

Prédiction de l'orientation des étudiants dans des filières d'études appropriées en utilisant les techniques de Data Mining

[Predicting students' orientation in suitable fields of study using data mining techniques]

KANDUKI KIVUYIRWA MYSTERE, HERITIER NSENGE MPIA, and MUTEKGHEKI BARAKA VINGI

Département d'Informatique de Gestion, Université de l'Assomption au Congo, B.P. 104, Butembo, Nord-Kivu, RD Congo

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

ABSTRACT: Several data mining techniques are used to extract hidden knowledge in educational data to help students make a useful decision for their university orientation. Indeed, every year, students are enrolling in universities, the massive arrival of these candidates poses the thorny problem of orientation. The hidden problem behind this orientation is the lack of information concerning the possibilities of orientation; or the lack of support from the close entourage. Having developed the survey questionnaire, the authors collected 712 responses. After analyzing these data, they trained the models and measured their performance with four evaluation measures: accuracy, precision, recall and the F-score. The results of these models showed that the SVM algorithm gave 70% accuracy, the Naïve Bayes 65% Accuracy, the Neural Network 64% and the decision tree gave only 52%. This allowed SVM to be selected as the model that predicted better than the others. Finally, the authors deployed the validated model in web technology using Flask.

KEYWORDS: Academic Guidance, field of study, Machine Learning, Data Mining, SVM, Naïve Bayes, Neural network, Decision tree.

RESUME: De nombreuses techniques de Data Mining sont utilisées pour extraire les connaissances cachées dans des données éducatives en vue d'aider les étudiants à prendre une décision utile pour leur orientation à l'université. En effet, chaque année, les étudiants se présentent aux portes des universités, l'arrivée massive de ces candidats pose l'épineux problème de leur orientation adéquate dans des filières d'études. La problématique cachée derrière cette orientation est le manque d'information concernant les possibilités d'orientation; ou encore le manque de soutien de la part de l'entourage proche. Ayant mis au point le questionnaire d'enquête, les auteurs ont collecté 712 réponses auprès des répondants. Après analyse de ces données collectées, les auteurs ont réalisé l'entraînement des modèles, après, les auteurs ont mesuré les performances de ces modèles avec quatre mesures d'évaluation qui sont accuracy, precision, recall et le F-score. Les résultats de ces modèles ont montré que l'algorithme SVM a donné 70 % d'accuracy, le Naïve Bayes 65% d'Accuracy, le Réseau de neurone 64% et l'arbre de décision n'a donné que 52%. Cela a permis de sélectionner SVM comme le modèle qui a prédit le mieux que les autres. Enfin, les auteurs ont implémenté le modèle validé dans la technologie web en utilisant Flask.

MOTS-CLEFS: Orientation scolaire, filière d'études, Apprentissage automatique, Exploration de données, SVM, Naïve Bayes, Réseau de neurones, Arbre de décision.

1 INTRODUCTION

L'homme étant en quête d'une vie harmonieuse et heureuse, cherche à tout prix à l'atteindre de diverses manières. En ayant un bagage solide ou des compétences dans tels ou tels autres domaines, l'homme peut y parvenir. Parfois, l'échec tend à décourager l'homme qui entreprend une activité. Ce dernier a tendance à diminuer le sentiment d'efficacité personnelle mais c'est surtout le traitement cognitif que l'individu fera de ce succès ou de cet échec qui déterminera son impact [1]. Dans la vie académique, l'orientation dans des domaines étudiants joue un grand rôle et il est la base dans la préparation à la vie professionnelle [2]. Aujourd'hui, les études supérieures techniques et informatiques sont aux initiés les plus privilégiées dans la mesure où elles ont été conçues pour apprendre le métier adapté aux facteurs et à la situation du jour. Toutefois, la mauvaise adaptation de filières techniques et informatiques est actuellement considérée comme le plus grave des défauts de l'enseignement technique et informatique [2]. D'après les recherches faites par Christine Fontanini dans son article sur les facteurs déterminant les choix d'études supérieures, les motivations avancées par les élèves concernent leurs différents choix de s'inscrire en classe préparatoire différent selon la spécialité convoitée [3]. En République Démocratique du Congo (RDC), chaque année, les étudiants se présentent aux portes des universités et Instituts Supérieurs dans le but d'une orientation dans un domaine quelconque. Et, fort malheureusement beaucoup d'entre eux choisissent des facultés par plaisir, par l'influence des amis, voire à cause des outils utilisés dans ce domaine. D'une manière générale, certains sont orientés ou forcés par les membres de leurs familles. Etant donné que le pays est en voie de développement, il y a ceux-là qui se sentent obligés de choisir un quelconque domaine sans leur propre volonté. Le candidat pourra, par exemple, choisir un domaine par manque d'autres domaines dans sa région, etc. L'orientation scolaire reste un vrai problème pour les étudiants car les résultats obtenus à l'école secondaire ne leurs permettent pas de s'inscrire aux domaines d'études souhaités [1]. Par conséquent, certains étudiants ne s'adaptent pas dans la faculté choisie, parce que leurs capacités intellectuelles ne leurs permettent pas. Ainsi, ces difficultés peuvent engendrer des troubles psychiques. De notre temps, le mauvais choix des étudiants constitue la cause des échecs à la fin de l'année, de l'abandon de cours, de l'incapacité d'entreprendre après le parcours, de l'inadaptation dans les entreprises, etc. Ce qui conduit la jeunesse à la facilité, à la corruption, à la fraude, au vol, à la rébellion, au terrorisme, etc. Bref, le mauvais choix du domaine entraîne la non évolution de soi, du pays et du monde entier [4].

Par ailleurs, Troyes Dijon, insiste sur le fait que le danger aux études est de penser que les études faites déterminent la carrière à exercer et que ces études suffisent à elles-mêmes. Si l'étudiant, au cours de son cursus, ne commence pas à se constituer un début de réseau professionnel, le passage à *l'après école* peut être compliqué et décourageant [5]. De ce fait, les nouveaux adhérents doivent prendre conscience de cette difficulté et assouplir un peu leurs règles. Ils doivent être déterminés et courageux dans leurs choix car ils sont la fondation du monde de demain. Tous les étudiants sont compétents et performants dans le cas où ils ont fait un bon choix du domaine à l'université. Pour ce faire, le présent travail a pour objectif principal de déterminer les modèles Machine Learning (ML) qui prédisent mieux l'orientation des futurs étudiants dans des facultés ou options techniques et informatique de la RDC en comparant les algorithmes de Support Vector Machine (SVM), Réseau de neurones, Naïve Bayes et l'Arbre de décision. D'une façon spécifique, cette recherche poursuit trois objectifs spécifiques et pour y répondre, trois questions de recherche ont constitués l'effort cognitif de cette étude.

Tableau 1. Mapping des objectifs spécifiques avec les questions de recherche

Objectifs spécifiques	Questions de recherche
Construire un cadre conceptuel capturant les facteurs conditionnant un choix judicieux des filières techniques et informatiques	Quels sont les facteurs qui prédisent le choix d'une filière technique et/ou informatique ?
Déterminer l'algorithme ML qui prédit mieux possible l'orientation des futurs étudiants dans des facultés techniques et informatique	Existe-il un algorithme d'apprentissage automatique capable de prédire la filière à suivre par les nouveaux adhérents aux portes des universités ?
Implémenter et déployer le modèle ML qui prédit le mieux l'orientation des futurs étudiants dans des facultés techniques et informatique de la RDC en utilisant la technologie web	Quelle technologie pouvons-nous utiliser pour déployer notre modèle ML développé ?

