Combining Fuzzy Logic and k-Nearest Neighbor Algorithm for Recommendation Systems

Paul Dayang, Cyrille Sepele Petsou, Damien Wohwe Sambo
Department of Mathematics and Computer Science, Faculty of Sciences, The University of Ngaoundéré, Cameroon
E-mail: pdayang@univ-ndere.cm, cyrillesepele3@gmail.com, wsdamieno@gmail.com

Abstract: Recommendation systems are a type of systems that are able to help users finding relevant and personalized content in a wide variety of possibilities. To help computers perform recommendations, there are several approaches used nowadays such as the Content-based approach, the Collaborative filtering approach and the Hybrid recommendation approach. However, these approaches are sometimes inappropriate for use cases where there is no prior large datasets of users' feedbacks or ratings needed for training Machine Learning models. Thus, in this work, we proposed a novel approach based on the combination of Fuzzy Logic and the k-Nearest neighbor algorithm (KNN). The proposed approach can be applied without any prior collected feedbacks of users and performs good recommendations. Moreover, our proposal uses Fuzzy Logic to infer values based on inputs and a set of rules. Furthermore, the KNN uses the output values of the Fuzzy Logic system to do some retrieval tasks based on existing distance measures. In order to evaluate our approach, we considered an expert system of food recommendation for people suffering from the two deadliest diseases in Cameroon: HIV/AIDS and Malaria. The obtained results are closed to the recommendation made by nutritionists. These results demonstrate how effective our approach can be used to solve a real nutrition problem for people suffering from Malaria or HIV/AIDS. Furthermore, this approach can be extended to other fields and even be used to perform any recommendation task where there is no prior collected user's feedback or ratings by using the proposed approach as a framework.

Index Terms: Fuzzy logic, k-Nearest Neighbor, Recommendation systems, Dish recommendation, Nutrition recommendation for HIV/AIDS, Nutrition recommendation for Malaria.

1. Introduction

Nowadays, the huge amount of information available online makes it difficult to access a specific or a personalized content. For that reason, the scientific community have developed information filtering techniques such as recommendation systems which helps users retrieved prioritized and personalized items in a wide variety of possibilities [1]. These recommendation systems are used in various fields among which we have e-commerce stores, e-learning platforms, e-library stores, e-business services [2] and the nutrition domain [3]. For developing these recommendation systems there are various techniques and approaches used but the main existing ones are the content-based approach, the collaborative filtering approach and the hybrid approach which is a combination of the two previous approaches. However, these approaches unfortunately assume the availability of users feedbacks datasets or ratings which cannot be sometimes accessible and therefore makes the application of these techniques not suitable. For performing a recommendation task, the main existing algorithms are the approaches based on Machine Learning such as Content based recommendation system, the collaborative filtering approach and finally the hybrid approach which combines the Content-based approach and the collaborative one. Since these algorithms are Machine Learning methods there is a need of creating datasets of previous recommendations or user's feedback. But in some context, there is no available datasets and therefore the application or use of these approaches are not possible. So, our aim is to develop a novel approach which can be applied without any prior collected datasets of user's feedback or ratings to tackle the problem of recommendation. To attain our objective, we proposed an approach which uses both Fuzzy logic and the k-Nearest Neighbour algorithm (KNN). This particular combination has been chosen because: firstly, Fuzzy logic can handle the classification of raw data without any thresholds and let users define explicitly rules to be applied to perform any task of recommendation, secondly the k-Nearest Neighbour algorithm is one of the simplest and most common classifiers, yet its performance competes with the most complex classifiers in the literature [4]. In this combination the KNN algorithm is used for computing similarities between data points using a similarity measure.

The rest of this paper is organized as follows. Section 2 gives an overview of existing similar approaches and their associated pros and cons. Section 3 is where we describe our architecture and show how it can efficiently be used for the task of recommendation. Section 4 gives the results and the validation of the application of our approach in the dish

recommendation for people living with HIV/AIDS and Malaria in Cameroon. Finally, Section 5 concludes and points out some future works.

2. Related Work on Recommendation Systems

Recommendation systems are a type of information filtering which facilitate users to swiftly locate desirable items without being inundated by irrelevant information [5]. For building recommendation systems, there are three main approaches used in the literature. The content-based system, the collaborative system and the hybrid system. The content-based system works by considering the data of each questioned item in the form of a set of descriptors or terms also denoted with similar set of descriptors and built-up by assessing the data of objects seen by the user. The collaborative filtering on the other hand indicates that people may prefer their items based on similarity between items they like and also based on the preferences or choices made by other who has the similar kind of taste. And finally, we have the hybrid system which combines the content-based approach and the collaborative filtering approach [6].

There are numerous attempts that have been used in the field of recommendation systems and more specifically when it comes to recommendations systems in the nutrition field on which we applied.

Kundan S. Rana [7] proposed a recipe recommendation application to help the user finding their favourite food and its nutritional value using the content-based approach. The author used unit testing to validate the approach.

Freyne and Berkovsky [8] used a Content Based algorithm (CB) to predict the rating value for a target recipe based on the corresponding information of ingredients that are included in those recipes.

For the purpose of improving health conditions of users, Ueta et *al.* [9] proposed a goal-oriented recipe recommendation. The recipe recommendation provides a list of dishes that contains the right type of nutrient to treat user's health problems. The approach used co-occurrence databases and was validated using F-Measure on 1000 collected dishes. Since this approach uses databases, it highly relies on prior collected dataset. To solve the dataset requirement problem, Sivilai et *al.* [10] proposed an approach to showcase the development and design of a personalized nutrition and food planning system for older people that uses a food-oriented ontology together with an expert system.

In research related to dealing with malnutrition for the elderly, Aberg [11] proposes in a menu-planning tool. He proposed a hybrid approach by combining the Collaborative Filtering (CF), the Content Based (CB) and the constraint-based recommendation.

Although, there are some works done in the field of recommendation systems related to health, the above solutions are quite satisfying but relies on prior collected data. Below we have listed some few works using Fuzzy Logic which does not necessarily relies on beforehand data.

Restu Arif Priyono and Kridanto Surendro [12] proposed and approach based on Fuzzy Logic which focuses on diabetic patients by considering the following parameters: Blood sugar levels, The Physical Activity Level. The inference system used was the Takagi-Sugeno-Kang inference model to assess food calories from unlabelled food.

Moreover, Nikahat Mulla et al. [13] proposed a novel fuzzy set approach to generate an intelligent personalized diet plan. Fuzzy sets and the Minmax algorithm method were used for fuzzification purposes. The application generates the optimal calorie range intake based on the user's profile using Fuzzy Sets.

3. Review of the Chosen Algorithms

Our work is based on two main components, the Fuzzy logic algorithm and the k-Nearest Neighbour algorithm. In this section we are going to take a deeper look at these two algorithms.

3.1. Fuzzy Logic

Propounded in 1965 by Lotfi Aliasker Zadeh, fuzzy logic can be defined as an extension of the Boolean logic based on the mathematical theory of fuzzy sets [14].

