en Mathématiques	Faible=0, Moyen=1, Bon=2, Très bon=3.
2	Moyenne en Français	Faible=0, Moyen=1, Bon=2, Très bon=3.
3	Moyenne en Sciences	Faible=0, Moyen=1, Bon=2, Très bon=3.

4	Moyenne en Techniques manuelles	Faible=0, Moyen=1, Bon=2, Très bon=3.
5	Moyenne en Éducation artistique	Faible=0, Moyen=1, Bon=2, Très bon=3.
6	Présence aux cours (%)	Continue
7	Redoublements antérieurs	Continue
8	Pré-test et avis d'orientation des enseignants	Continue
3. Variables liées au contexte scolaire et familial		
1	Gestion de l'école	Publique =0, Privée=1,
2	Situation géographie école	Urbaine =0, Rurale=1
3	Niveau d'études du père	Aucun=0, Primaire=1, D6=2, G=3, L=4, D=5
4	Niveau d'études de la mère	Aucun=0, Primaire=1, D6=2, G=3, L=4, D=5
5	Profession du père	Sans emploi=0, Travaux champêtres=1, Tous travaux=2, Ouvrier=3, commerçant/indépendant=4, employé de bureau=5, fonctionnaire=6
6	Profession de la mère	Sans emploi=0, Travaux champêtres=1, Tous travaux=2, Ouvrier=3, commerçant/indépendant=4, employé de bureau=5, fonctionnaire=7
7	Revenu mensuel du père	Très faible revenu=0, Faible revenu=1, Moyen=2, Elevé=3, Très élevé=4
8	Revenu mensuel de la mère	Très faible revenu=0, Faible revenu=1, Moyen=2, Elevé=3, Très élevé=5
9	Nombre de frères/sœurs scolarisés	Contenue
10	Avis d'orientation des parents	Général=0, Professionnel=1, Spécial=2
4. Variables psychosociales et comportementales		
1	Motivation scolaire (autoévaluée)	Très faible=0, Faible=1, Moyenne=2, Bonne=3, Très bonne=4,
2	Activités extrascolaires	Aucune=0, sport=1, musique=2, Religieuses=3, Ménage=4, Business=5, agropastorales=6, informatique=7

3	Projet de carrière déclaré	Santé=0, enseignement=1, technique=2, Droit=3, Commerce=4, Art=5, Agriculture=6, informatique=7
4	Avis d'orientation des enseignants	Général=0, Professionnel=1, Spécial=2
5	Résultat au test psychotechnique	Faible=0, Moyen=1, Bon=2, Très bon=3.
5. Variables cibles (à prédire)		
1	Cycle d'orientation	Général=0, Professionnel=1, Spécial=2
2	Option d'études	Scientifique=0, Littéraire=1, Pédagogie=2, Technique commerciale=3, Technique coupe=4, Technique electroélectricité=5

Source : Auteur

4. Résultats et interprétations

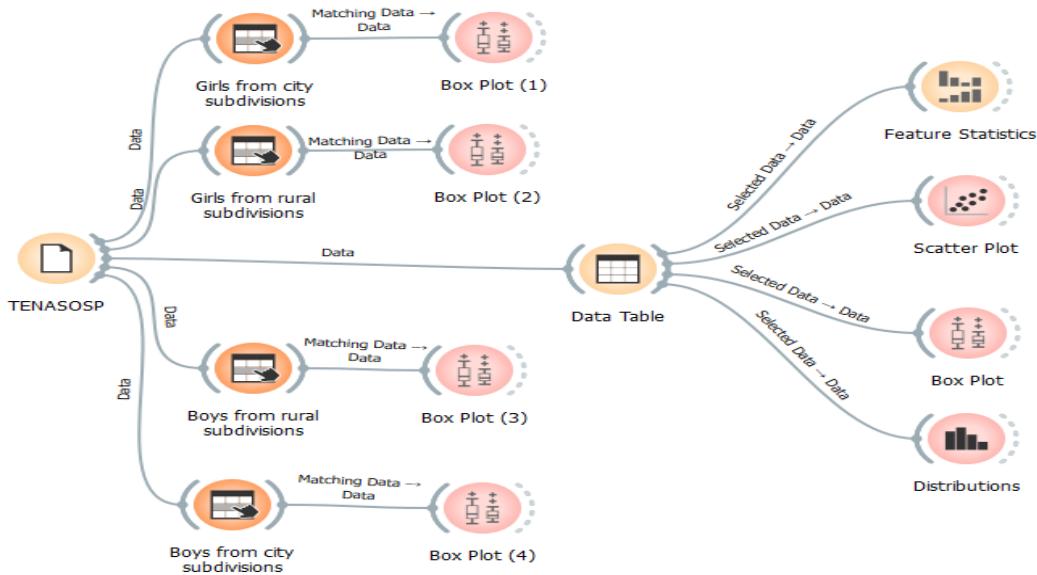
Notre premier objectif spécifique à atteindre consistait à collecter et analyser les données antérieures du TENASOSP afin d'en découvrir les points forts à consolider et les points faibles à remédier. Alors, après la collecte et l'analyse de données au moyen des techniques de Data Mining, les résultats et les interprétations suivants ont été dégagés.

4.1 Méthodes de collecte et correction du TENASOSP

Les données récoltées au près des conseillers d'orientation scolaire, Chefs d'établissement, parents et enseignants par interviews directs ont montré à 100% que, les intervenants dans le processus du TENASOSP recourent aux techniques et méthodes manuelles. Cela, de la collecte à la publication des résultats en passant par l'analyse et la recommandation de l'orientation scolaire et professionnelle. Ceci nous amène à affirmer notre première hypothèse selon laquelle, tout le processus d'orientation scolaire, du début à la fin serait manuel.

4.2 Schématisation de données

Figure N°3 : Schéma de la présentation des outils d'analyse de données

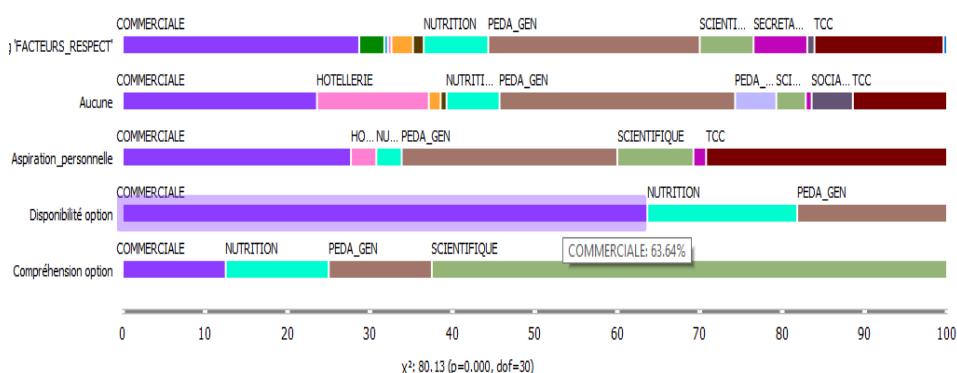


Source : Auteur.