2 REVUE DE LITTERATURE

2.1 REVUE DE LITTERATURE THEORIQUE

2.1.1 ORIENTATION SCOLAIRE

L'orientation scolaire est un accompagnement qui consiste à aider les jeunes dans le milieu éducatif et professionnel. Bénéficiaire d'un coaching en orientation scolaire aide les jeunes à réussir leur parcours scolaire [6]. Pierre Doray et ses compagnons, les définissent comme un processus qui se trouve à l'intersection de l'action des systèmes éducatifs et des dispositions et attributs sociaux et culturels des élèves et des étudiants. C'est l'élaboration des aspirations scolaires (choix d'un niveau d'études) ainsi que dans le choix d'une filière d'études, d'une discipline, d'un programme d'études ou d'une carrière [6].

2.1.2 INTELLIGENCE ARTIFICIELLE

L'Intelligence Artificielle (IA) est une branche du domaine informatique, ayant pour but la création de machines capables d'effectuer des tâches nécessitant traditionnellement une intelligence humaine. Son objectif est de répliquer ou de simuler l'intelligence humaine dans les machines [7]. L'IA définie par Louis Frecon et Okba Kazar comme l'art de simuler l'intelligence à l'aide d'ordinateurs, limite parfois son objectif à l'interprétation d'une situation ou d'un texte. Elle se situe d'abord entre la Psychologie Cognitive, dont elle s'inspire quant aux buts, et l'Informatique (ou plus précisément le Génie Logiciel) quant aux moyens matériels. Elle entretient ensuite des relations importantes avec la Linguistique soit officiellement lorsqu'elle s'occupe de traitement des langues naturelles, soit de façon plus banale lorsqu'il s'agit d'interaction avec l'utilisateur sous forme verbale, par exemple dans les didacticiels [8].

2.1.3 MACHINE LEARNING

Le ML est une sous-discipline de l'IA, née grâce aux technologies de reconnaissance de pattern et à la théorie selon laquelle les ordinateurs peuvent apprendre sans être programmés pour effectuer des tâches spécifiques [9]. Aussi appelé Apprentissage Automatique en français, est une forme d'IA permettant aux ordinateurs d'apprendre sans avoir été programmés explicitement à cet effet [7]. Cette technologie permet de développer des programmes informatiques pouvant changer en cas d'exposition à de nouvelles données. Par le biais d'algorithmes capables d'apprendre de manière itérative, le ML permet aux ordinateurs de découvrir des insights cachés sans être programmés pour savoir où les chercher [9].

2.1.4 DATA MINING

Data Mining est un terme anglo-saxon qui peut être traduit par exploration de données ou extraction de connaissances via des données [10]. C'est une famille d'outils automatiques ou semi-automatiques permettant l'analyse d'une grande quantité de données contenues dans une base de données. Son objectif est de faire apparaître des corrélations entre des phénomènes en apparence distincts afin d'anticiper des tendances [10].

2.1.5 QUELQUES ALGORITHMES DE ML

SVM

Une SVM, traduction littérale pour Support Vector Machine, est un algorithme d'apprentissage automatique supervisé utilisé souvent à des fins de classification et de régression [11]. Cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle. Elle est basée sur l'utilisation de fonction dites noyau qui permettent une séparation optimale des données. Le principe fondateur des SVM est d'intégrer à l'estimation de contrôle de la complexité, c'est-à-dire le nombre de paramètres qui est associé dans ce cas au nombre de vecteurs supports [11].

RESEAU DE NEURONES

Un réseau de neurones artificiel est une association d'un certain nombre des neurones formels. Le réseau est généralement organisé en couches successives où les sorties d'une couche constituent les entrées d'une autre [12]. Les entrées du premier niveau modélisent le problème que l'on veut résoudre (par exemple, en associant chaque pixel d'une image dont on veut reconnaître le contenu à une entrée) et les sorties du dernier niveau représentent la réponse attendue (par exemple 1 pour

dire que le système a reconnu un chat et 0 pour dire qu'il n'en a pas reconnu). Entre les deux, les valeurs calculées sont transmises d'une couche à la suivante. Les couches intermédiaires construisent ainsi une fonction mathématique qui doit représenter le système de reconnaissance [12].

ARBRE DE DECISION

L'Arbre de décision utilise les nœuds et les branches composés d'un arbre. Chaque nœud représente des caractéristiques dans une catégorie à classer et chaque sous-ensemble définit une valeur qui peut être prise par le nœud. Les nœuds internes sont représentés par des cercles, tandis que les feuilles sont désignées par des triangles. Soulignons que cet arbre de décision intègre à la fois des attributs nominaux et numériques. Avec ce classificateur, l'analyste peut prédire et comprendre les caractéristiques comportementales de l'ensemble de la population cible. Chaque nœud est étiqueté avec l'attribut qu'il teste, et ses branches sont étiquetées avec les valeurs correspondantes [13]. Le but premier d'un arbre de décision pour développer un modèle d'apprentissage est de prédire la valeur ou la classe d'une cible en comprenant les règles de décision fondamentales prises à partir de données plus anciennes, que les programmeurs appellent également données d'apprentissage [8].

NAÏVE BAYES

Naïve Bayes, couramment utilisé dans l'apprentissage automatique, est une collection d'algorithmes de classification basés sur le théorème de Bayes. Ce n'est pas un algorithme unique, mais une famille d'algorithmes. Tous ces algorithmes partagent tous un principe commun, à savoir que chaque caractéristique classée est indépendante de la valeur de toute autre caractéristique [14]. C'est un algorithme d'apprentissage supervisé utilisé pour la classification. Il est particulièrement utile pour les problématiques de classification de texte [14].

2.2 REVUE DE LITTÉRATURE EMPIRIQUE

Beaucoup de chercheurs se sont intéressés au domaine de la recherche sur l'exploitation de données éducatives, raison pour laquelle, cette recherche s'appuie d'une manière ou d'une autre sur leurs investigations. Neelam Naik et Seema Purohit avaient travaillé sur la prédiction du résultat final et placement des étudiants à l'aide d'un algorithme de classification dans l'objectif d'utiliser des algorithmes d'exploration de données pour prédire le résultat final et les placements des étudiants diplômés en fonction du dossier scolaire antérieur. Leur échantillon contenait environ 325 étudiants des instituts MCA à Mumbai [15]. La version XLMiner 3.2.8 avait été utilisée pour appliquer l'algorithme d'Arbre de classification sur les données. Enfin, leur conclusion était que l'erreur globale survenue lors de la classification des données de validation avaient été de 38,46 % (pour dire qu'il a réussi à 61.54%), tandis que pour la validation de l'arbre de classification de prédiction de placement, avait donné 45.38 % (Donc il avait prédit à 54.62%). Cette prédiction avait montré que le résultat final et le placement des étudiants peuvent être prédits avant de les admettre au cours [15]. Ely Rosely et Paramita Mayadewi dans leur article: "The decision support system of majors placement for students in high school using data mining C4.5 Algorithm" ont eu pour objectif d'aider les élèves à obtenir un enseignement adéquat et ciblé en fonction du potentiel de chacun. Les attributs de base étaient le nom de l'étudiant, le niveau de compétence en mathématique, sciences naturelles, sciences sociales et un test à l'examen du dossier d'études supérieures appelé « Tes Potensi Akademik » [16]. Ces deux chercheurs avaient utilisé l'algorithme C4.5 dans l'outil WEKA. Pour vérifier l'exactitude du modèle, ont évalué en utilisant la matrice de confusion. Ils ont construit leur application Web en utilisant le framework CodeIgnitor [16]. La prédiction de leur modèle a atteint 92.5% [16].