It is an approach used to model the uncertainties and vagueness of human reasoning using fuzzy sets where the degree of truth varies between 0 and 1, with 0 meaning completely false and 1 completely true. In our proposed approach, the Fuzzy Logic component act as any fuzzy inference system, but in our case the used one is the Mamdani's fuzzy inference system. The general architecture of a Fuzzy Logic inference systems is mainly composed of the following parts [15]:
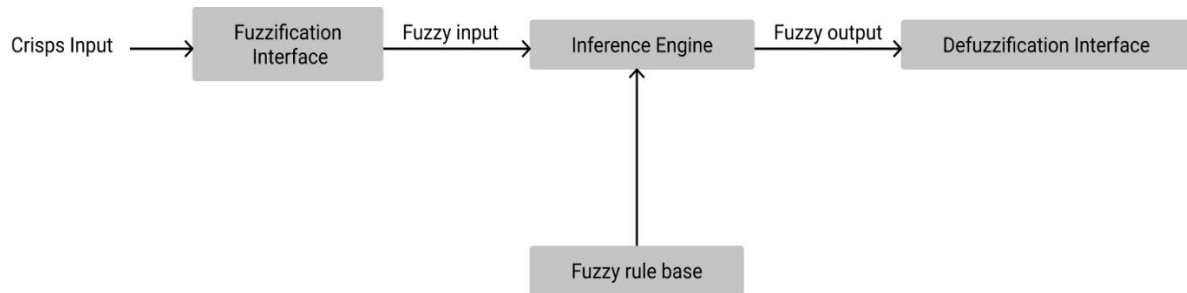
Fig.1. Fuzzy inference system

A. Fuzzification interface

Fuzzification is the process of decomposing a system input into one or more fuzzy sets with the use of membership functions such as triangular mf, trapezoidal mf etc. [16]. Here the information is converted into a form which the system can search for in its Knowledge Base.

B. Knowledge base or fuzzy rule base

The fuzzy rule base is the place where all the rules of the Fuzzy Logic set are stored and the rules are in the form of *if. . . then.* So, whenever the system tries to solve a new problem, the *if. . . then* ladder is executed and the system then works on the rule that it gets from the matched condition.

C. Inference Engine

The inference engine is responsible for drawing a valid result by analysing and concluding all the information that it gets from the fuzzification process. It can be viewed as the aggregation module of valid rules chosen. The final conclusion is then sent for further modification to the defuzzification component.

D. Defuzzification interface

Defuzzification is the inversion of fuzzification where the mapping is done to convert the crisp results into fuzzy results, but for defuzzification, the mapping is done to convert the fuzzy results into crisp results. This process is capable of generating a nonfuzzy control action which illustrates the possibility distribution of an inferred fuzzy control action [17]. There are numerous approaches of defuzzification: the center of area, the last of maxima, the mean of maxima etc. [18].

As example of a defuzzification technique we have the center of area which computes the position under the curve where the areas on both sides are equal. The center of area generates the action that partitions the area into two regions with the same surface.

Recommendation systems without any prior datasets or user's feedback and ratings is often subject to uncertainties or vagueness. Because when recommending a particular item, the formulation might be quite vague as recommending a dish rich in protein, where rich can't be quantified. Therefore, the only method that can address this level of vagueness is the Fuzzy Logic approach.

3.2. The k-nearest neighbour algorithm

The k-nearest neighbour algorithm is a non-parametric supervised machine learning algorithm used for both classification and regression. The input of the system consists of k closest training examples in a feature space and the output depends on the purpose of the algorithm.

The overall algorithm of the KNN algorithm is as follows [19]:

Algorithm: k-Nearest-Neighbor Algorithm
Input: Training dataset
Load Data
Initialize k to your chosen number of neighbors
for *example in the data* **do**
Calculate the distance between the query example and the current example from the data.
Add the distance and the index of the example to an ordered collection
end
Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
Pick the first k entries from the sorted collection
Get the labels of the selected k entries
If *regression, return the mean of the k labels*
If *classification, return the mode of the k labels*

The k-Nearest Neighbor Algorithm in our case has been used because of its simplicity and high performance when it comes to classification or regression. Thus, there are some other classification algorithms but the k-Nearest Neighbor Algorithm is one of the simplest and it outperforms the other existing ones in the literature.

4. Proposed Architecture, Results and Discussion

Given a recommendation task, the goal of a recommendation system is to find relevant items in a research space. Our proposed approach has two main components to solve this problem: A Fuzzy Logic component and a k-nearest neighbour algorithm.
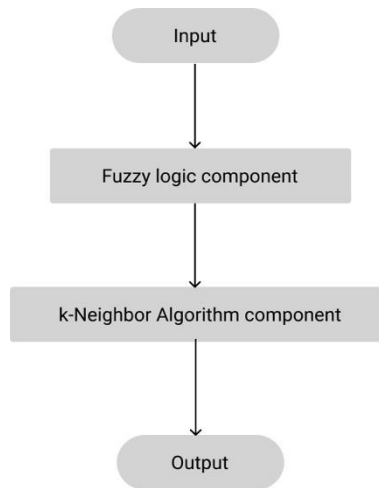

Fig.2. Overall architecture of the proposed approach

Fig.2. presents the overall architecture of the proposed approach. The architecture has 4 main components, the input of the system, the Fuzzy logic component, the k-Nearest Neighbor algorithm and finally the output of the system.

4.1. Input of the overall architecture

Since this paper is dealing with a recommendation problem, there is a need to consider the input of the proposed approach which can also be viewed as the filtering criterion or research criterion. In our case the input of the system represents the elements to be considered when performing the recommendation task. For instance, if the recommendation problem targets an item like a smartphone, the input can be the brand, the price, the size etc.

In our case we have considered the dish recommendation for people suffering from HIV/AIDS or Malaria. The following parameters have been taken using the nutrition guidelines provided by the World Health Organization [20, 21].

The *Age* in years: represents the age of a patient

- *The Body Mass Index (BMI)*: computed using the height and the weight of the patient following the formula

$$BMI(kg/m2) = \frac{Mass(kg)}{Height(m)^2} \tag{1}$$

- *Physical Activity Level:* range goes from *0* to *4* and represents the degree of physical activity per day (*Very light, Light, Heavy and Very heavy*)
- *Gender:* Is either male or female
- *Health status:* In our case it takes either HIV/AIDS or Malaria
- *Health state:* Symptomatic or Asymptomatic especially for People Living with HIV/AIDS (PLWHA).

4.2. Fuzzy logic component

The fuzzy inference system used here is the Mamdani's fuzzy inference because in our work the output of the system are fuzzy sets which needs defuzzification.

For the input of the inference system, the considered variables are: *Age*, *Gender*, *BMI*, *Health status*.

A. Inputs of the fuzzy system

- *Age:* The age has been divided in three categories *Young*, *Adult*, *Old*; *Young* is *in [0, 21]*, *Adult* is in *[18, 55]*, *Old* is in [50, ∞]. These values were taken according to [22, 23].

