

# Implementation of a machine learning-based system for the orientation of secondary school students

KABEYA TSHISEBA Cedric

Associate Professor at DRC Pedagogic National University

Faculty of Sciences

Department of Mathematics and Computing Science

DOI: <https://doi.org/10.5281/zenodo.7462544>

Published Date: 20-December-2022

---

**Abstract:** Machine learning is a form of artificial intelligence that focuses on creating systems that learn, or improve their performance, based on the data they process. Artificial intelligence is a broad term that refers to systems or machines that simulate a form of human intelligence.

Today, we use machine learning in all areas. When we interact with banks, shop online or use social media, machine learning algorithms come into play to optimize, streamline and secure our experience. Machine learning and the technology around it are developing rapidly, and we are only beginning to glimpse its capabilities. It is exactly within this same framework that we proposed to be able to set up an intelligent system based on artificial intelligence, or precisely on machine learning, to allow schools in difficulty to have one or more orientation advisers on a permanent basis., to be equipped with this service, in order to help secondary school students moving on to the humanities, to be able to choose their options.

**Keywords:** machine learning, artificial intelligence, orientation advisers, secondary school students.

---

## I. INTRODUCTION

The school population of Kinshasa being very dense, by reading and following all the observations and criticisms on the quality of education, then understanding that school orientation constitutes an important factor in the life of the learner, we have seen fit as an actor and scientist to mark our step. Our intellectual curiosity led us to ask ourselves the following series of questions:

- Is there a guidance policy within the primary, secondary and technical education services (EPST), at the level of the city-province?
- Are the school guidance methods and techniques used effective?

Considering the above questions and the realities on the ground of our investigation or observation, we noted that:

- The guidance service exists and has less qualified staff;
- The virtual non-existence of educational guidance for learners in schools;
- Lack of a powerful software tool that can help guidance counselors provide options to learners across the entire Branch and across the ministry.

Through this research, we want to:

- Initiate an Artificial Intelligence Project that can be used in the educational field;
- Design and implement a system capable of predicting a learner's option in terms of school results;

## II. KNN (K nearest neighbors)

Before focusing on the KNN algorithm, it is necessary to go over the basics. What is a supervised learning algorithm?

### A. Supervised learning

In this mode of learning, the data is labeled and the system learns to classify according to a classification model. Ex: discriminant analysis or SVMs.

### B. Unsupervised learning

The algorithm discovers itself, the system only has examples but not labels and that the number of classes and their nature are not predetermined. No expert is required. The knn The K-nearest neighbors (kNN) algorithm is a Machine Learning algorithm that belongs to the class of simple and easy-to-implement supervised learning algorithms that can be used to solve classification problems.

In supervised learning, an algorithm receives a set of data that is labeled with corresponding output values on which it will be able to train and define a prediction model. This algorithm can then be used on new data to predict their corresponding output values.

### C. K nearest neighbors

The intuition behind the K nearest neighbors algorithm is one of the simplest of all supervised machine learning algorithms:

Step 1: Select the number K of neighbors

Step 2: Calculate the Euclidean distance

Step 3: Take the K nearest neighbors according to the calculated distance.

Step 4: Among these K neighbors, count the number of points belonging to each category.

Step 5: Assign the new point to the category most present among these K neighbors.

Step 6: Our model is ready:

## III. GUIDANCE COUNSELING BASED ON A MACHINE LEARNING SYSTEM

In this part of this article, we present the way in which the reflection in question in this work was made.

Our Dataset:

```
Entrée [4]: import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
df = pd.read_excel('C:/OPTION.xlsx')
df.head()

Out[4]:
```

|   | Numero | NOM      | POST NOM | ECOLE  | LANGUES | SCIENCES | C.GENERALES | MOYENNE | code_Option | OPTION             |
|---|--------|----------|----------|--------|---------|----------|-------------|---------|-------------|--------------------|
| 0 | 1      | LIONGO   | ESENGO   | ELIKYA | 8       | 4        | 3           | 6.0     | 1.01        | Latin-Philosophie  |
| 1 | 2      | KAYEMBE  | DIKA     | ELIKYA | 5       | 8        | 3           | 6.4     | 1.02        | Math-Phys          |
| 2 | 3      | MOTEMO   | EMEKA    | ELIKYA | 3       | 8        | 2           | 5.2     | 1.03        | Biologie-Chimie    |
| 3 | 4      | GIDA     | LIFIO    | ELIKYA | 8       | 3        | 2           | 5.2     | 1.04        | GREC-LATIN         |
| 4 | 5      | DEDETEMO | SETIA    | ELIKYA | 8       | 6        | 2           | 6.4     | 1.05        | LATIN MATHEMATIQUE |

```
Entrée [ ]:
```

Fig N° 1: Display of our Dataset.