K. Sreenivasa Rao et les autres avaient présenté un cas d'étude intitulé "Educational data mining for student placement prediction using machine learning algorithms". Les données collectées pour leur dataset sont le numéro de série, SSC %, B.Tech % et le cube score. Pour construire le modèle, quelques algorithmes étaient exécutés dans l'outil WEKA. Il s'agit de J48, Naive Bayes, Random Forest et Decision Tree [17]. Dans R studio, ils avaient exécuté la régression linéaire, régression logistique, et arbre d'inférence conditionnelle. Les algorithmes de forêt aléatoire ont donné une précision de 100%, l'algorithme J48 à une précision de 88,89% et Naive Bayes 61,10%. Après analyse dans la phase de validation, deux algorithmes ont été retenus, notamment Random Forest qui avait donné 100% de précision et Decision Tree qui avait donné une précision de 90% de prédiction [17]. Partant des résultats de ces chercheurs antérieurs, les auteurs de cette présente étude ont remarqué que la majorité des recherches existantes ont considéré seulement les étudiants finalistes comme population de leurs études. Par conséquent, cette présente étude a pris pour unité d'analyse tous les étudiants qui ont déjà une expérience d'au moins une année dans des universités congolaises.

2.3 CADRE CONCEPTUEL

Cette recherche a utilisé 28 variables dont 27 indépendantes et une variable dépendante qui est la filière d'études à suivre à l'université. Les variables indépendantes sont regroupées en cinq facteurs, à savoir: le facteur démographique composé de trois variables qui sont: le genre, la confession et le cas de perte d'un des parents; le facteur cursus secondaire avec les variables telles que: le tuteur pendant le parcours secondaire, l'option suivie au secondaire, le grade obtenu à l'examen d'Etat, la province d'étude secondaire, l'Age avec lequel la personne a fini le secondaire, la révision des notes de cours, la nature de l'école secondaire, redoublant à sixième année secondaire; le facteur Admission à l'Université constitué de cinq variables à l'occurrence la filière d'études voulue avant l'université, la personne qui a orienté le candidat dans cette filière d'études, la volonté de s'inscrire dans cette filière d'études, le nombre d'années de repos après les études secondaires et de l'Age d'admission à l'université; le facteur cursus universitaire rassemblant les variables comme la nature de l'université, le type de l'institution, la province où se situe cette institution, le tuteur pendant les études universitaires et le niveau d'études actuel du répondant; et enfin le facteur compétences scolaires dans les cours qui met l'accent sur les variables en relation avec les compétences en mathématique, physique, biologie, chimie, informatique et dessin. D'où, le cadre conceptuel illustré par la figure ci-dessous:

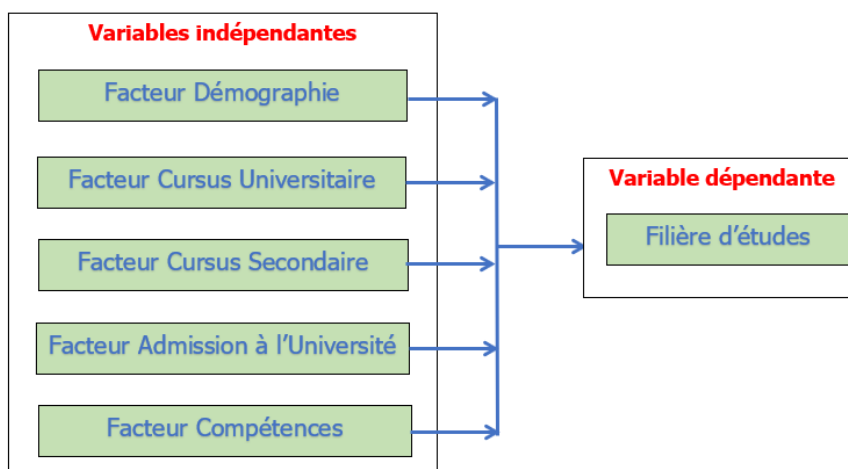


Fig. 1. Cadre conceptuel graphique

3 METHODOLOGIE

La méthode, en science, est la pierre angulaire de toute recherche car elle balise le chemin qu'emprunte le chercheur pour réaliser son projet [18]. Cette recherche décline et lie un certain nombre d'étapes qui peuvent être itératives et qui sont associées principalement à la finalité de cette recherche, à son orientation et à son cadre méthodologique [19].

3.1 CONCEPTION DE LA RECHERCHE

Cette recherche s'était basée sur la méthode quantitative. La terminologie quantitative désigne plusieurs choses, en particulier des données ou encore des techniques de collecte et de traitement de ces données. La recherche quantitative renvoie à des travaux, dont les données sont statistiquement analysables et surtout elle étudie de gros échantillons [19]. Les étapes de l'ensemble de cette investigation sont résumées et visualisée dans le diagramme sous-dessous:

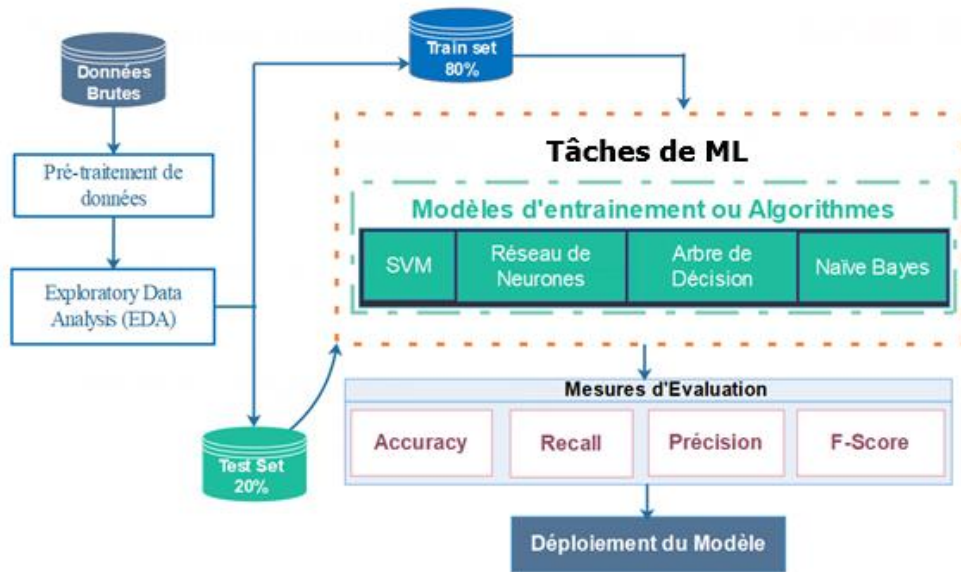


Fig. 2. Processus de mise en place du projet

Après l'élaboration du questionnaire de recherche de de cette étude, l'étape qui a suivi était la collecte des données. Cette collecte a permis les auteurs d'obtenir les données primaires. De ce qui précède, les données brutes ou primaires ont été vérifiées avec soin afin de déceler d'éventuelles erreurs. Le but a été d'éliminer les données de mauvaise qualité (comme les données redondantes ou incorrectes) et de commencer à créer les données de haute qualité qui peuvent garantir la qualité des modèles ML [20]. L'analyse exploratoire de donnée (AED) a été l'étape qui a permis les chercheurs de découvrir des relations dans les données et leurs valeurs. Une fois terminée l'AED, les fonctions établies ont été près pour être utilisées dans des analyses de données plus complexes, des modélisations plus sophistiquées, y compris l'apprentissage automatique [20]. L'ensemble de données ont été divisé en Train set et Test set. Spécifiquement, 80% de données au Train set et le 20% au Test set. Il s'en est suivi l'entraînement des modèles. Cette procédure a permis de sélectionner quatre algorithmes parmi lesquelles les auteurs ont choisi celui qui convient le mieux. L'étape de l'entraînement des modèles a été suivie par celle d'évaluations en utilisant diverses mesures telles que l'accuracy, le recall, le f-score et la précision. L'accuracy est le pourcentage de classifications correctes. Cette mesure est obtenue par la formule suivante [21]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Avec TP qui représente les valeurs des vrais positifs (true positive), FN les valeurs des faux négatifs (false negative), FP les faux positifs (false positive) et TN les valeurs des vrais négatifs (True negative). Au-delà de l'accuracy, les chercheurs ont également calculé le recall dont la formule est la suivante [22]:

$$\text{Recall} = \frac{TP}{TP+FN}$$

La précision est la proportion de résultats de reconnaissance qui sont de vraies détections de l'espèce cible et est calculée comme suit [23]:

$$\text{Precision} = \frac{TP}{TP+FP}$$

Le F-Score est la combinaison de la précision et du recall. Il est obtenu via la formule suivante:

$$\text{F-score} = \frac{(\beta^2+1) * \text{Precision} * \text{Recall}}{\beta^2 * \text{Precision} + \text{Recall}}$$