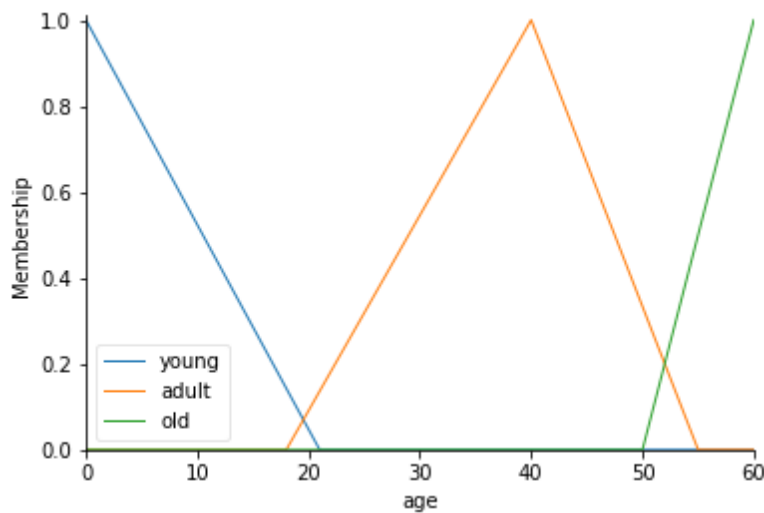The membership function is shown below.

Fig.3. Age membership function

- *BMI:* {*Underweight, Normal, Overweight, Obese 1, Obese 2, Obese 3*} which values vary between *0* to *40*; according to the World Health Organization [24].

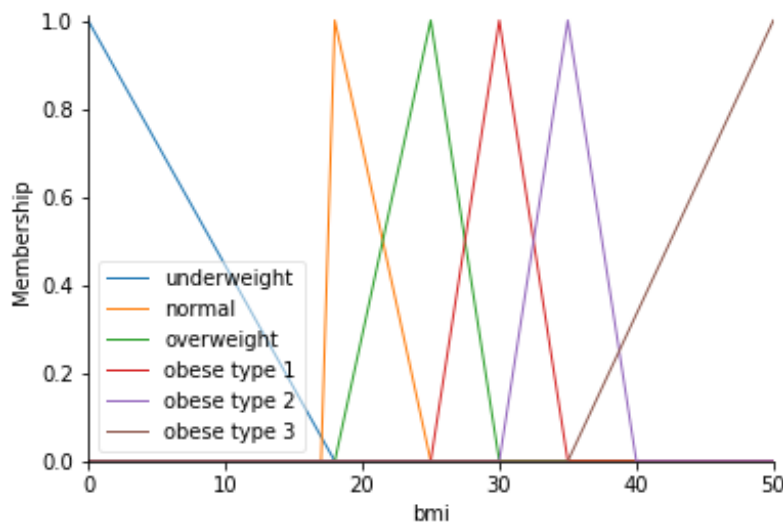The membership function is shown below.

Fig.4. BMI membership function

- *PAL*: {*Sedentary, light, Moderate, Heavy, Very Heavy*} ranges from *0* to *4* where *0* means sedentary and *4* very heavy.
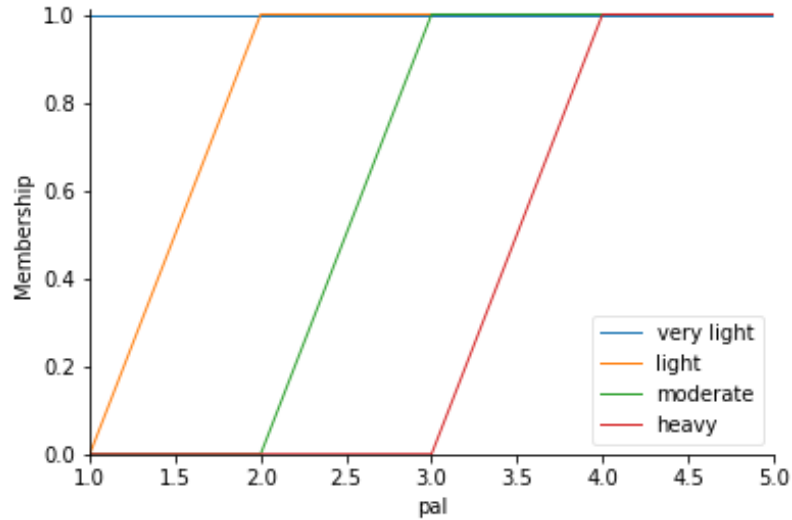
The membership function is shown below Figure 5.

Fig.5. Physical Activity Level membership function

As output of the Fuzzy Logic component, we got the following values.

- *Energy: {Low, Medium, High}* ranges from *0* to *600* [25].
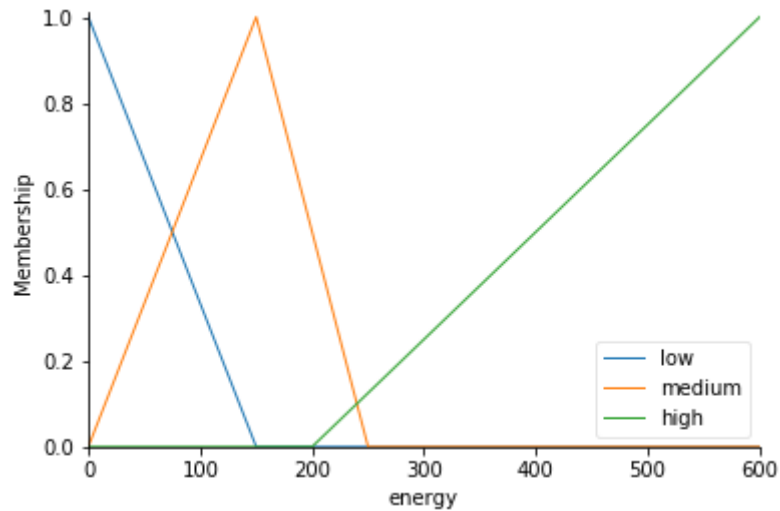
The membership function is shown below.

Fig.6. Energy membership function

- *Protein: {Low, Medium, High}* ranges from *0* to *40* [26].

The membership function is shown below.
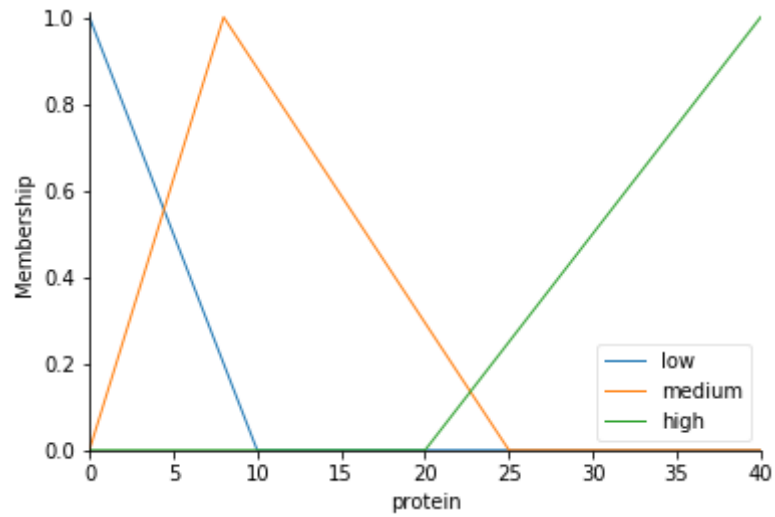
Fig.7. Protein membership function