4.3 Orientation la plus choisie selon le milieu et facteurs de respect

a. Par les filles en milieux urbains :

Figure N°4 : Orientation la plus voulue par les filles en ville

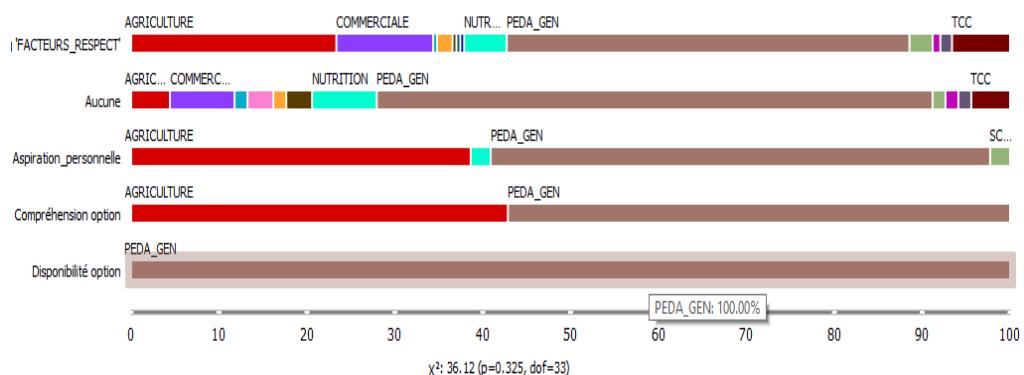


Source : Auteur.

De par ces résultats, l'option la plus désirée par les élèves filles évoluant en milieux urbains est la commerciale et gestion (63,64%) avec pour facteur de respect, disponibilité de l'option dans les milieux éducatifs urbains.

b. Par les filles en milieux ruraux :

Figure N°5 : Orientation la plus voulue par les filles au village

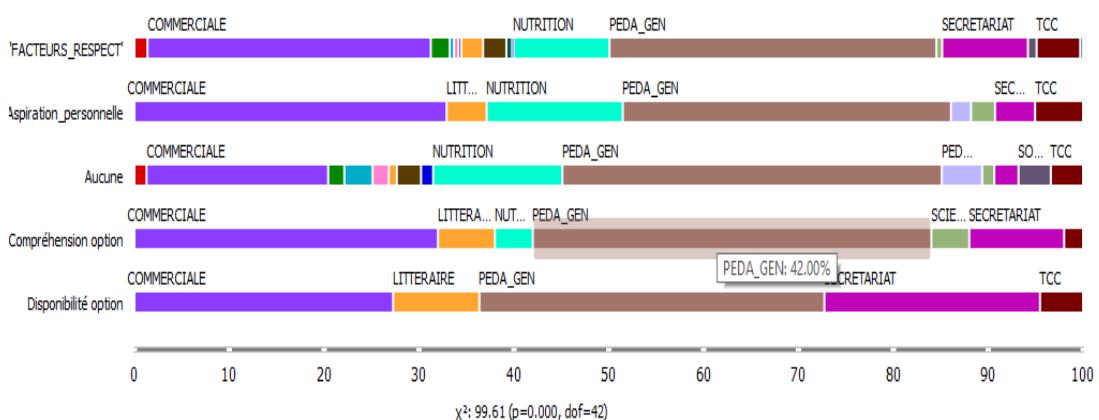


Source : Auteur.

Selon les résultats présentés ci-haut, l'option la plus suivie par les élèves filles évoluant en milieux ruraux est la commerciale et gestion (100%) avec pour facteur de respect, disponibilité de l'option dans les milieux éducatifs ruraux,

c. Par les garçons en milieux urbains :

Figure N°6 : Orientation la plus voulue par les garçons en ville

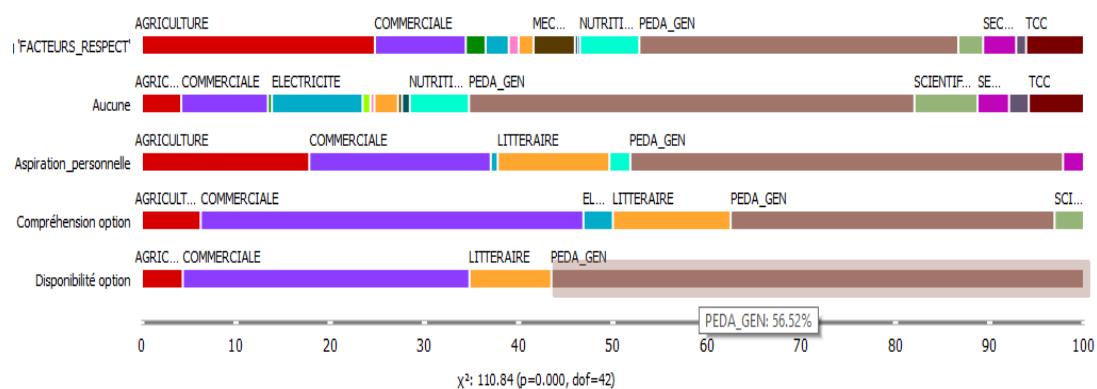


Source : Auteur.

Les résultats présentés ci-haut prouvent que, le choix d'option des élèves garçons évoluant en milieux urbains chevauche entre la commerciale et la pédagogique. Mais, c'est cette dernière qui prime (42%) avec pour facteur de respect « compréhension de l'option » contre la commerciale qui suit (32%) avec pour facteur de respect compréhension de l'option.

d. Par les garçons en milieux ruraux :

Figure N°7 : Orientation la plus voulue par les garçons au village

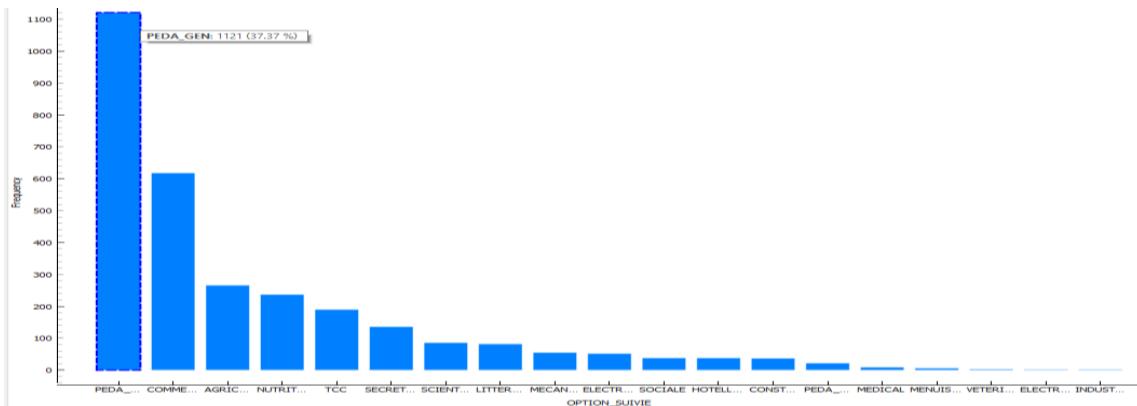


Source : Auteur.

Par contre, les résultats présentés ci-haut prouvent que, le choix d'option des élèves garçons évoluant en milieux ruraux prime sur la pédagogie (56,52%) avec pour facteur de respect disponibilité de l'option dans les milieux éducatifs ruraux.