Où β est une métrique définie par l'utilisateur qui permet de donner la priorité à la précision par rapport au recall, ou vice-versa [24]. Les chercheurs ont utilisé Python dans l'implémentation du système.

3.2 POPULATION CIBLE

Comme cette étude visait à implémenter un modèle ML qui prédit l'orientation des nouveaux étudiants dans les domaines appropriés, le seul critère d'éligibilité pour participer à l'étude était d'être étudiants ou ressortissants des universités ou instituts supérieurs en RDC. Particulièrement, la population cible de cette recherche a été les étudiants du premier cycle, hormis ceux de première année de Licence (L1), autrement dit des promotions de L2, L3 et tous ceux qui ont déjà leurs diplômes de Licence dans les domaines informatiques et techniques.

3.3 COMPREHENSION ET PROCEDURE D'ECHANTILLONAGE

Dans le but de tirer des conclusions au sujet de la population cible, les chercheurs ont utilisé la technique d'échantillonnage aléatoire simple [25] et la formule de Cochran les a aidés à trouver l'échantillon [26]:

$$n_0 = \frac{Z^2 * p * q}{e^2}$$

Où Z^2 représente la valeur z trouvée dans la table z à un degré de confiance donné [27]. Le degré de confiance est de 95%, dans ce cas estimé à (1.96); p représente la proportion estimée d'un attribut qui est présent dans la population de cette recherche. Il est estimé à 0.5, q équivaut à $1-p$. Le degré de précision désiré est symbolisé par e^2 et les chercheurs ont pris 5%. Ainsi, la taille minimale de l'échantillon a été:

$$n_0 = \frac{(1.96)^2 (0.5) (1-0.5)}{(0.05)^2}$$

$$n_0 = \frac{3,8416 * 0,5 * 0,5}{0,0025}$$

Après le calcul, l'échantillon s'élève à 384,16 qui est arrondi à 384 répondants. Bien que l'échantillon calculé ait été de 384 répondants, les auteurs ont pu collecter 712 questionnaires. Les auteurs ont ainsi utilisé un dataset ayant 712 enregistrements.

3.4 INSTRUMENT DE RECHERCHE

Les chercheurs ont utilisé l'outil Google Forms pour administrer le questionnaire online. Ce questionnaire était composé de 28 questions fermées dont cinq questions binaires (qui ne proposent que deux options de réponse) et vingt-trois questions à échelle allant de 3 à 18. Ces questions sont reprises dans le tableau ci-dessous:

Tableau 2. Questions du questionnaire et leurs échelles de Likert

	Question	Echelle de Likert
1	Genre	2
2	Avec qui habitez-vous pendant votre parcours secondaire ?	8
3	Avez-vous déjà perdu un de vos deux parents ?	2
4	Quel est votre confession ?	10
5	Quelle est la nature de l'école secondaire où vous avez étudié votre école secondaire ?	8
6	Vous aviez fait quelle option à l'école secondaire ?	15
7	Grade obtenu à l'examen d'état	4
8	Province de votre école secondaire	18
9	A quel âge avez-vous fini votre école secondaire ?	11
10	Révisiez-vous vos notes de cours pendant votre parcours du secondaire ?	4
11	Que voulez-vous devenir avant la fréquentation de l'université ?	12
12	Qui vous a orienté dans le département fréquenté à l'université ?	9
13	Aviez-vous la volonté de faire ce département ?	2
14	Combien d'année de repos avant la fréquentation universitaire ?	6
15	Age de l'admission à l'Université ?	11
16	Quelle est la nature de l'université où vous avez fait votre cycle de graduat ?	8
17	Quel est le type de l'institution où vous avez étudié ?	2
18	Province de votre Université ou Institut Supérieur	18
19	Avec qui habitiez-vous pendant votre parcours universitaire ?	7
20	Quel est votre actuel niveau ?	3
21	Aviez-vous refait la sixième année secondaire ?	2
22	Niveau de compétence en mathématique quand vous étiez au secondaire	3
23	Niveau de compétence en physique quand vous étiez au secondaire	3
24	Niveau de compétence en biologie quand vous étiez au secondaire	3
25	Niveau de compétence en chimie quand vous étiez au secondaire	3
26	Niveau de compétence en informatique quand vous étiez au secondaire	3
27	Niveau de compétence en Dessin quand vous étiez au secondaire	3
28	Dans quel domaine avez-vous fini votre graduat (Licence) ?	15

3.5 PROCEDURE DE LA COLLECTE DES DONNEES

Toute la population n'avait pas le moyen de répondre au questionnaire en ligne, c'est l'une des causes qui ont poussé les chercheurs à imprimer le questionnaire à quelques-uns et à d'autres envoyer via Google Forms. Cet outil de Google a permis aux chercheurs de faire l'administration des questionnaires. Lors de la collecte des données à travers Google Forms, les chercheurs ont utilisé deux voies, à savoir: le téléphone portable et la machine ordinateur.

3.6 OUTILS DE TRAITEMENT ET D'ANALYSE DES DONNEES UTILISES

Tout le code python de cette étude a été développé en utilisant Jupyter Notebook. L'utilisation de Jupyter Notebook pour la science des données en Python, notamment pour l'apprentissage profond, se justifie par sa puissance, sa flexibilité et sa facilité d'utilisation [28]. Les bibliothèques Pandas, NumPy, Matplotlib, seaborn, Scikitlearn et pickle a permis de bien comprendre chaque variable, mais aussi de découvrir les effets croisés dans les données. Ces bibliothèques sont des modules ou bibliothèques de base que les chercheurs ont utilisée pour l'analyse et le développement des modèles ML.

Après entraînement des modèles avec python, les auteurs ont déployé le modèle validé dans la technologie web pour qu'il soit utilisé. Python et le web ne pouvait pas communiquer sans intermédiaire. Pour assurer cette communication, les auteurs ont utilisé la technologie Flask. Flask est un petit Framework web Python léger, qui fournit des outils et des fonctionnalités utiles qui facilitent la création d'applications web en Python [29]. De là, les auteurs ont disposé donc de tous les nécessaires pour répondre aux objectifs assignés au début de la recherche.