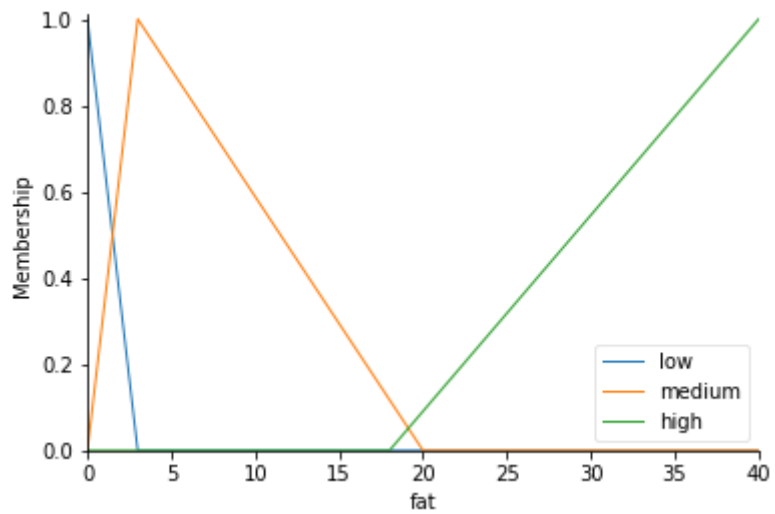- *Fat: {Low, Medium, High}* ranges from *0* to *40* [27].

Fig.8. Fat membership function

- *Carbs: {Low, Medium, High}* ranges from *0* to *50* [28].

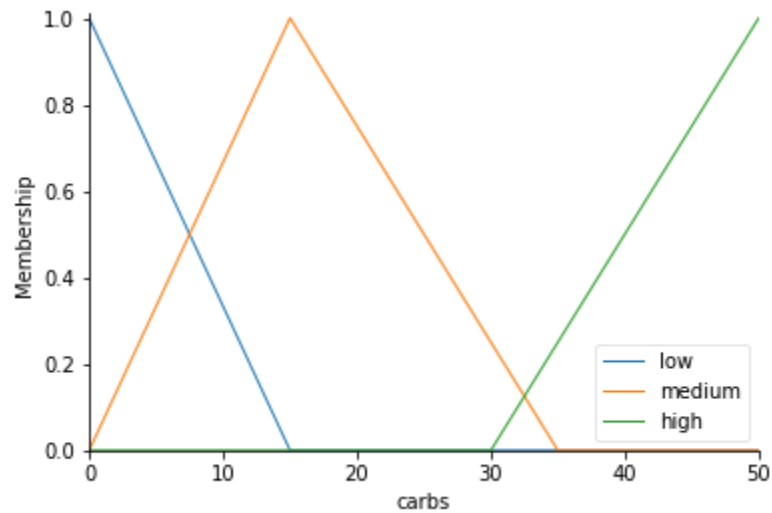The membership function is shown below.

Fig.9. Carbs membership function

- *Iron:* {*Low, Medium, High*} ranges from *0* to *10* [29].

The membership function is shown below.
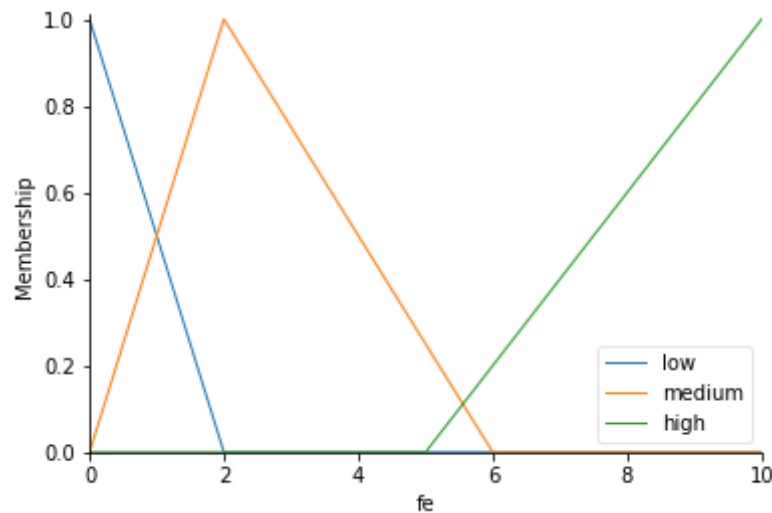
Fig.10. Iron Membership function

B. Fuzzy inference rules

To be able to work properly, a fuzzy inference system needs to implement some rules, which here are mainly extracted from WHO source. The table below shows a list of the rules used in our work [30, 31, 32].

- Examples of Rules for PLWHA

Since the rules declared in a Fuzzy Logic system uses the form *if antecedent then consequence,* the rule base presented below uses *BMI, Age, and Physical Activity Level* as antecedents and *Energy, Protein, Carbs, Fat, Iron* as consequences. As example we have the rule 1 which states that if the BMI of the person if Underweight then all the output (consequences) must be high in values.

No	Rules
1	rule1 = ctrl.Rule(bmi['underweight'], (energy['high'], protein['high'],carbs['high'] , fat['high'], fe['high']))
2	rule2 = ctrl.Rule(bmi['normal'], (energy['medium'], protein['high'],carbs['medium'] , fat['medium'], fe['medium']))
3	rule3 = ctrl.Rule(bmi['normal'], (energy['medium'], protein['high'],carbs['medium'] , fat['medium'], fe['medium']))
4	rule4 = ctrl.Rule(bmi['overweight'], (energy['medium'], protein['medium'], carbs['medium'] , fat['low'], fe['medium']))
5	rule5 = ctrl.Rule(bmi['obese type 1'], (energy['medium'], protein['medium'], carbs['low'] , fat['low'], fe['low']))
6	rule6 = ctrl.Rule(bmi['obese type 2'], (energy['low'], protein['low'], carbs['low'] , fat['low'], fe['low']))
7	rule7 = ctrl.Rule(bmi['obese type 3'], (energy['low'], protein['low'], carbs['low'] , fat['low'], fe['low']))
8	rule8 = ctrl.Rule(age['young'] & bmi['underweight'], (energy['high'], protein['high'],carbs['high'] , fat['high'], fe['high']))
9	rule9 = ctrl.Rule(age['young'] & bmi['normal'], (energy['high'], protein['high'],carbs['high'] , fat['medium'], fe['high']))
10	rule10 = ctrl.Rule(age['young'] & bmi['overweight'], (energy['medium'], protein['medium'],carbs['medium'] , fat['low'], fe['medium']))
11	rule11 = ctrl.Rule(age['young'] & bmi['obese type 1'], (energy['medium'], protein['medium'], carbs['low'] , fat['low'], fe['medium']))
12	rule12 = ctrl.Rule(age['young'] & bmi['obese type 2'], (energy['low'], protein['medium'], carbs['low'] , fat['low'], fe['medium']))
13	rule13 = ctrl.Rule(age['young'] & bmi['obese type 3'], (energy['low'], protein['low'],carbs['low'] , fat['low'], fe['medium']))
14	rule14 = ctrl.Rule(age['adult'] & bmi['underweight'], (energy['high'], protein['high'],carbs['high'] , fat['high'], fe['high']))
15	rule15 = ctrl.Rule(age['adult'] & bmi['normal'], (energy['high'], protein['high'],carbs['high'] , fat['medium'], fe['high']))
16	rule16 = ctrl.Rule(age['adult'] & bmi['overweight'], (energy['medium'], protein['medium'],carbs['medium'] , fat['low'], fe['medium']))
17	rule17 = ctrl.Rule(age['adult'] & bmi['obese type 1'], (energy['low'], protein['medium'], carbs['low'] , fat['low'], fe['medium']))
18	rule18 = ctrl.Rule(age['adult'] & bmi['obese type 2'], (energy['low'], protein['medium'], carbs['low'] , fat['low'], fe['medium']))
19	rule19 = ctrl.Rule(age['young'] & bmi['obese type 3'], (energy['low'], protein['low'],carbs['low'] , fat['low'], fe['medium']))
20	rule20 = ctrl.Rule(age['adult'] & bmi['underweight'], (energy['high'], protein['high'],carbs['low'] , fat['low'], fe['high']))
21	rule21 = ctrl.Rule(age['adult'] & bmi['normal'], (energy['medium'], protein['medium'],carbs['low'] , fat['low'], fe['high']))