De manière générale, l'option la plus voulue et suivie est présentée de la manière suivante :

Figure N°8 : Orientation la plus voulue par les élèves

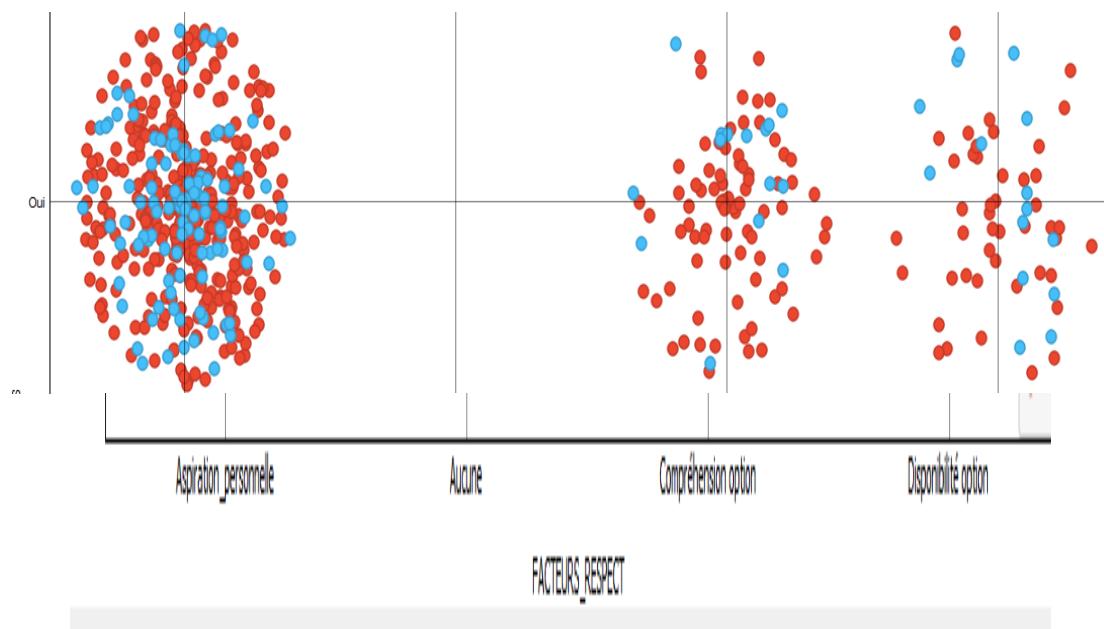


Source : Auteur.

Partant de notre deuxième hypothèse selon laquelle, c'est l'option pédagogie générale qui serait la plus suivie parmi tant d'autres, car, il y aurait plus de chance de trouver le travail dans l'enseignement national, nous affirmons à la lumière de résultats ci-haut que, pour trois éditions et pour les élèves filles et garçons et pour les milieux urbains et ruraux, l'option la plus suivie est pédagogie générale (37,37%) avec facteur de respect disponibilité de l'option dans les établissements scolaire. Cela prouve en suffisance que, dans tous les milieux éducatifs Congolais, plusieurs écoles organisent l'option pédagogique et il y a forte probabilité de trouver un travail d'enseignement.

4.4 Respect d'orientation recommandée par les conseillers d'orientation et facteurs de respect

Figure N°9 : Respect des orientations recommandées

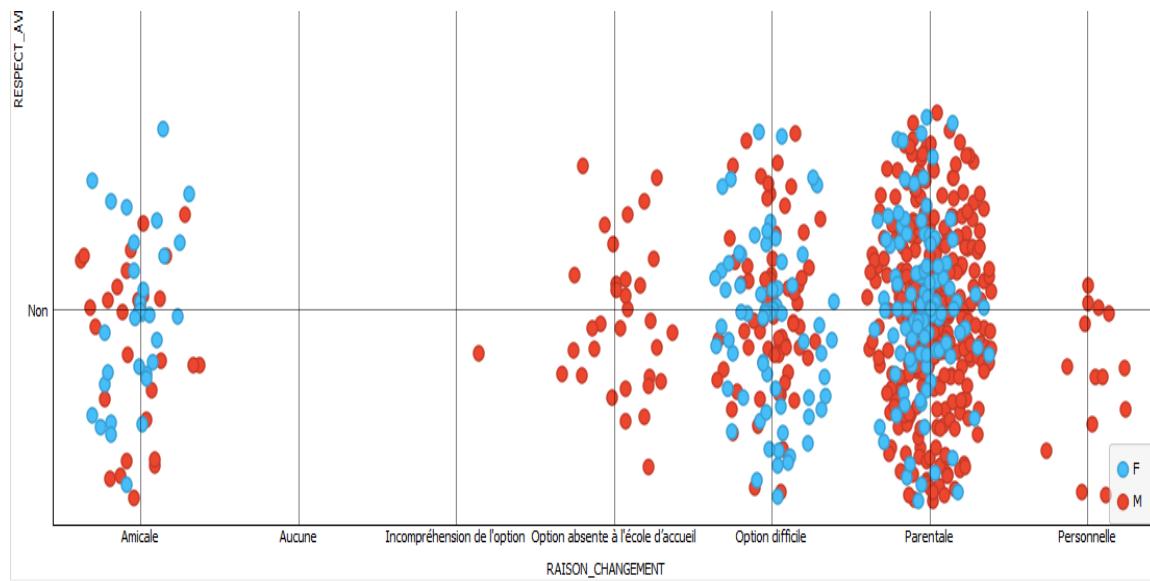


Source : Auteur.

En moyenne sur trois éditions du TENASOSP, 69,03 % des élèves acceptent l'orientation recommandée par les conseillers d'orientation, en majorité par aspiration personnelle.

4.4 Changement d'orientation recommandée et raison changement

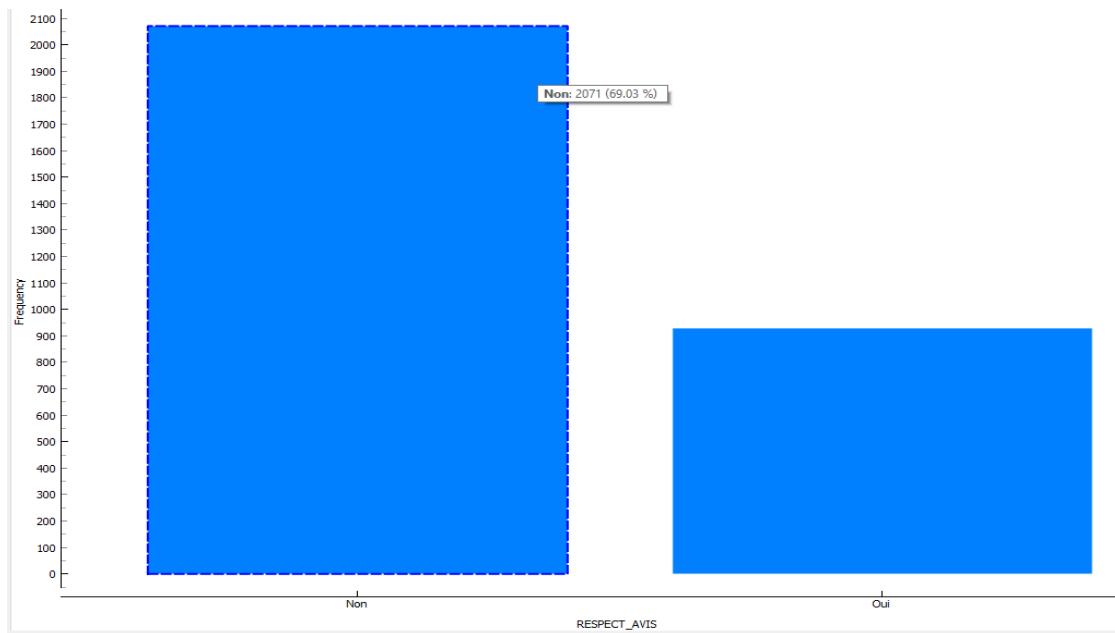
Figure N°10 : Changement des orientations recommandées



Source : Auteur.