4 ANALYSE DES DONNEES, RESULTATS ET DISCUSSIONS

Cette section est consacrée d'abord à l'analyse des données, ensuite à la présentation des résultats comparativement aux objectifs spécifiques assignés dès le départ, et enfin à la discussion des résultats obtenus. De prime, des 28 items composant le questionnaire, les auteurs ont pu conduire la technique de feature selection pour sélectionner dans le jeu de données les variables qui prédisent le mieux l'orientation d'un étudiant dans une filière d'études [30]. Se basant sur les données récoltées pour cette étude, la figure ci-dessous donne une vue générale sur les 28 variables de cette étude et leur degré de pertinence dans la prédiction de l'orientation des étudiants dans des domaines techniques et informatiques après application de feature selection:

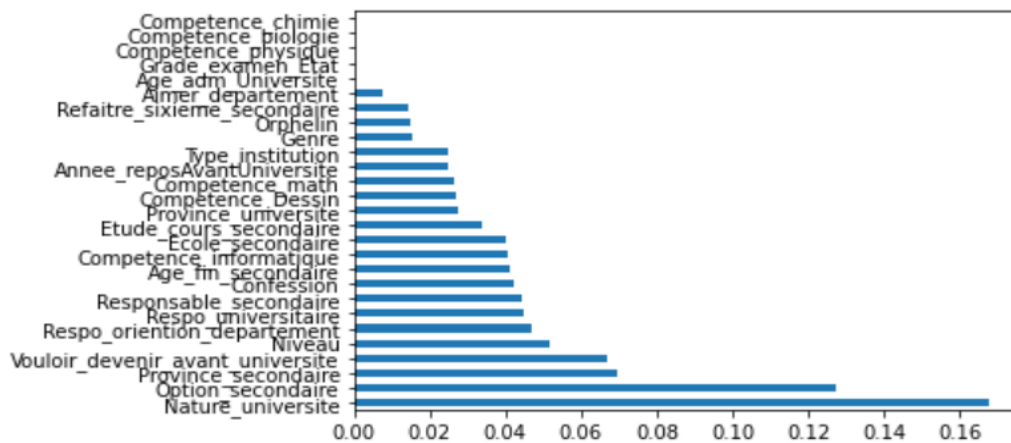


Fig. 3. Résultats de feature selection et visualisation des variables avec leur degré de pertinence

Partant de la figure ci-haut, on peut constater que les six premières variables (Competence_chimie, Competence_biologie, Competence_physique, Grade_examenEtat, Age_adm_Universite, Aimer_department) n'influencent pas la cible de cette étude. Par conséquent, les auteurs ne les ont pas considérées comme prédicteurs de l'orientation aux filières d'études. Ci-dessous se trouve le tableau reprenant seulement 22 meilleures variables et leur format de codage:

Tableau 3. Features utilisés et leur format de codage

Feature	Item	Codage
Genre	Genre	0=Masculin, 1=Féminin
Responsable_secondaire	Avec qui habitez-vous pendant votre parcours secondaire ?	0= seul (e), 1=les deux parents, 2= Maman, 3= Papa, 4= conjoint (e), 5= Amis, 6= internat, 7= Autre
Orphelin	Avez-vous déjà perdu un de vos deux parents ?	0=Non, 1=Oui
Confession	Quelle est votre confession ?	0=Catholique, 1=Protestante, 2=Musulmane, 3=Kimbanguiste, 4=Adventiste, 5= Evangélique, 6= Pentecôtiste, 7=Témoin de Jéhovah, 8=Salutiste, 9=Autre
Ecole_secondaire	Quelle est la nature de l'école secondaire où vous avez étudié votre école secondaire ?	0=Catholique, 1=Protestante, 2=Musulmane, 3=Kimbanguiste, 4=Adventiste, 5=Publique, 6=Privée, 7=Autre
Option_secondaire	Vous aviez fait quelle option à l'école secondaire ?	0=Commerciale, 1=Math-Physique, 2=Littéraire, 3=Biochimie, 4=Secrétariat et administratif, 5=Technique sociale, 6=Technique industrielle, 7=Technique informatique, 8=Pédagogie générale, 9=Technique agricole, 10=Technique professionnelle, 11= Coupe et couture, 12=Technique artistique, 13=Construction, 14= Autre

Province_secondaire	Province de votre école secondaire	0=Nord-Kivu, 1=Ituri, 2=Sud-Kivu, 3=Kinshasa, 4=Haut-Uele, 5= Kasai oriental, 6=Haut-Katanga, 7=Equateur, 8=Maniema, 9=Tshopo, 10=Kongo Central, 11=Bas-Uele, 12=Kasai, 13=Kwango, 14=Nord-Ubangi, 15=Lualaba, 16=Haut-Lomami, 17=Autre
Age_fin_secondaire	A quel âge avez-vous fini votre école secondaire ?	0=16 ans, 1=17 ans, 2= 18 ans, 3= 19 ans, 4= 20 ans, 5= 21 ans, 6= 22 ans, 7= 23 ans, 8=24 ans, 9=25 ans, 10=Plus
Etude_cours_secondaire	Révisiez-vous vos notes des cours pendant votre parcours du secondaire ?	0=Souvent, 1=Régulièrement, 2=Un peu, 3=Non
Vouloir_devenir_avant_universite	Que vouliez-vous devenir avant la fréquentation de l'université ?	0=Gestionnaire, 1=Ingénieur, 2=Entrepreneur, 3=Comptable, 4=Journaliste, 5= Capitaine de Bateaux, 6=Chercheur, 7=Commerçant, 8=Homme de droit, 9=Chef du projet, 10=Pilote, 11=Autre
Respo_orientation_departement	Qui vous a orienté dans le département fréquenté à l'université ?	0=Seul (e), 1=Papa, 2=Conjoint (e), 3=Amis, 4=Maman, 5=Deux parents, 6=Autorité académique, 7=Confrère (consoeur), 8=Autre
Annee_reposAvantUniversite	Vous aviez fait combien d'années de repos avant de fréquenter l'université ?	0=Moins d'un an, 1=1, 2=2, 3=3, 4=4, 5=Plus
Nature_universite	Quelle est la nature de l'université (institut supérieur) où vous avez fait votre cycle de graduat ?	0=Catholique, 1=Publique, 2=Protestante, 3=Privée, 4=Musulmane, 5=Adventiste, 6=Kimbanguiste, 7=Autre
Type_institution	Quel est le type de l'institution où vous avez étudié ?	0=Université, 1= Institut supérieur
Province_universite	Province de votre Université ou Institut Supérieur	0=Nord-Kivu, 1=Ituri, 2=Sud-Kivu, 3=Kinshasa, 4=Haut-Uele, 5= Kasai oriental, 6=Haut-Katanga, 7=Equateur, 8=Maniema, 9=Tshopo, 10=Kongo Central, 11=Bas-Uele, 12=Kasai, 13=Kwango, 14=Nord-Ubangi, 15=Lualaba, 16=Haut-Lomami, 17=Autre
Respo_universitaire	Avec qui habitez-vous pendant votre parcours universitaire ?	0=Seul (e), 1=Les deux parents, 2=Amis, 3=Maman, 4=Conjoint (e), 5=Papa, 6=Autre
Niveau	Quel est votre actuel niveau ?	2=Déjà Gradué, 1=G3, 0=G2, 1=L3, 2=Déjà Gradué ou licencié=2, 0=L2
Refaire_sixieme_secondaire	Aviez-vous refait la sixième année secondaire ?	0=Non, 1=Oui
Compétence_math	Niveau de compétence en mathématique quand vous étiez au secondaire	0=Bas, 1=Moyen, 2=Elevé
Compétence_informatique	Niveau de compétence en informatique quand vous étiez au secondaire	0=Bas, 1=Moyen, 2=Elevé
Compétence_Dessin	Niveau de compétence en Dessin quand vous étiez au secondaire	0=Bas, 1=Moyen, 2=Elevé
Domaine_Universite	Dans quel domaine avez-vous fini votre graduat ?	0=Informatique de gestion (Gestion informatique), 0=Télécommunication, 0=Génie informatique, 1=Génie civil, 1=Electronique, 1=Génie mécanique, 1=Electricité, 1=Architecture, 1=Bâtiment, 1=Construction, 2=Math-physique, 2=Math-Info, 3=Biologie, 3=Chimie, 3=Chimie industrielle, 4= Pas encore

4.1 INFORMATIONS DEMOGRAPHIQUES DES REpondANTS

Cette sous-section présente quelques informations socio-démographiques sur les répondants. Ainsi, le profil des répondants selon le mode de recueil choisi moyennant les graphiques explicatifs des réponses a été établi. Se basant du facteur genre, les auteurs ont constaté que les hommes ont dominé dans la recherche avec 59.13% et les femmes n’ont représenté que 40.87%. Soulignons que la variable "Orphelin" est celle qui informe plus sur l’existence ou la perte des parents du répondant. En effet, sur 712 réponses recueillies dans cette recherche, 70.08% ont encore la chance d’avoir les parents et 29.92% ont déjà perdu leurs parents.