Fig.11. Screenshot of some rules for people living with HIV/AIDS

- Examples of rules for people suffering from Malaria

In this case the recommendation relies heavily on the amount of *Iron* to be recommended.
The rules below shows that all the output of the Iron consequence is high because Malaria causes Iron deficiency.

No	Rules
1	rule1 = ctrl.Rule(bmi['underweight'], (energy['high'], protein['high'],carbs['high'] , fat['high'], fe['high']))
2	rule2 = ctrl.Rule(bmi['normal'], (energy['medium'], protein['high'],carbs['medium'] , fat['medium'], fe['high']))
3	rule3 = ctrl.Rule(bmi['normal'], (energy['medium'], protein['high'],carbs['medium'] , fat['medium'], fe['high']))
4	rule4 = ctrl.Rule(bmi['overweight'], (energy['medium'], protein['medium'], carbs['medium'] , fat['low'], fe['high']))
5	rule5 = ctrl.Rule(bmi['obese type 1'], (energy['medium'], protein['medium'], carbs['low'] , fat['low'], fe['high']))
6	rule6 = ctrl.Rule(bmi['obese type 2'], (energy['low'], protein['low'], carbs['low'] , fat['low'], fe['high']))
7	rule7 = ctrl.Rule(bmi['obese type 3'], (energy['low'], protein['low'], carbs['low'] , fat['low'], fe['high']))
8	rule8 = ctrl.Rule(age['young'] & bmi['underweight'], (energy['high'], protein['high'],carbs['high'] , fat['high'], fe['high']))
9	rule9 = ctrl.Rule(age['young'] & bmi['normal'], (energy['high'], protein['high'],carbs['high'] , fat['medium'], fe['high']))
10	rule10 = ctrl.Rule(age['young'] & bmi['overweight'], (energy['medium'], protein['medium'],carbs['medium'] , fat['low'], fe['high']))
11	rule11 = ctrl.Rule(age['young'] & bmi['obese type 1'], (energy['medium'], protein['medium'], carbs['low'] , fat['low'], fe['high']))
12	rule12 = ctrl.Rule(age['young'] & bmi['obese type 2'], (energy['low'], protein['medium'], carbs['low'] , fat['low'], fe['high']))
13	rule13 = ctrl.Rule(age['young'] & bmi['obese type 3'], (energy['low'], protein['low'],carbs['low'] , fat['low'], fe['high']))
14	rule14 = ctrl.Rule(age['adult'] & bmi['underweight'], (energy['high'], protein['high'],carbs['high'] , fat['high'], fe['high']))
15	rule15 = ctrl.Rule(age['adult'] & bmi['normal'], (energy['high'], protein['high'],carbs['high'] , fat['medium'], fe['high']))
16	rule16 = ctrl.Rule(age['adult'] & bmi['overweight'], (energy['medium'], protein['medium'],carbs['medium'] , fat['low'], fe['high']))
17	rule17 = ctrl.Rule(age['adult'] & bmi['obese type 1'], (energy['low'], protein['medium'], carbs['low'] , fat['low'], fe['high']))
18	rule18 = ctrl.Rule(age['adult'] & bmi['obese type 2'], (energy['low'], protein['medium'], carbs['low'] , fat['low'], fe['high']))
19	rule19 = ctrl.Rule(age['young'] & bmi['obese type 3'], (energy['low'], protein['low'],carbs['low'] , fat['low'], fe['high']))
20	rule20 = ctrl.Rule(age['adult'] & bmi['underweight'], (energy['high'], protein['high'],carbs['low'] , fat['low'], fe['high']))
21	rule21 = ctrl.Rule(age['adult'] & bmi['normal'], (energy['medium'], protein['medium'],carbs['low'] , fat['low'], fe['high']))

Fig.12. Screenshot of some rules for people living with Malaria

C. Defuzzification

Once the user or patient enters some personal information, the system picks a rule corresponding to this information and then apply it, but the result of these rules is in linguistic terms so we need to get back a crisp value which is done by the defuzzification module. The method used in our approach is the centroid method. In this component the system outputs a list of crisps values corresponding to the amount of *Energy* in kcal, *Carbs* in grams, *Iron* in mg, *Protein* in grams. *Fats* in grams which are necessary for the dish retrieval process.

4.3. K-Nearest Neighbour computation algorithm

The k-Nearest Neighbor algorithm is a non-parametric method used for classification and regression purposes by

computing the similarity between points, but in our case instead of using this algorithm for classification or regression, we used it here to retrieve the closest dishes to the values obtained by the Fuzzy Logic component.

The KNN component take as input the following values derived from the Fuzzy Logic component:

- *Energy* in *kcal*
- *Carbs* or *carbohydrates* in *g*
- *Iron* in *mg*
- *Protein* in *g*
- *Fats* in *g*

The dataset used in this work [33] was mainly extracted from a study that reviews published data and contributes to the knowledge of the ingredients and nutrients of Cameroon traditional dishes. Macronutrient energy, *carbohydrates*, *protein*, t*otal fat*, *fiber* and *ash*) and mineral (*iron, zinc, magnesium, calcium, phosphorus, copper, manganese, potassium, sodium and selenium*) data are presented for *117* commonly consumed dishes from three ecoregions. The dataset gave us a list of most consumed dishes and their nutrients value for *100g* of each dish.