En moyenne sur trois éditions du TENASOSP, 31,97 % d'élèves demandent un changement d'option ou carrément changent d'option motivés par leurs parents qui ne sont pas du tout associés à l'orientation de leurs enfants comme l'indique la constitution du pays que les parents ont le droit de choisir le mode d'éducation à donner à leurs enfants (JOURNAL OFFICIEL de la RDC, CONSTITUTION DE LA RDC, , Article 43, 2011). Ceci nous amène à, de un affirmer notre troisième hypothèse selon laquelle, certains élèves orientés au TENASOSOP auraient sollicité et obtenu le changement d'orientation et de deux infirmer notre quatrième hypothèse selon laquelle la majorité d'élèves ayant changé d'orientation l'aurait fait influencés par ses paires, car, les résultats prouvent le contraire, c'est la raison parentale qui prime sur les autres. Ces résultats de trois éditions du TENASOSP, respectivement pour le respect d'orientation et pour le non-respect d'orientation (69,03 % contre 71,97%) sont présentés dans l'histogramme suivant :

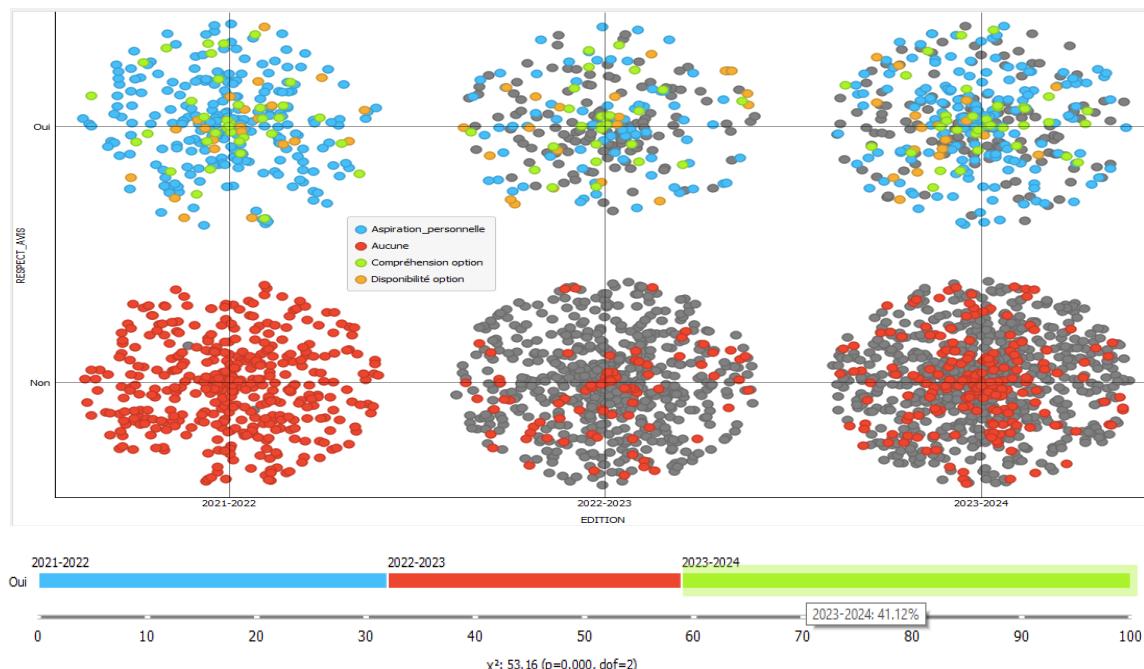
Figure N°11 : Respect et changement des orientations recommandées



Source : Auteur.

4.6 Évolution des cas de respect des avis d'orientation par édition et leurs facteurs

Figure N°12 : Niveau évolutif de respect des avis d'orientation par édition et leurs facteurs

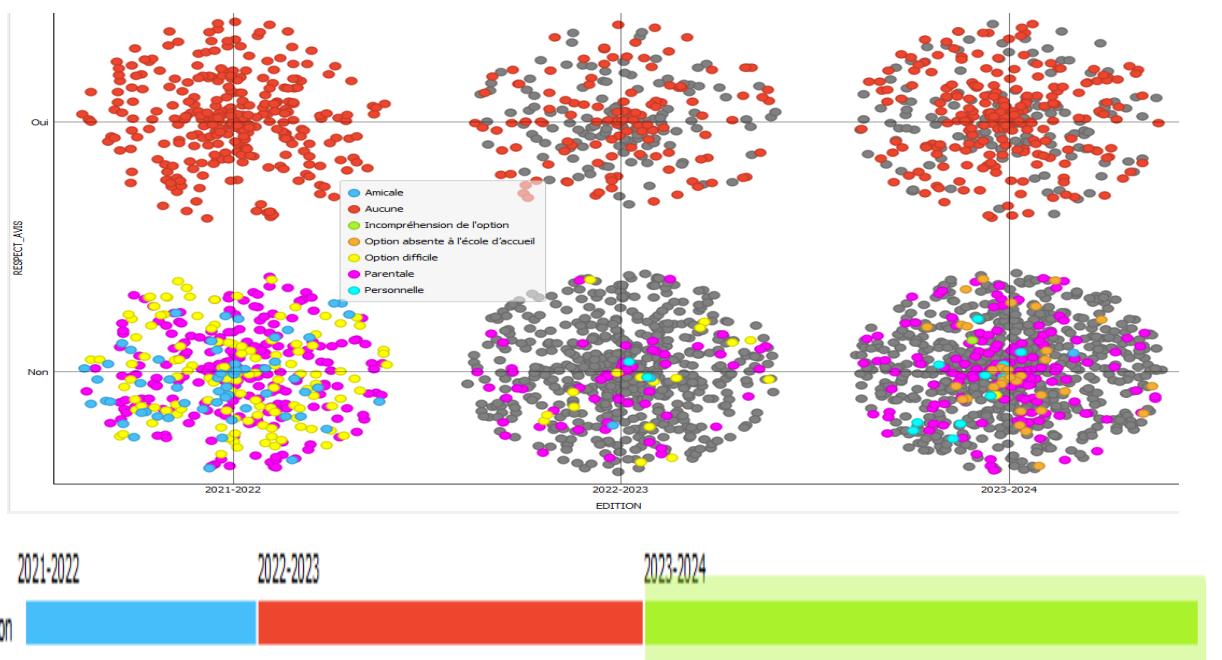


Source : Auteur.

Partant de ces résultats de trois éditions du TENASOSP, c'est l'édition 2023-2024 qui a connu plus des cas d'acceptation (41,12%) des avis d'orientations émis par les conseillers d'orientation pour tous les deux milieux éducatifs, suivie de la première (2021-2022). Les facteurs de respects des orientations scolaires recommandées sont repris dans les légendes de la figure 10.