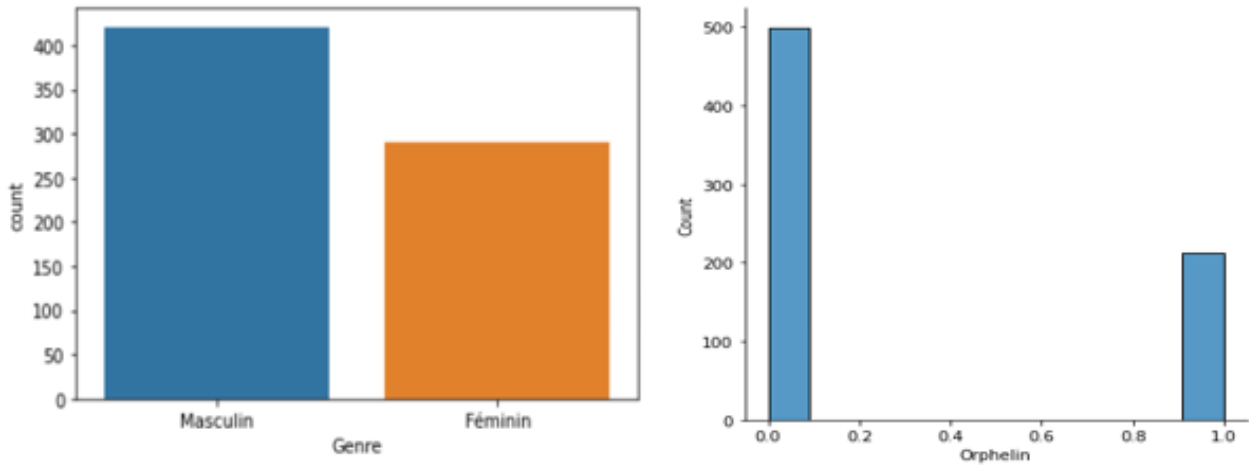


Fig. 4. L’image de gauche sont les statistiques des genres des répondants et celle de droite celles de la variable Orphelin

La variable confession a été également considéré comme facteur démographique des répondants. Les réponses obtenues montrent que les enquêtés de confession catholique représentent 44.24%, ceux de confession protestante ont été 25.14%, ceux de confession adventiste présentait un taux de 7.44%, ceux de confession évangélique 6.60%. Les musulmans et les kimbanguistes ont été 2.81%, les pentecôtistes 2.39%, les témoins de Jéhovah 2.11%, les salutiste 0.84%. Enfin, les répondants des autres confessions représentent 5.62%. D’où, le graphique ci-après:

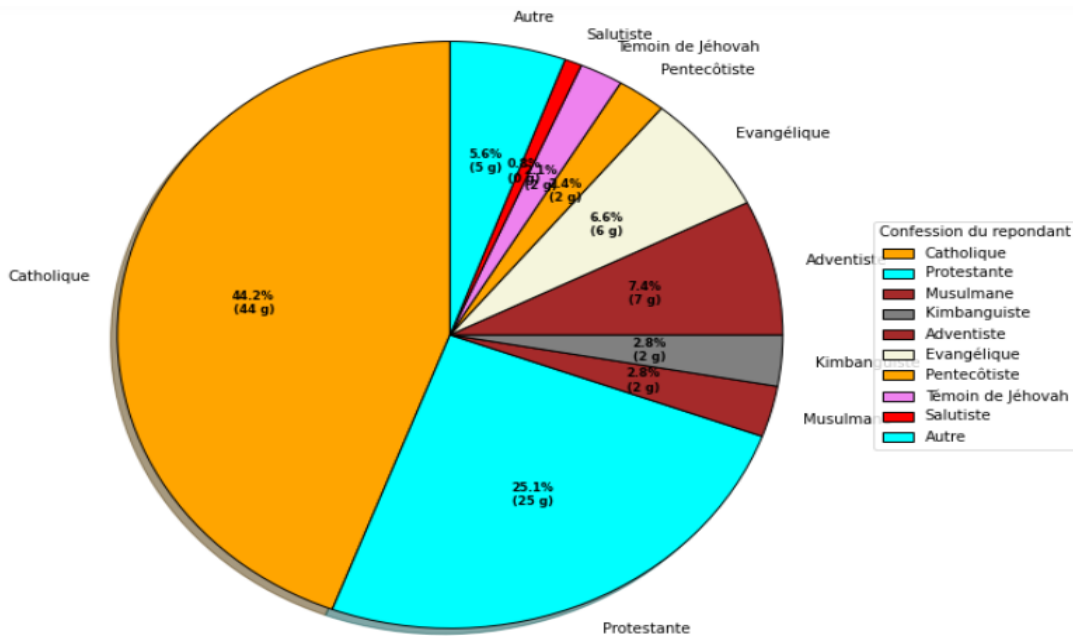


Fig. 5. Les statistiques des répondants par confessions

4.2 RESULTATS DE LA RECHERCHE

4.2.1 RESULTATS RELATIFS AU PREMIER OBJECTIF SPECIFIQUE

Cette section est une réponse au premier objectif qui a consisté à comprendre les théories qualitatives illustrant le processus d'orientation scolaire envie de construire un cadre conceptuel capturant les facteurs conditionnant un choix judicieux des filières techniques et informatiques en milieu universitaire. Pour atteindre cet objectif, les chercheurs ont lu quelques travaux connexes dans le pour comprendre pourquoi les recherches antérieures ont sélectionné tel ou tel autre facteur pour mener leurs études. Ces études ont aidé cette étude à construire le cadre conceptuel de cette étude. De ces facteurs, les auteurs ont repris quelques-uns qui ont été jugés important dans cette étude afin de construire leur cadre conceptuel qui est illustré dans la figure 1. Partant de ce cadre conceptuel, les auteurs ont pu construire l'instrument de leur recherche afin de collecter les données. Le tableau 2, dans sa deuxième colonne, illustre les 28 items qui ont fait objet des questions de l'instrument de recherche.

4.2.2 RESULTATS RELATIFS AU DEUXIEME OBJECTIF SPECIFIQUE

Les expériences ont été menées dans le but d'atteindre le deuxième objectif spécifique qui consistait à déterminer l'algorithme ML qui prédit le mieux possible l'orientation des futurs étudiants dans des filières techniques et informatiques.

SVM

Dans cette expérience, l'algorithme SVM a été utilisé sur diverses fonctions pour des proportions variées de données et les résultats des mesures de performance ont été comparés pour obtenir le accuracy de 70%; le score de precision a été de 0.64, le recall était de 0.7 et le f-score de 0.66.

NAÏVE BAYES

L'algorithme Naïve Bayes a eu pour score d'accuracy de 65%; le score de precision a été de 1, le recall a été également de 1 et le F-score a donné 1.