Below are listed and presented the different columns that we used in our work:

- *Usual names:* describing the name of a particular dish in the local tongue
- *Type:* presenting the type of the dish, m for meal, s for snack, and n for . . .
- *Main Ingredients:* showing a list of main ingredients used during the dish preparation
- *Energy in kcal:* the amount of energy contained in in *100g* of a dish. The energy varies from *12* to *403*.
- *Water(g): t*he amount, in gram, of water contained in the dish which varies from *29.8* to *95g*.
- *Protein x 6.25 (g):* varies from *0.10g* to *20.90g* present the amount of protein contained in the dish
- *Total fat (g):* ranging from *0.05000 g* to *35g*, present the amount of fat contained in the dish.
- *Available carbohydrates (g):* shows the amount of carbohydrates which varies from *0.40* to *40.32g*.
- *Fe(mg):* 0 to *7.29g* the amount of iron contained in the dishes

	0	Usual names	Type	Main Ingredients	Energy (kcal)	Water(g)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Crude fibre (g)	...	Fe (mg)	Zn (mg)	Mg (mg)	Ca (mg)	P (mg)	Cu (mg)	Mn (mg)	K (mg)	Na (mg)	Se (mg)
0	1	Banane-malaxée	m	Banane, arachide, poisson, ail, herbes vertes,...	84	83.0	1.93	4.92	7.16	1.66	...	0.78	2.05	21.0	49.0	49.0	0.047	0.455	196.0	NaN	NaN
1	2	Ragoût de Haricot	s	Haricot rouge, tomate, écrevisse, oignon, huil...	240	71.9	0.20	26.60	0.90	0.20	...	0.10	0.10	2.0	5.0	5.0	NaN	0.000	31.0	5.0	0.2
2	3	Bouillie-maïs	m, sn	Farine de maïs, arachide grillée, sucre, extra...	64	84.5	0.95	0.44	13.97	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	4	Bouillie-infantile 1	m	Farine de maïs, sucre, lait, arachide, jaune d...	51	86.9	0.77	0.38	11.86	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	5	Bouillie infantile 2	m	Pâte de maïs, sucre, lait, arachide, jaune d'œuf	48	87.5	0.70	0.30	11.40	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fig.13. Screenshot of the dish dataset.

A. Dataset preprocessing

In this section we pass our dataset through a preprocessing step. Fig.14 presents the heatmap of our dish database. The abscissa represents the nutrients and the ordinate represents number of entries automatically generated by the *Matplotlib* library. The purpose (heatmap) is to get an insight of the null (*NaN*) values contained in the dataset. So, in our case, the colored bars represent the null values for each column. However, in order to apply the KNN algorithm on our dataset, we didn't use all the columns on the datasets because some columns had too many null values and could not be used for the distance metric computation.
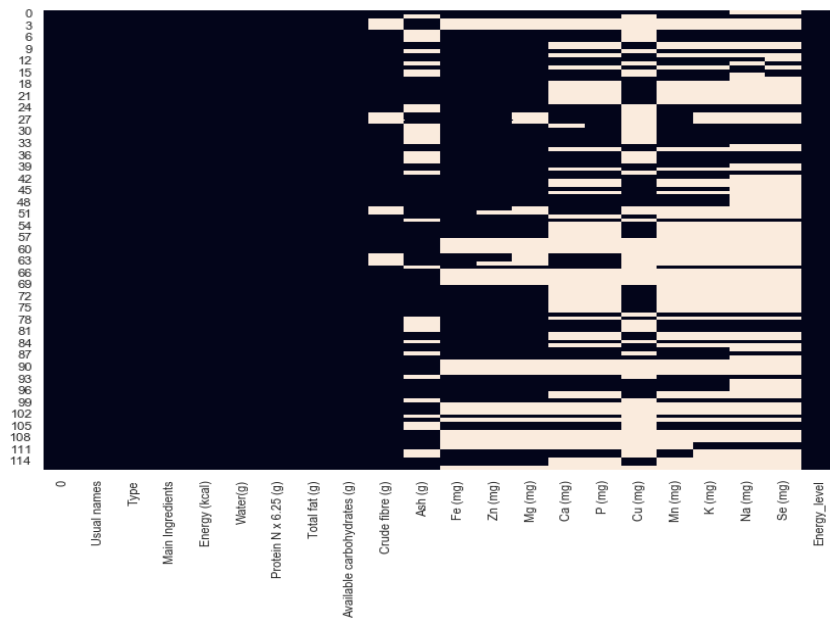
Fig.14. Heatmap of the dataset on using Matplotlib

To handle the null values, we deleted the columns which contained the *NaN* values except for the Iron (*Fe*) which was really important because of the malaria disease. So, the *NaN* for the *Fe* columns were fill with zeros.

B. Outputs of the KNN algorithm

To retrieve the dishes, a Euclidean distance measure is computed between the outputted values from the Fuzzy Logic component and the preprocessed dataset.

After this step, a list of dishes is sorted by the minimum distance to the maximum one. In our approach we used a *k* value of 4 to let the user have a wide range of choices concerning the dishes.

In summary, we presented our approach which consisted in two main components: The Fuzzy Logic component and the k-nearest neighbor. Our approach used Fuzzy Logic to recommend or compute the values of *Energy, Protein, Carbs, Fats, Iron* then the k-nearest neighbor used these values to retrieve the dishes from the preprocessed dataset.

In this paper, we presented our architecture combining Fuzzy Logic and K-Nearest Neighbor, to obtain any results, the whole approach has been used to recommend dishes for people living with HIV/AIDS or Malaria. As example when any new recommendation task -for a person living with HIV/AIDS- comes in the system, it goes through each layer till the system output a result. The user first provides personal information such as Gender, Age, Weight, Height, his Physical activity level. These values are fed to the Fuzzy Logic component which applies our predefined rules then output a list of nutrients values to be consumed by the user, the output is a list of numerical values. Then the obtained numerical values are used by the k-Nearest Neighbor Algorithm to retrieve the dishes which nutrients values are closest to the output of the Fuzzy Logic component. In the next section, we present results for some of the real use cases that we used.

4.4. Dishes suggested to people suffering from HIV/AIDS

The test case in this section deals with a young *male* person which weight is *80kg*, height is *150cm*, the state of the disease is Asymptomatic and his physical activity level is Very Light meaning this person doesn't exercise that much.

Test case

- *Age = 15*
- *Weight = 80*
- *Height = 150*
- *Physical Activity Level = Very Light*
- *State = Asymptomatic*
- *Gender = Male*

The output is

```
1  #input_age, input_height, input_weight, input_pal, input_state = "SYMPTOMATIC", input_gender = "MALE", input_disease
2
3  M1 = FuzzyRecommendationModel(15, 150, 80, 'Very Light', "ASYMPTOMATIC", "MALE", "HIV")
4
5  M1.daily_energy_needs()
6  M1.fuzzy_model()
7
8
```

```
Daily Energy needs:
 2384
input bmi is:  35.55555555555556
Nutrients values from fuzzy logic model :
 [50.55555555555559, 8.184828105408217, 1.011111111111111, 5.0555555555555545, 2.005349846390585]
Recommended dish list:  Eru/fufu
Recommended dish list:  Bouillie-infantile 1
Recommended dish list:  Bouillie infantile 2
Recommended dish list:  Bouillie-maïs
```

Fig.15. Output of a test case for a HIV/AIDS patient

Fig.15. represents the output of the proposed system in the *jupyter notebook* environment for the test case of a HIV/AIDS patient. Here we can notice that the system outputs 04 main elements. The daily energy needs in kcal, the Body Mass Index (BMI), the nutrients values in a vector, and finally a list of potential dishes that can be consumed.