4.7 Évolution des cas de changement des avis d'orientation par édition et leurs raisons

Figure N°13 : Niveau évolutif de respect des avis d'orientation par édition et leurs facteurs

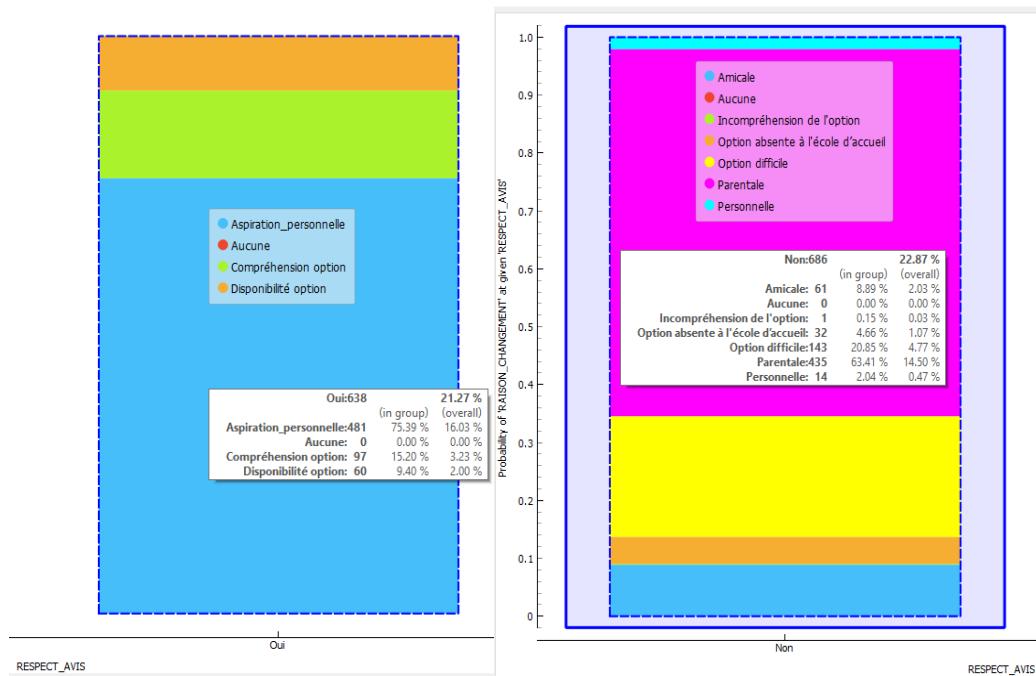


Source : Auteur.

Partant de résultats présentés ci-haut de trois éditions du TENASOSP, c'est l'édition 2023-2024 qui a connu plus des cas changement d'option (47,27%) des avis d'orientations émis par les conseillers d'orientation pour tous les deux milieux éducatifs, suivie de la deuxième (2022-2023). Les raisons de changement des orientations scolaires recommandées sont reprises dans les légendes de la figure 13. Cela nous conduit à infirmer notre cinquième

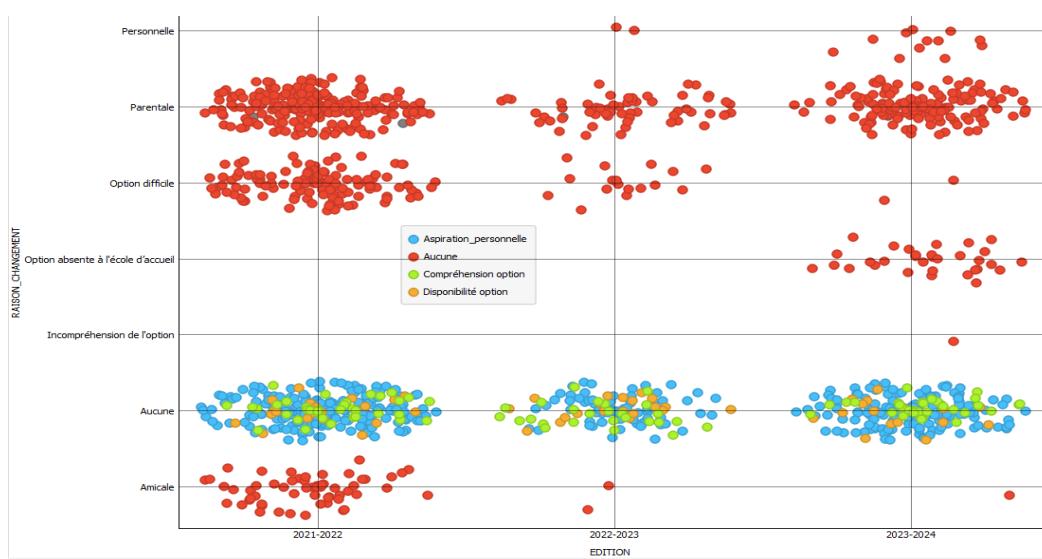
hypothèse, selon laquelle, c'est l'édition 2021-2022 qui aurait connu plus de cas de changement d'orientation scolaire par le fait que c'était une édition expérimentale.

Figure N°14 : Facteurs de respect et raisons de changement des avis d'orientation de 3 éditions



Source : Auteur.

Figure N°15 : Raisons de changement d'orientation par édition



Source : Auteur.

De par les résultats de la figure 15 ci-haut, la raison parentale occupe le top dans l'influence de changement des avis d'orientation scolaire et cela pour toutes les trois éditions passées. Cela est dû par le fait qu'ils n'ont jamais été associés par les conseillers d'orientation des sous-divisions lors du processus d'orientation scolaire. Dans ces résultats, figurent aussi, les autres élèves qui se sont inscrits sans passer par le test. Ceux-ci (64% en 2021-2022, 51% en 2022-2023 et 61%) n'ont aucune raison à avancer.

4.8 Application de modèles de Machine Learning

Notre deuxième objectif spécifique était de comparer les modèles de Machine Learning afin d'en découvrir celui qui prédit mieux l'orientation scolaire et professionnelle dans le contexte du TENASOSP. Pour atteindre cet objectif, nous avons procédé à une comparaison des algorithmes : Machine à Vecteurs de Support (SVM : Support Vector Machine), classificateur de Bayes naïf (Naïve bayes), Arbre de décision (Decision Tree), Forêt Aléatoire (Random Forest) et Réseau de neurone (Neural Network). Cette comparaison est faite sur base des différentes mesures de performance (métriques d'évaluation) évoquées par (BERTRAND, S., 2023, pp.20-24) :

- AUC (Area Under the Curve) : C'est une aire sous la courbe ROC qui mesure la capacité du modèle à distinguer les classes positives et négatives. Le meilleur modèle est celui qui donne la valeur plus proche de 1.
- CA (Classification Accuracy / Exactitude) : c'est la proportion de prédictions correctes sur l'ensemble des prédictions. Elle est donnée par la formule suivante :

$$CA = \frac{TP+TN}{TP+TN+FP+FN}$$
 . Elle est simple à comprendre, mais trompeuse si les classes à prédire sont déséquilibrées.