ARBRE DE DÉCISION

Après la construction de l'arbre de décision, les mesures de performances appliquées ont donné des résultats de 52% d'accuracy, 0.45 de precision, 0.52 de recall et un score de 0.46 pour la mesure de performance F-score.

RÉSEAU DE NEURONES

Dans cette expérience, après avoir trouvé les poids du modèle qui minimise la fonction coût, le réseau de neurones a été utilisé dans la prédiction. Après avoir entraîné ce modèle, les mesures de performances ont donné les taux de 64% d'accuracy, 0.35 de précision, 0.41 de recall et 0.37 de F-score.

Enfin, les chercheurs ont présenté et synthétisé les résultats des mesures de performance pour ces algorithmes dans le tableau ci-dessous:

Tableau 4. Sommaire des mesures de performance des modèles entraînés dans cette étude

	Modèle	Accuracy (%)	Recall	Precision	F-score
1	SVM	70	0.7	0.64	0.66
2	Naïve Bayes	65	1	1	1
3	Arbre de décision	52	0.52	0.45	0.46
4	Réseau de neurones	64	0.41	0.35	0.37

D'après le résultat de la comparaison dans le tableau ci-dessus, il a été évident que les résultats obtenus pour l'algorithme SVM ont été les meilleurs par rapport aux trois autres classificateurs du point de vue accuracy. Concernant les trois autres métriques, bien que Naïve Bayes ait atteint le score de 1 partout, les auteurs n'ont pas opté pour ce modèle par peur d'un

potentiel overfitting. Ce qui a mené à conclure que l'algorithme SVM est celui qui a prédit mieux que les autres dans notre recherche. Par conséquent, ce modèle (SVM) a été déployé dans un environnement Web.

4.2.3 RESULTATS RELATIFS AU TROISIEME OBJECTIF SPECIFIQUE

Après avoir opté pour le modèle SVM, les auteurs l'ont déployé en utilisant la technologie web. Le but de cet objectif était de développer une interface graphique de visualisation des tests du modèle retenu pour permettre aux nouveaux étudiants, désirant s'inscrire à une quelconque filière technique ou informatique dans des universités et instituts supérieurs de la RDC, de pouvoir prédire la filière technique ou informatique adéquate. De ce qui précède, les auteurs ont développé une interface Web en utilisant HTML et Flask. Ci-dessous on voit la page d'accueil permettant aux utilisateurs d'exploiter le modèle en vue de prédire leurs orientations :

The screenshot shows a web application interface for student orientation prediction. The page title is "Orientation des étudiants à l'Université". The navigation menu includes "Accueil", "Prediction", and "Contact". The main heading is "Predire votre orientation dans des sciences Techniques et Informatique à l'Université". The form contains 20 input fields for personal and academic information, and a "Predire" button.

Votre Genre	A quel âge avez-vous fini le secondaire?	Vous aimez les sciences techniques et Informatiq
Votre Responsable au Secondaire	Révisiez-vous vos notes de cours ?	Province de votre futur Université
Est-Vous Orphelin?	Que voulez-vous devenir ?	Futur responsable à l'université
Quelle est votre Confession?	Qui va vous orienté à l'université?	Aviez-vous refait la sixième année ?
Nature de votre Ecole Secondaire	Années de répos après votre Diplome	Vos compétences en Dessin
Votre Option au secondaire	Nature de l'université de votre choix	Vos compétences en Informatique
Province de votre Secondaire	Type de l'institution de votre choix	Vos compétences en Mathématique
Quel est votre Niveau actuel?	Predire	

Fig. 6. Illustration de la page d'accueil de l'application Web

Une fois accédé à la page d'accueil ci-haut, l'utilisateur va compléter le formulaire et puis cliquer sur le bouton *Predire*. Soulignons que les champs de ce formulaire sont des ComboBox ou de type choix multiple (pour dire qu'il suffit de cliquer dans le champ et choisir une des reponses). Le resultat de la prédiction se presente doublement, avant et à la fin du formulaire comme suit:

Orientation des étudiants à l'Université

Accueil Prediction Contact

Predire votre orientation dans des sciences Techniques et

Masculin	19 ans	Oui
Les deux parents	Souvent	Nord-Kivu
Oui	Ingénieur	Seul(e)
Catholique	Seul(e)	Non
Catholique	Moins d'un an	Elevé
Math-Physique	Catholique	Elevé
Nord-Kivu	Université	Moyen
Déjà Diplômé		

Predire

Votre orientation est la Polytechnique

Copyright 2022. Kanduki Kivuyirwa Mystère. Reserver à tous. kivuyirwamystere@gmail.com

Fig. 7. Illustration des résultats d'une prédiction

Pour le cas de l'exemple de la figure 7, on voit que les résultats d'un étudiant ayant pour genre masculin, pour responsable au secondaire les deux parents, qui est orphelin, dont la confession religieuse est catholique, qui a fait son école secondaire dans une école catholique, qui avait suivi l'option math-physique au secondaire, dans la province du Nord-Kivu, qui est déjà diplômé, qui avait fini l'école secondaire à 19 ans, qui révisait ses notes de cours souvent, qui voulait devenir ingénieur, qui avait pris la décision seul de son orientation scolaire à l'université directement après le diplôme d'Etat, dans une école catholique de type université, qui aimait les sciences techniques et informatiques, qui se débrouillait seul à l'université, qui n'avait pas échoué la sixième année secondaire, qui avait des compétences élevées en dessin, en informatique et des compétences moyen en mathématique, le modèle a prédit qu'il est destiné à s'inscrire dans le domaine de *polytechnique*.

5 CONCLUSION ET RECOMMANDATIONS

Cette recherche a permis aux auteurs d'aborder les problèmes liés à l'orientation des étudiants dans des filières des techniques et celles d'informatique dans des universités, problèmes qui étaient un vrai embarras pour les étudiants car les résultats obtenus au niveau des études secondaires ne leurs permettent pas souvent de s'inscrire dans une des telles filières. Par conséquent, tels étudiants parviennent à faire des choix arbitraires des domaines inadéquats. Ces mauvais choix des étudiants constituaient la cause des échecs à la fin de l'année académique, de l'abandon des cours à caractère scientifique et technique, de l'incapacité d'entreprendre après le parcours académique, de l'inadaptation dans des entreprises, etc. Ainsi, cette recherche s'est basée sur la méthode quantitative pour développer un modèle ML capable de prédire la filière technique ou informatique appropriée pour des tels étudiants partant des données obtenues auprès des diplômés en informatique et en technique et auprès des étudiants de L2 et L3. Pour aboutir aux résultats finaux, cette recherche lié un certain nombre d'étapes.

Par surcroît de la méthode quantitative, cette étude s'est fondée sur 712 enregistrements des données collectées en utilisant google Forms comme outil d'enquête. Les auteurs ont analysé, implémenté et déployé le modèle SVM.

Du point de vue contributions, la première observation distinctive faite dans le cadre de la présente étude se base sur les travaux connexes qui prenaient seulement les étudiants finalistes comme population de leurs études alors que cette étude a prouvé que tous les étudiants qui ont déjà une expérience d'au moins une année à l'université impacte une telle recherche.

Après les recherches et une analyse approfondie, les chercheurs ont compris que les variables des études antérieures ne suffisaient pas pour une recherche de ce genre; ainsi, ils ont ajouté une liste des variables qui ont été groupés en cinq facteurs afin de construire un nouveau cadre conceptuel adapté. Un grand constat a été fait dans la mesure de performance des modèles. Les prédécesseurs se contentaient d'une seule mesure de performance pour évaluer leurs modèles de prédictions, par contre, cette étude a utilisé quatre modèles comme souligné dans la section des résultats. Après comparaison de différents modèles, les auteurs ont déployé dans la technologie web en utilisant Flask leur modèle, ce qui n'avait pas été fait dans les études antérieures.