Fig.16. presents the nutrients values of the retrieved dishes. The main purpose of this figure is to show the amount of nutrients of the recommended dish list which we can notice it's quite close to the outputted vector from Fig.15.

In [43]: `1 df["Usual names"] == "Eru/fufu"][["Usual names", "Type", "Main Ingredients", "Energy (kcal)", "Protein N x 6.25 (g)", "Total`

Out[43]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
17	Eru/fufu	m	Pâte de manioc fermentée, feuille de Gnetum af...	46	2.01	1.81	4.32	1.2

In [22]: `1 ", "Main Ingredients", "Energy (kcal)", "Protein N x 6.25 (g)", "Total fat (g)", "Available carbohydrates (g)", "Fe (mg)"] ]`

Out[22]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
3	Bouillie-infantile 1	m	Farine de maïs, sucre, lait, arachide, jaune d...	51	0.77	0.38	11.86	NaN

In [24]: `1 df.loc[df["Usual names"] == "Bouillie infantile 2"][["Usual names", "Type", "Main Ingredients", "Energy (kcal)", "Protein N`

Out[24]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
4	Bouillie infantile 2	m	Pâte de maïs, sucre, lait, arachide, jaune d'œuf	48	0.7	0.3	11.4	NaN

In [28]: `1 df.loc[df["Usual names"] == "Bouillie-maïs"][["Usual names", "Type", "Main Ingredients", "Energy (kcal)", "Protein N x 6.25`

Out[28]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
2	Bouillie-maïs	m, sn	Farine de maïs, arachide grillée, sucre, extra...	64	0.95	0.44	13.97	NaN

Fig.16. Nutrients values of the retrieved dishes.

From the above output (Fig.15.), the fuzzy logic model gives some approximate values of the nutrients values which may be contained in the consumed dishes and from these values we get a list of dishes ranging from the closest to the farthest of these nutrients' values. In our case the patient should consider the consumption of firstly the *Eru/fufu*, then *Bouillie infantile 1*, *Bouillie infantile 2* and finally *Bouillie-mais*. Here the dishes are not too rich in kcal because of the B*ody Mass Index* which is 35; so, the person is obese and must avoid dishes too rich in *fat* and *energy*.

4.5. Dishes proposed to People suffering from Malaria

For People suffering from Malaria in this section, we took people suffering from Malaria and we also took a look at one test case as shown below. The nutrition recommendation here is mainly related to the amount of iron contained in the dishes. For the test purpose we took a person whose personal information are listed below:

Test case

- *Age = 22*
- *W eight = 70*
- *Height = 168*
- *Physical Activity Level = Heavy*
- *Gender = Female*

The output is:

```
1  #input_age, input_height, input_weight, input_pal, input_state = "SYMPTOMATIC", input_gender = "MALE", input_disease
2
3  M1 = FuzzyRecommendationModel(22, 168, 70, 'Heavy', "", "FEMALE", "MALARIA")
4
5  M1.daily_energy_needs()
6  M1.fuzzy_model()
7
8
```

```
Daily Energy needs:
 2958.0
input bmi is:  24.801587301587304
Nutrients values from fuzzy logic model :
 [154.80507159037091, 11.598472466021384, 3.4927897180018053, 17.06608909366467, 6.802534028962517]
Recommended dish list:  Kon
Recommended dish list:  Corn chaff
Recommended dish list:  Site-sembun-bo-kenà
Recommended dish list:  Mbedi 2
```

Fig.17. Output of a test case for a HIV/AIDS patient

Fig.17. represents the output of the proposed system in the *Jupyter notebook* environment for the test case of a malaria patient. We can also denote here that the system outputs also 04 main elements. The daily energy needs in kcal, the Body Mass Index (BMI), the nutrients values in a vector, and finally a list of potential dishes that can be consumed.

Fig.18. represents the nutrients values of the retrieved dishes. As explained in the previous test case, we have the amount of nutrients of the recommended dish list presented in Fig.18.

In [34]: `1 df.loc[df["Usual names"] == "Kon"][["Usual names", "Type", "Main Ingredients", "Energy (kcal)", "Protein N x 6.25 (g)", "Tot`

Out[34]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
40	Kon	s	Haricot rouge, huile de palme, tomate, sel, po...	161	7.38	7.05	15.15	2.24

In [35]: `1 df.loc[df["Usual names"] == "Corn chaff"][["Usual names", "Type", "Main Ingredients", "Energy (kcal)", "Protein N x 6.25 (g)`

Out[35]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
9	Corn chaff	m	Grains de maïs, haricot rouge, huile de palme,...	149	3.76	5.65	20.29	1.95

In [36]: `1 df.loc[df["Usual names"] == "Site-sembun-bo-kenà"][["Usual names", "Type", "Main Ingredients", "Energy (kcal)", "Protein N x`

Out[36]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
102	Site-sembun-bo-kenä	m	Arachide, tubercule de manioc, huile de palme,...	150	2.59	6.86	19.04	NaN

In [37]: `1 df.loc[df["Usual names"] == "Mbedi 2"][["Usual names", "Type", "Main Ingredients", "Energy (kcal)", "Protein N x 6.25 (g)",`

Out[37]:

	Usual names	Type	Main Ingredients	Energy (kcal)	Protein N x 6.25 (g)	Total fat (g)	Available carbohydrates (g)	Fe (mg)
51	Mbedi 2	m	Feuille de manioc, tubercule de manioc, arachi...	159	2.92	6.67	21.87	1.9

Fig.18. Nutrients values of the retrieved dishes for Malaria recommendation

For the patient suffering from Malaria, the aliments that should be considered for consumption are *Kon*, *Corn chaff*, *Site − sembun − bo − kenà* and *Mbedi 2*. We notice that for a person suffering from malaria and whose *BMI* is about 25 (*normal*) the values obtained here are mainly related to the last element (*Iron*) of the nutrient values which is about 7mg, so the dishes obtained are richer in iron. This result clearly demonstrate that this particular patient should consumed dishes that are richer in iron.

Since the main objective of this work was to propose an approach able to perform recommendation task without any training sets, in this work we used the combination of Fuzzy Logic and the k-Nearest Neighbor Algorithm to tackle the problem. The results demonstrated how effective the proposed approach is because we applied the approach to a dish recommendation problem for people suffering from Malaria or HIV/AIDS and the results were quite satisfying and can be used in a real-world scenario.

The proposed approach should be taken as a framework where the fuzzification methods, the inference engine, the distance measure and other parameters can be changed to solve any given recommendation problem.

5. Conclusion

In this work we proposed a novel approach of recommendation based on the combination of Fuzzy Logic and the k-nearest neighbour Algorithm. This approach provides a new alternative to develop recommendation systems because unlike the available ones in the literature, it does not need any prior collected data and still performs good recommendations. We applied this approach to address the problem of nutrition for people suffering from HIV/AIDS and Malaria. Until now in order to build recommendation systems, we usually rely on huge datasets of user's interactions with the item to be recommended. The available approaches used Machine Learning models to train the recommendation system, but in our case, there is no need to collect huge datasets or perform some model training. The proposed approach can be applied in any field and any recommendation problem through a good formulation. First using the Fuzzy Logic to infer an output using a predefined rule base then use the output to retrieve the item to be recommended in the whole search space.