### A. PREDICTION RULES

The Congolese education system organizes about fifty options. In order to predict the learners' option, we propose the following rules:

- The Orientation Test or TENASOSP is based on the following 03 learning areas: Languages, Sciences and General Cultures;
- Pupils who have obtained at least 50% of the points in the 3 different sub-fields are eligible for the options: Mathematics-Physics, Latin-Philosophy, Social, Normal Pedagogy, Commercial and Administrative;
- Students who have exclusively obtained at least 50% in Languages and General Cultures are eligible for the options: Secretarial, Latin-Philosophy, Greek-Latin, Physical Education, Music, Accommodation, Tourism, Hostess, Hospitality and catering, Pedagogy kindergarten drama.
- Students who have exclusively obtained at least 50% in Languages and Sciences are eligible for the options: Scientists, Biology-Chemistry, Mathematics-Physics, Commercial and Management, Commercial and Computer Science, Computer Secretariat, Agroforestry, Agriculture, Horticulture, Veterinary, Food industry, Nutrition, fishing, Agricultural economics, Agricultural mechanics, General mechanics, Electricity, Construction, Industrial chemistry, Printing, Electronics, Radio-transmission, Meteorology, Civil aviation, Mechanical drawing, Hydropneumatics, Automotive mechanics, Metal construction, Carpentry, Cutting and sewing, Aesthetics and Hairdressing, plastic arts, Petrochemicals.
- Pupils who have obtained less than 50% of the points in the 3 different sub-fields are declared ineligible for the humanities and are obliged to repeat the 8th year of the Terminal Cycle of Basic Education.
- Pupils who have obtained at least 70% of the points in Sciences and less than 50% in the other 2 with an overall average greater than or equal to 50% of the total maximum of the test are exclusively eligible for the options: Scientists, Industrial and agricultural techniques .
- Pupils who have obtained at least 70% of the points in Languages and less than 50% in the other 2 with an overall average greater than or equal to 50% of the total maximum of the test are exclusively eligible for the options: Normal pedagogy, General pedagogy, Kindergarten, Social, Hostess, Accommodation, Hotels and restaurants, Literature.
- Students who have obtained at least 70% of the points in General Cultures and less than 50% in the other 2 with an overall average greater than or equal to 50% of the total maximum of the test are exclusively eligible for the options: Tourism, Music and Social.

### B. Example of option prediction

Example of option prediction, for a student who presents himself to the system, and who has obtained the following ratings:

- 5 in language,
- 5 in Science and
- 5 in General Culture

```
prediction_option = knn.predict([[5,5,5]])
option[prediction_option[0]]
```

```
{1: 'Latin-Philosophie', 2: 'Math-Phys', 3: 'Biologie-Chimie', 4: 'GREC-LATIN', 5: 'LATIN MATHEMATIQUE', 6: 'SCIENTIFIQUE', 7: 'Pédagogie Générale', 8: 'EDUCATION PHYSIQUE', 9: 'PEDAGOGIE NORMALE', 10: 'PEDAGOGIE MATERNELLE', 11: 'Commerciale et Gestion', 12: 'SECRETARIAT', 13: 'Commerciale et informatique', 14: 'SECRETARIAT INFORMATIQUE', 15: 'Technique Sociale', 16: 'ARTS PLASTIQUES', 17: 'ARTS DRAMATIQUES', 18: 'MUSIQUE', 19: 'ESTHETIQUE ET COUTURE', 20: 'Coupe et couture', 21: 'HOTESSE D'ACCEUIL', 22: 'HOTELLERIE ET RESTAURATION', 23: 'HEBERGEMENT', 24: 'TOURISME', 25: 'AGRICULTURE', 26: 'HORTICULTURE', 27: 'VETERINAIRE', 28: 'INDUSTRIE ALIMENTAIRE', 29: 'NUTRITION', 30: 'FORESTERIE', 31: 'PECHE ', 32: 'ECONOMIE AGRICOLE', 33: 'AGROFORSTERIE', 34: 'MECANIQUE AGRICOLE', 35: 'MECANIQUE GENERALE', 36: 'MECANIQUE MACHINES-OUTILS', 37: 'ELECTRICITE', 38: 'CONSTRUCTION', 39: 'CHIMIE INDUSTRIELLE', 40: 'ELECTRONIQUE', 41: 'IMPRIMERIE', 42: 'COMMUTATION', 43: 'RADIO TRANSMISSION', 44: 'METEOROLOGIE', 45: 'AVIATION CIVILE', 46: 'MECANIQUE DESSIN', 47: 'HYDROPNEUMATIQUE', 48: 'PETROCHIMIE'}
```

```
Out[10]: 'TOURISME'
```