- MCC (Matthews Corrélation Coefficient) : elle mesure la corrélation entre les prédictions et les vraies classes. Sa formule est donnée par :
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$
. Sa valeur varie de -1 (prédiction totalement fausse) à +1 (parfaite). La valeur 0 est une prédiction aléatoire.
- F1-Score : elle est une moyenne harmonique entre la précision et le rappel. Elle est utile quand on veut un équilibre entre la précision et le rappel. Sa valeur varie entre 0 et 1 (plus proche de 1 = meilleur). Elle est donnée par la formule suivante :

$$F1 = 2 \times \text{Precision} \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

➤ Précision (PREC / Precision) : c'est la balance entre les instances prédictes pour trouver celles qui sont réellement positives. Elle est utile quand le coût d'un faux positif est élevé. Sa formule est la suivante : $PREC = \frac{TP}{TP+FP}$

➤ Rappel (RECALL / Sensitivity / Recall) : c'est la balance entre les instances prédictes pour trouver parmi les vrais positifs, celles qui ont été correctement détectés. Il est utile quand le coût d'un faux négatif est élevé. Sa formule est donnée par :

$RECALL = \frac{TP}{TP+FN}$. Enfin, le tableau ci-dessous présente en synthèse les résultats des mesures de performance pour ces algorithmes dans le tableau ci-dessous :

Tableau N°16 : Résumé des mesures de performance des modèles entraînés

Model	AUC	CA	F1	Prec	Recall	MCC
Naive Bayes	0.579	0.189	0.153	0.156	0.189	0.038
SVM	0.561	0.226	0.154	0.148	0.226	0.009
Tree	0.586	0.234	0.199	0.200	0.234	0.055
Random Forest	0.597	0.228	0.200	0.196	0.228	0.052
kNN	0.558	0.208	0.189	0.185	0.208	0.052

Source : Auteur.

D'après le résultat de la comparaison des modèles dans le tableau ci-haut, il a été évident que l'algorithme d'arbre de décision a été meilleur par rapport aux quatre autres classificateurs du point de vue toutes les mesures de performance sauf pour Random Forest dont l'aire sous la courbe ROC (AUC) est supérieure (0.59%) à celle de l'arbre de décision (0.58%). Ce qui amène à conclure que l'algorithme arbre de décision est celui qui a prédit mieux que les autres dans notre contexte de l'orientation scolaire et professionnelle au TENASOSP. Par conséquent, ce modèle sera déployé dans un environnement Web.

4.9 Implémentation du modèle valide de Machine Learning

Notre troisième objectif spécifique à atteindre consistait à implémenter l'algorithme de l'apprentissage automatique qui prédit mieux l'orientation scolaire et professionnelle des candidats au TENSAOSP. Alors, après avoir opté pour le modèle d'arbre de décision, nous

allons le déployer en utilisant la technologie web. Le but est de développer une interface graphique de visualisation des tests du modèle retenu pour permettre aux nouveaux élèves des années avenir, de soumettre leurs identités et leurs avis d'orientation en ligne, dans des établissements de l'Education Nationale et Nouvelle Citoyenneté de la RDC, de pouvoir prédire l'orientation scolaire adéquate d'un candidat. De ce qui précède, nous avons développé une interface Web en utilisant le HTML, CSS et le Serveur Flask de Python.

Figure N° 17 : Illustration du lancement du modèle d'entraînement

```
C:\tenasosp20250K>python train_model.py
Accuracy: 1.0
      precision    recall   f1-score   support
Commercial 1.00 1.00 1.00 1
Coupe 1.00 1.00 1.00 1
Litteraire 1.00 1.00 1.00 1
Scientifique 1.00 1.00 1.00 2
Social 1.00 1.00 1.00 1
Technique 1.00 1.00 1.00 1
accuracy 1.00 1.00 1.00 7
macro avg 1.00 1.00 1.00 7
weighted avg 1.00 1.00 1.00 7
Modèle enregistre dans decision_tree_pipeline.joblib
```

Source : Auteur.

Nous avons considéré un échantillon par défaut d'options organisées dans le système éducatif Congolais, le plus choisie et le plus subie.

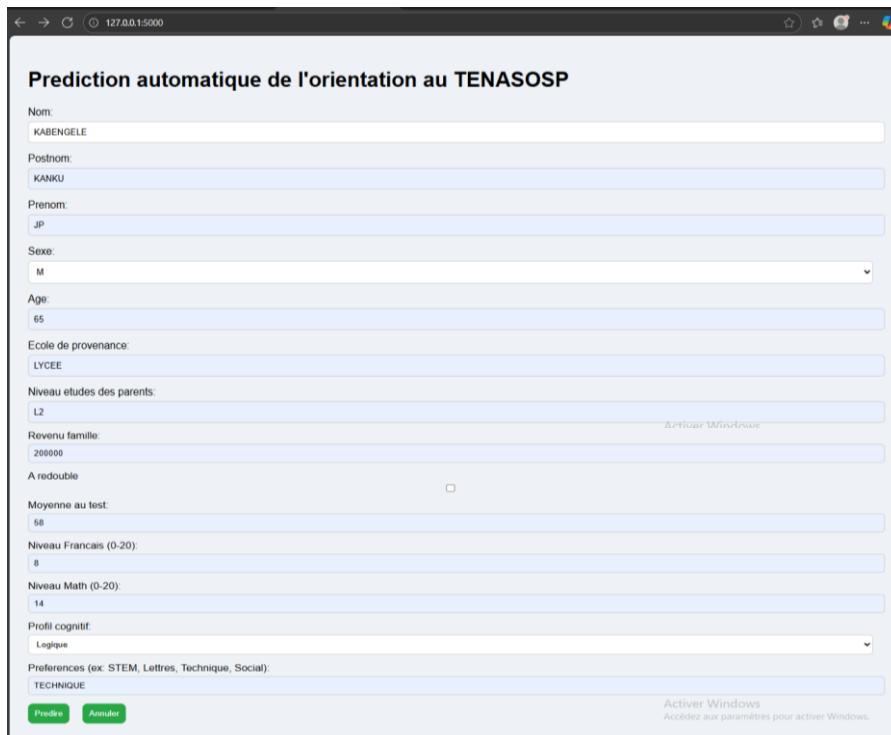
Figure N° 18 : Illustration du lancement du serveur Flash et ses autres services

```
C:\tenasosp20250K>python app.py
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on all addresses (0.0.0.0)
* Running on http://127.0.0.1:5000
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 253-585-497
```

Source : Auteur.

Dans cette interface, tous les services du serveur Flash de Python sont démarrés, activés et prêts à être utilisés à partir du navigateur web.

Figure N° 19 : Illustration de la page d'accueil de l'application Web



Prediction automatique de l'orientation au TENASOSP

Nom: KABENGELE

Postnom: KANKU

Prenom: JP

Sexe: M

Age: 65

Ecole de provenance: LYCEE

Niveau etudes des parents: L2

Revenu famille: 200000

A redouble:

Moyenne au test: 58

Niveau Francais (0-20): 8

Niveau Math (0-20): 14

Profil cognitif: Logique

Preferences (ex: STEM, Lettres, Technique, Social): TECHNIQUE

Prédire **Annuler**

Activer Windows
Accédez aux paramètres pour activer Windows.

Source : Auteur.

4.9.1 Scénario de cas d'utilisation

Ce formulaire s'affiche lorsque nous saisissons et validons l'adresse <http://127.0.0.1> dans la barre d'adresse du navigateur. Il permet à l'élève de fournir ses identités, ses préférences et ses réalisations scolaires (scores). Une fois validé, l'orientation s'affiche dans le navigateur web et directement, elle est aussi enregistrée dans la base de données pour un usage ultérieur. Nous soulignons que, ce formulaire est composé des champs de saisie simple, des ComboBox ou de type choix multiple (pour dire qu'il suffit de cliquer dans le champ et choisir une des réponses), des check box et des Command Buttons. Pour évaluer notre étude ou prototypage, les figures N°19 ci-haut et N°20 ci-dessous présentent le scénario d'utilisation du prototype que nous avons développé.