Des nombreuses recherches menées depuis longtemps ont estimé que la question du processus d'orientation dans des filières d'études supérieures était multidimensionnelle et complexe. Pour vouloir dire que conceptualiser la notion de choix d'étude n'est pas aisée et renvoie à des enjeux théoriques et méthodologiques complexes [3].

Partant des résultats de cette présente étude, il est à souligner qu'il y a plusieurs nécessités de continuer à améliorer cette recherche car la problématique développée dans la section introduction reste pleine et entière et permet de dégager plusieurs pistes de réflexions. Au futur, il faudrait récolter plus des données en vue d'augmenter la performance des modèles d'apprentissage. En plus, on pourra déployer ce modèle dans une autre technologie telle que la technologie mobile, etc.

REFERENCES

- [1] Lardy L., Bressoux P., Lima P. Factors influencing students' academic success in a technological university course: A freshman year study. *Orientation scolaire et professionnelle*, 2016, 44 (4).
- [2] Mpia H.N. Des griefs de l'échec de l'enseignement de l'informatique dans des universités de la RDC: Enjeux socio-didactiques de la crise. *Etincelle*, 2020, (22), pp.107-131.
- [3] Christine F. Les facteurs déterminant les choix d'études supérieures. *Orientation et parcours des filles et des garçons dans l'enseignement supérieur*, 2016, pp.112-174. <https://doi.org/10.4000/books.purh.1554>.
- [4] Haas V., Morin-Messabel C., Fieulaine N., Demoures A. L'entrée à l'université et ses difficultés: Regards croisés à partir d'un dispositif du plan réussite licence. *Orientation scolaire et professionnelle*, 2015, 41 (4), pp.1-23. <https://doi.org/10.4000/osp.3925>.
- [5] Troyes D. En architecture, le danger est de penser que l'école est le reflet du monde professionnel et que les études se suffisent à elles-mêmes, *Ecole des grands projets*, Paris, Edition ESTP, 2018.
- [6] Picard F., Olympio N., Masdonati J., Bangali M. Justice sociale et orientation scolaire: l'éclairage de l'approche par les 'capabilités' d'Amartya Sen. *Orientation scolaire et professionnelle*, 2015, 44 (1). <https://doi.org/10.4000/osp.4515>.
- [7] Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health*, 2018, 8 (2): 020303. <https://doi.org/10.7189/jogh.08.020303>.
- [8] Frecon L., Kazar O. Manuel d'Intelligence Artificielle, *Presses Polytechniques et Universitaires Romandes*, METIS LyonTechn, Première édition, 2009.
- [9] Sree Divya K., Bhargavi P., Jyothi S. Machine Learning Algorithms in Big data Analytics. *International Journal of Computer Sciences and Engineering*, 2018, 6 (1), pp.63-70.
- [10] Miao F., Xie X., Wu Y., Zhao F. Data Mining and Deep Learning for Predicting the Displacement of «Step-like» Landslides. *Sensors*, 2022, 22 (2): 481. <https://doi.org/10.3390/s22020481>.
- [11] Chloé-Agathe A., Introduction au Machine Learning, Dunod, Paris, 2018.
- [12] Mpia H.N., Inipaivudu B.N. L'Algorithme de rétro-propagation de gradient dans le perceptron multicouche: Bases et étude de cas. *International Journal of Innovation and Applied Studies*, 2021, 32 (2), pp. 271-290.
- [13] Bahzad T.J, Adnan M.A, Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2021, 2 (1), pp. 20-28. <https://doi.org/10.38094/jastt20165>.
- [14] Viet T.N, Le Minh H., Le Cong H., Anh T.H. The Naïve Bayes Algorithm for Learning Data Analytics. *Indian Journal of Computer Science and Engineering*, 2021, 12 (4), pp.1038-1043. <https://doi.org/10.21817/indjcse/2021/v12i4/211204191>.
- [15] Neelam N., Seema P. Prediction of final result and placement of students using classification algorithm. *International Journal of computer applications*, 2012, 56 (12), pp.35-40. <https://doi.org/10.5120/8945-3111>.

- [16] Mutiara S.S., Ely R., Paramita M. The decision support system of majors placements for students in high school using data mining C4.5 algorithm. *Science & technology*, 2016, 4 (1), pp.1-10.
- [17] Kuncham S.R., Swapna N., Praveen Kumar P. Educational data mining for student placement prediction using machine learning algorithms. *International journal of engineering & technology*, 2017, 7 (1-2), pp.43-46. <https://doi.org/10.14419/ijet.v7i1.2.8988>.
- [18] Denise T. Méthodes qualitatives, Quantitatives et Mixtes, L'approche Delphi: Application dans la conception d'un outil clinique en réadaptation au travail en santé mentale, Presses de l'université du Québec, Paris, 2014.
- [19] Yvonne G., Alain J. Pourquoi je préfère la recherche quantitative/Pourquoi je préfère la recherche qualitative ? *Revue internationale P.M.E*, 2016, 29, pp.7-17.
- [20] Kabita S., Abhaya K.S., Jitendra P., Subhendu K.P. Exploratory Data Analysis using Python. *International Journal of Innovative Technology and Exploring Engineering*, 2019, 8 (12), pp. 4727-4735. <https://doi.org/10.35940/ijitee.L3591.1081219>.
- [21] Maxwell A.E., Warner T.A., Guillén L.A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. *Remote Sensing*, 2021, 13 (13): 2450. <https://doi.org/10.3390/rs13132450>.
- [22] Padilla R., Passos W.L., Dias T.L.B., Netto S.L., da Silva E.A.B. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics*, 2021, 10 (3): 279. <https://doi.org/10.3390/electronics10030279>.
- [23] AlZoman R.M., Alenazi M.J.F. A Comparative Study of Traffic Classification Techniques for Smart City Networks. *Sensors*, 2021, 21 (14): 4677. <https://doi.org/10.3390/s21144677>.
- [24] Knight E.C. et al., Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology*, 2017, 12 (2): 14. <https://doi.org/10.5751/ACE-01114-120214>.
- [25] Firdion, J.M. Construire un échantillon. L'enquête sociologique, Presses universitaires de France, Paris, 2010, pp.71-92.
- [26] Mballo A.D. et al. Pastoralisme et Défis statistiques: Vers une méthodologie d'échantillonnage en milieu mobile. Cas des transhumants du Ferlo (Sahel Sénégal), 2019, UQO. <https://agris.fao.org/agris-search/search.do?recordID=FR2019153937>.
- [27] Anderson et al. Statistiques pour l'économie et la gestion, Données et statistiques, 5^{ème} édition, Paris, 2015.
- [28] Mpia H.N., Mwendia S.N., Mburu L.W. Predicting Employability of Congolese Information Technology Graduates Using Contextual Factors: Towards Sustainable Employability. *Sustainability*, 2022, 14 (20): 13001. <https://doi.org/10.3390/su142013001>.
- [29] Uzo I.U. et al. Exploring a Secured Socket Python Flask Framework in Real Time Communication System. *Asian Journal of Research in Computer Science*, 2021, 8 (1), pp. 77-87. <https://doi.org/10.9734/AJRCOS/2021/v8i130194>.
- [30] Albattah W., Khan R.U., Alsharekh M.F., Khasawneh S.F. Feature Selection Techniques for Big Data Analytics. *Electronics*, 2022, 11 (19): 3177. <https://doi.org/10.3390/electronics11193177>.