Fig N° 2: Prediction for 5, 5 and 5.

### C. SOURCE CODE

```

import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
df = pd.read_excel('C:/OPTION.xlsx')
df.head()
option = dict( zip (df.Number.unique(), df.OPTION.unique()))
print(option)
x = df [['LANGUAGES', 'SCIENCES', 'C. GENERAL']]
y = df['Number']
x_train, x_test, y_train, y_test = train_test_split(x, y, random_state=0)
from sklearn.neighbors import KNeighborsClassifier
#instantiation and definition of k
knn = KNeighborsClassifier(n_neighbors = 3)
#learning
knn.fit(x_train,y_train)
#Prediction of the option for a future student who has 5 in language, 5 in Science, 5 in General Culture
prediction_option = knn.predict([[5,5,5]])
option[prediction_option[0]]

```

### IV. CONCLUSION

This small reflection focused on the implementation of an artificial intelligence system, mainly using the nearest k algorithm for the prediction of the learner's option on the basis of his school results, and constitutes a sustainable solution. and effective in improving the quality of education so decried by the scientific public and will avoid any subjectivity in the educational orientation of these learners.

Several perspectives arise in the face of this question addressed here, in particular that of the complexity of our dataset on the one hand, and on the other hand of the use of other algorithms well beyond the K nearest neighbor used here.

### REFERENCES

- [1] Discala, The basics of computing and programming, 1208 pages.
- [2] Anonymous; Universal Dictionary, Hachette, Paris, 2010.
- [3] Anonymous; Illustrated Larousse Dictionary, Special Edition DRC, Kinshasa, 2010.
- [4] Doctor G. Cimbela KABONGO Professor, Special questions of research methodology oriented in pure and applied sciences at the 3rd cycle, UPN? Department of Physics and Applied Sciences, Kinshasa 2014, Unpublished.
- [5] Professor Jean Paul YAWIDI, Ethics and Deontology of Professional Guidance Counselor, Editions Saint Paul, Kinshasa-Limete.
- [6] J. Afi-Pine, Machine Learning, M2DM 2019/2020, page 38. Accessed at [www.google.com](http://www.google.com), 03/15/2022 at 1:00 p.m.

**International Journal of Novel Research in Computer Science and Software Engineering**

Vol. 9, Issue 3, pp: (24-28), Month: September - December 2022, Available at: [www.noveltyjournals.com](http://www.noveltyjournals.com)

- [7] KPAKPO DONGE Everard, Course Notes of Principles of Educational and Professional Orientations, G3 All, ISP/Gemena, 2020-2021, Unpublished.
- [8] NTUKADI MOMBO A., Doctor, KABAMBA MBIKAYI Hervé, Head of Works, Artificial Intelligence Course Notes, L1 Informatique, ISC, 2019, Unpublished.
- [9] Professor Doctor KABEYA TSISEBA Cédric, Artificial Intelligence Course Notes, L1 Mathematics-Computer Science, ISP/Gemena, 2020-2021, Unpublished.
- [10] Professor Doctor KABEYA TSISEBA Cédric, Lecture Notes in Computer Systems Design, L2 Mathematics-Computer Science, ISP/Gemena, 2020-2021, Unpublished.
- [11] Professor Doctor KABEYA TSISEBA Cédric, Course Notes for language, automata and compilation, L2 Mathematics-Computer Science, ISP/Gemena, 2020-2021, Unpublished.