Figure N° 20 : Illustration de la base de données stockant les prédictions au TENASOSP

nom	postnom	prenom	sexe	age	ecole_provenance	niveau_etudes_parents	revenu_famille	a_redouble	score	niveau_math	profil_cognitif	preferences	prediction_option
1 KABONGO	BANDOWE	PIERRE	M	45	INSTITUT KANANGA ESTE	D6	4588.00	1	5	8	Logique	lettres	Scientifique
2 KABONGO	BANDOWE	PIERRE	M	45	LYCEE DISUMINYINA	MASTER	48500.00	1	5	8	Logique	lettres	Scientifique
3 KANKU	KALONJI	PIERRE	M	45	INSTITUT MPASHI	MASTER	45.00	1	5	8	Logique	lettres	Scientifique
4 MUKUNA	MUKENDI	LOLIS	F	50	INST KATUBUA	G3	5000.00	0	4	4	Pratique	TECHNIQUE	Technique
5 MUKUNA	MUKENDI	LOLIS	F	50	INST KATUBUA	G3	5000.00	0	4	4	Pratique	TECHNIQUE	Technique
6 MUKUNA	MUKENDI	LOLIS	F	50	INST KATUBUA	G3	5000.00	0	4	4	Pratique	TECHNIQUE	Technique
7 MUKUNA	MUKENDI	LOLIS	F	50	INST KATUBUA	G3	5000.00	0	4	4	Pratique	TECHNIQUE	Technique
8 MUKUNA	MUKENDI	LOLIS	F	50	INST KATUBUA	G3	5000.00	0	4	4	Pratique	TECHNIQUE	Technique
9 MUKUNA	KALALA	LOLIS	M	50	INST KATUBUA	G3	Activer Windows Accédez aux paramètres pour activer Windows	5000.00	0	4	4	Social	Activer Windows TECHNIQUE Social Accédez aux paramètres pour activer
Console de requêtes SQL													

Source : Auteur.

5. Discussion des résultats

Notre premier objectif spécifique était de collecter, analyser les données antérieures du TENASOSP et interpréter les résultats afin d'en découvrir les points forts à consolider et les points faibles à remédier. De par les résultats obtenus après la collecte et l'analyse des données, nous avons observé que, l'option la plus désirée par les élèves filles évoluant en milieux urbains est la commerciale et gestion (63,64%) avec pour facteur de respect, la disponibilité de l'option dans les milieux éducatifs urbains. Par contre, selon les résultats obtenus, l'option la plus suivie par les élèves filles évoluant en milieux ruraux est la commerciale et gestion (100%) avec pour facteur de respect, la disponibilité de l'option dans les milieux éducatifs ruraux. Les résultats ont aussi montré que, le choix d'option des élèves garçons évoluant en milieux urbains chevauche entre la commerciale et la pédagogique. Mais, c'est cette dernière qui prime (42%) avec pour facteur de respect compréhension de l'option

contre la commerciale qui suit (32%) avec pour facteur de respect compréhension de l'option. Par contre, les résultats obtenus ont prouvé que, le choix d'option des élèves garçons évoluant en milieux ruraux prime sur la pédagogie (56,52%) avec pour facteur de respect disponibilité de l'option dans les milieux éducatifs ruraux. Nous avons, à la lumière des résultats obtenus trouvé que, c'est l'option pédagogie générale qui est la plus suivie parmi tant d'autres, car, il y a plus de chance de trouver le travail dans l'enseignement national. Pour trois éditions et pour les élèves filles et garçons et pour les milieux urbains et ruraux, l'option la plus suivie est la pédagogie générale (37,37%) avec facteur de respect disponibilité de l'option dans les établissements scolaires. Cela prouve en suffisance que, dans tous les milieux éducatifs Congolais, plusieurs écoles organisent l'option pédagogique et il y a forte probabilité de trouver un travail d'enseignement.

En moyenne sur trois éditions du TENASOSP, 69,03 % des élèves acceptent l'orientation recommandée par les conseillers d'orientation, en majorité par aspiration personnelle. Et sur trois éditions du TENASOSP, 31,97 % d'élèves demandent un changement d'option ou carrément changent d'option motivés par leurs parents qui ne sont pas du tout associés à l'orientation de leurs enfants comme l'indique la constitution du pays que les parents ont le droit de choisir le mode d'éducation à donner à leurs enfants (CONSTITUTION DE LA RDC, , Article 43, 2011). Ces résultats nous amènent à, de un affirmer que certains élèves orientés au TENASOSOP ont sollicité et obtenu le changement d'orientation et de deux, la majorité d'élèves ayant changé d'orientation l'ont fait influencés par la raison parentale qui prime sur les autres. Partant des résultats de trois éditions du TENASOSP, c'est l'édition 2023-2024 qui a connu plus des cas d'acceptation (41,12%) des avis d'orientations émis par les conseillers d'orientation pour tous les deux milieux éducatifs, suivie de la première (2021-2022). Les résultats de nos recherches de trois éditions du TENASOSP ont montré que c'est l'édition 2023-2024 qui a connu plus des cas changement d'option (47,27%) des avis d'orientations émis par les conseillers d'orientation pour tous les deux milieux éducatifs, suivie de la deuxième (2022-2023). Alors, c'est l'édition 2021-2022 qui a connu plus de cas de changement d'orientation scolaire par le fait que c'était une édition expérimentale. Nos recherches ont prouvé que, (64% en 2021-2022, 51% en 2022-2023 et 61%) d'élèves ont pris l'inscription sans passer par le test et sans aucune raison avancée.

Notre deuxième objectif spécifique était de comparer les modèles de Machine Learning afin d'en découvrir celui qui prédit mieux l'orientation scolaire et professionnelle dans le contexte du TENASOSP. Pour atteindre cet objectif, nous avons procédé à une comparaison des algorithmes : Machine à Vecteurs de Support (SVM : Support Vector Machine), classificateur de Bayes naïf (Naïve bayes), Arbre de décision (Decision Tree), Forêt Aléatoire (Random Forest) et Réseau de neurone (Neural Network). D'après le résultat de la comparaison des modèles, il a été évident que c'est l'algorithme d'arbre de décision qui est meilleur par rapport aux quatre autres classificateurs du point de vue toutes les mesures de performance sauf pour Random Forest dont l'aire sous la courbe ROC (AUC) est supérieure (0.59%) à celle de l'arbre de décision (0.58%). Par conséquent, ce modèle est déployé dans un environnement Web. De ce qui précède, nous avons développé une interface Web en utilisant le HTML, le JavaScript et le Serveur Flask de Python. Ci-dessous nous voyons la page d'accueil permettant aux utilisateurs d'exploiter le modèle en vue de prédire leurs orientations scolaires.

6. Limites de notre solution proposée et perspectives d'avenir

Le système que nous avons proposé au public et général et au Ministère de l'Education Nationale et Nouvelle Citoyenneté présente certaines limites. Entre autre, un élève peut entraîner ses données plus d'une fois. Ce qui faucherait la prédiction. Dans nos perspectives d'avenir, la collecte des données (identités, préférences (de l'élèves, parents et conseillers d'orientation de l'école de provenance), scores, etc.) sera digitalisée et en ligne, à l'issue de laquelle, un numéro unique sera généré pour chaque élève qui lui permettra le jour de la passation du test, de l'utiliser comme mot de passe. Une fois accédé, il peut faire son test en ligne une et une seule fois et la correction se fait automatiquement après avoir cliqué sur un bouton « Corriger ». A la même occasion, le candidat au test clique sur un autre bouton Prédire l'orientation, immédiatement l'orientation est prédite.

Conclusion

La présente étude a mis en lumière après l'analyse exploratoire les points forts que nous avons consolidé et les points faibles du TENASOSP que nous avons corrigé en implémentant un modèle de prédiction automatique à base de l'algorithme d'arbre de décision. Lequel modèle s'est montré excellant dans la prédiction automatique de l'orientation scolaire et professionnelle. De par les résultats, le test d'orientation scolaire et professionnelle n'a pas résolu le problème d'orientation scolaire car, l'élève orienté parvient à changer son orientation scolaire motivé par ses parents, ses paires ou l'école d'accueil. En outre, les acteurs locaux de l'orientations scolaire (parents, enseignants orienteurs des écoles) ne sont pas du tout associées. Donc, le système actuel du TENASOSP n'a pas atteint ses objectifs terminaux et il a mérité ce que nous venons de réaliser comme travail et ce que nous aurons à réaliser dans un future proche.

Bibliographie

1. ARRETE MINISTERIEL N°MINEPST/CABMIN/005/2022 du 13/01/2022,
2. AZENCOTT, C-A., (2017). Introduction au Machine Learning 1. Livre 2. <http://tinyurl.com/ma2823-20173>,
3. BADACHE, I., et al., (2024) Intelligence artificielle: usage pédagogique et esprit critique, HAL Id: hal-04659335 <https://hal.science/hal-04659335v1>,
4. BARBARO, F., (2022). Analyse exploratoire et classification de textes, Thèse de doctorat de l'Université Paris 1 Panthéon Sorbonne, HAL Id: tel-03708173, <https://hal.science/tel-03708173v1>,
5. BERTRAND, S., (2023). « Détection de menaces internes par apprentissage automatique non supervisé », Mémoire de Master, Université de NAVAL,
6. CHARDIN, D., (2023). Étude de différentes méthodes d'apprentissage supervisé pour le développement de tests diagnostiques basés sur des données métabolomiques, Thèse, HAL Id: tel-04143914 <https://theses.hal.science/tel-04143914v1> Submitted on 28 Jun 2023
7. COUSINEAU, D., et al., (2010). « Analyse exploratoire des données : Concepts et outils ». Article, Revue des sciences de l'éducation,
8. DANVERS, F., (2022). S'orienter dans la vie (5 tomes), Livre, PUPS/Harmattan,
9. DORAY, P., (2020). « L'orientation scolaire sous l'emprise des rapports sociaux de sexe. La situation dans l'enseignement postsecondaire au Québec, Article, URL : <https://journals.openedition.org/osp/11962> DOI : 10.4000/osp.11962 ISSN : 2104-3795
10. GEORGES, B., (2019). « Intelligence artificielle : de quoi parle-t-on ? Article, dans Constructif 2019/3 N° 54, Éditions Fédération Française du Bâtiment ISSN 1950-5051 DOI 10.3917/const.054.000, Article disponible en ligne à l'adresse <https://shs.cairn.info/revue-constructif-2019-3-page-5?lang=fr>,
11. GUICHARD, D., (2005). « Choisir son avenir : guide pratique pour réussir son orientation ». Nathan,
12. JOURNAL OFFICIEL de la RDC, CONSTITUTION DE LA RDC, 2011, Article 43,
13. KANDUKI, M., et al., (2023). « Prédiction de l'orientation des étudiants dans des filières d'études appropriées en utilisant les techniques de Data Mining », Article, International Journal of Innovation and Applied Studies ISSN 2028-9324 Vol. 39 No. 1 Mar. 2023, pp. 193-208 © 2023 Innovative Space of Scientific Research Journals <http://www.ijias.issr-journals.org/>,
14. KAVUGHO, F., (2024). « Des avis d'orientation émis au TENASOSP en fin de la 8ième année d'éducation de Base et le rendement des élèves : observations faites dans les écoles de la Sous Division de Butembo 1 », Article, Université Officielle de Ruwenzori Butembo, (P.ISSN: 2992-247X), Volume 07 sur <https://ijojournals.com>,
15. LAHRACHE R. & BEKKAOUI A. (2024). « L'impact de l'intelligence artificielle sur la prise de décision», Article, Revue Internationale des Sciences de Gestion « Volume 7 : Numéro 3 », www.revue-isg.com
16. LAZIB Ep., (2020). Un système pour la E-Orientation Scolaire Intégrant Un l'Analyse d'Opinion, THESE DE DOCTORAT LMD, République Algérienne Démocratique et Populaire,

17. LOI-CADRE n° 14/004 du 11 Février 2014 DE L'ENSEIGNEMENT NATIONAL, p. http://www.leganet.cd/Legislation/Droit%20Public/enseignement/Loi14.004.11.02.20_04.htm
18. METRAL, J-F., (2018). « L'orientation des élèves au lycée : un travail collectif », Article, URL : <http://journals.openedition.org/osp/8896> DOI : 10.4000/osp.8896 ISSN : 2104-3795,
19. MOKDAR , S., (2024). « La mise en œuvre des algorithmes d'apprentissage automatique en assurance non-vie : une revue de littérature », Article, International Journal of Accounting, Finance, Auditing, Management and Economics - IJAFAFAME ISSN: 2658-8455 Volume 5,
20. MOUMOULA, I., et al., (2005). « L'orientation scolaire et professionnelle au Burkina Faso ». Article, Revue internationale de l'éducation de Sèvres, au Burkina Faso, consulté le 15/07/2025,
21. NAOUFI S., (2023). « La décision d'orientation : quels enjeux ? », Article, Revue Marocaine d'Evaluation et de la Recherche Educative. N°10 Revue scientifique internationale, ISSN Print: 2550-5688. E-ISSN : 2658-9079 avec DOI et Indexée par IMIST Rabat.
22. NSIALA KABEYA, D., (2020). « Système intelligent d'aide à l'orientation scolaire en RDC : cas des élèves de 3e année secondaire **Appart** : Travaux congolais sur la prédiction de l'orientation scolaire par IA, en utilisant des scores et profils scolaires ». **Institution** : Mémoire de Master, Université de Kinshasa.
23. WANIEZ, P., (2021). Analyse exploratoire des données, Collection Reclus Modes d'Emploi n°7, Livre, MAISON DE LA GÉOGRAPHIE, MONTPELLIER, horizon.documentation.ird.fr
24. WEIXLER, F., (2021). « L'ORIENTATION SCOLAIRE PARADOXES, MYTHES ET DÉFIS », Article, Grenoble, Inspection Générale de l'éducation du Sport, République Française,
25. ZOUHRI. A., (2019). « Le Potentiel De L'intelligence Artificielle Au Maroc, Une Question De Perception : Cas Des Entreprises Marocaines D'assurances», Article, Revue du contrôle, de la comptabilité et de l'audit, « Numéro 10 : Septembre 2019 / Volume 4 : numéro 2 », www.revuecca.com.