Future Work

An interesting lead for future research is the use of this approach to collect user's data and form a dataset to explore other machine learning algorithms. Furthermore, since this approach was used to address the HIV/AIDS and Malaria diseases, we would eventually extend it to other diseases and even other domains.

References

[1] Y. O. Isinkaye, Folajimi, B.A. Ojokoh, "Recommendation systems: Principles, methods and evaluation." In: Egyptian Informatics Journal16.3 Nov.2015, pp. 261–273, Doi: 0.1016/j.eij.2015.06.005.

[2] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, "Recommender system application developments: A survey." In: Decision Support Systems, 2015, 74, pp. 12–32, Doi: 10.1016/j.dss.2015.03.008.

[3] T. N. T. Tran, M. Atas, A. Felfernig, M. Stettinger, "An overview of recommender systems in the healthy food domain." In Intell Inf Syst50.3 June 1, 2018, pp. 501–526, doi:10.1007/s10844- 017- 0469-0.

[4] V. B. S. Prasath, H. A.A. Alfeilat, A.B.A. Hassanat, O. Lasassmeh, A.S. Tarawneh, M.B. Alhasanat, H.S.E. Salman, "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier – A Review." In: Big Data7.4 Dec. 1, 2019, pp. 221–248, Doi:10.1089/big.2018.0175. arXiv: 1708.04321.

[5] Saudagar L. Jadhav, Manisha P. Mali, "Pre-Recommendation Clustering and Review Based Approach for Collaborative Filtering Based Movie Recommendation", International Journal of Information Technology and Computer Science (IJITCS), Vol.8, No.7, pp.72-80, 2016. DOI: 10.5815/ijitcs.2016.07.10

[6] Raghavendra C K, Srikantaiah K.C, Venugopal K. R, " Personalized Recommendation Systems (PRES): A Comprehensive Study and Research Issues", International Journal of Modern Education and Computer Science(IJMECS), Vol.10, No.10, pp. 11-21, 2018.DOI: 10.5815/ijmecs.2018.10.02

[7] K. S. Rana, "Food Recommendation System based on Content filtering Algorithm". Bachelor's Degree in Computer Science. Tribhuwan University, August 2016.

[8] J. Freyne, S. Berkovsky, "Recommending Food: Reasoning on Recipes and Ingredients." In: User Modeling, Adaptation, and Personalization. Ed. by Paul De Bra, Alfred Kobsa, and David Chin. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2010, pp. 381–386, Doi:10.1007/978-3-642-13470-8_36.

[9] T. Ueta, M. Iwakami, T. Ito, "Implementation of a Goal-Oriented Recipe Recommendation System Providing Nutrition Information." In: International Conference on Technologies and Application of Artificial Intelligence Nov. 1, 2011, Doi:10.1109/TAAI.2011.39.

[10] S. Sivilai, C. Snae, M. Brückner, "Ontology-Driven Personalized Food and Nutrition Planning System for the Elderly." In: the 2ndInternational Conference in Business Management and Information Sciences 19-20 January 2012 at: Phitsanulok, Thailand. Jan. 19, 2012.

[11] J. Aberg, "Dealing with Malnutrition: A Meal Planning System for Elderly." In: Argumentation for Consumers of Healthcare, Papers from the 2006AAAI Spring Symposium, Technical Report SS-06-01, Stanford, California, USA, March 27-29, 2006. Jan. 1, 2006, pp. 1–7.

[12] R. A Priyono, K. Surendro, "Nutritional Needs Recommendation based on Fuzzy Logic". In: Procedia Technology. 4th International Conference on Electrical Engineering and Informatics, ICEEI 2013 11 Jan. 1,2013, pp. 1244–1251, Doi: 10.1016/j.protcy.2013.12.320.

[13] N. Mulla, S: Kurhade, N. Bakereywala, "An Intelligent Application for healthcare Recommendation using Fuzzy Logic." In: 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA). June 1, 2019, pp. 466–472. Doi:10.1109/ICECA.2019.8821959

[14] F. Dernoncourt, "Introduction to fuzzy logic." In: Massachusetts Institute of Technology Fuzzy Logic Lecture, 01 jan 2013 (Jan. 1, 2013).

[15] M. Sharma, Fuzzy Logic System Architecture in Artificial Intelligence. URL: https://www.includehelp.com/ml-ai/fuzzy-logic-system-architecture-in-artificial-intelligence.aspx (visited on 12/05/2020).

[16] Z. D. Xu, Y.Q Guo, J.T. Zhu, F.H. Xu, "Fuzzification - an overview". Intelligent Vibration Control in Civil Engineering Structures. Academic Press. 2017, p. 21-67.

[17] GeeksforGeeks. Difference between Fuzzification and Defuzzification. Section: Difference Between. Aug. 28, 2019. URL: https://www.geeksforgeeks.org/difference-between-fuzzification-and-defuzzification/ (visited on 12/06/2020)

[18] D. Samanta, Indian Institute of Technology Kharagpur. "Defuzzification Methods". URL: https://cse.iitkgp.ac.in/~dsamanta/courses/archive/sca/Archives/Chapter%205%20Defuzzification%20Methods.pdf (visited on 12/25/2020).

[19] O. Harrison, Towards Data Science. "Machine Learning Basics with the K-Nearest Neighbors Algorithm". Medium. July 14, 2019. URL: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761 (visited on 01/10/2021).

[20] World Health Organization. "Nutrient requirements for people living with HIV/AIDS". Report of a technical consultation. Geneva, Switzerland. 2003. ISBN 9241591196.

[21] World Health Organization. "HIV/AIDS: A Guide For Nutrition Care and Support". Food and Nutrition Technical Assistance (FANTA) Project. 2001.

[22] World Health Organization. Adolescent health. URL: https://www.who.int/westernpacific/health-topics/adolescent-health (visited on 01/11/2021)

[23] Wikipedia. "Old age". Page Version ID: 998897633. Jan. 7, 2021. URL: https://en.wikipedia.org/w/index.php?title=Old_age&oldid=998897633 (visited on 01/11/2021).

[24] World Health Organization. "Body mass index – BMI". URL: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi (visited on 11/13/2020).

[25] D. Whitbread, "Top 10 Foods Highest in Calories". USDA Nutrition Data. URL: https://www.myfooddata.com/articles/highest-calorie-foods.php (visited on 01/11/2021).

[26] D. Whitbread, 2Top 10 Foods Highest in Protein". USDA Nutrition Data. url: https://www.myfooddata.com/articles/foods-highest-in-protein.php (visited on 01/11/2021).

[27] Heart UK. "Saturated fat". URL: https://www.heartuk.org.uk/low- cholesterol-foods/saturated-fat (visited on 01/11/2021).

[28] Whitbread, D. "Top 10 Foods Highest in Iron". USDA Nutrition Data. Top 10 Foods Highest in Carbohydrates (To Limit or Avoid). URL: https://www.myfooddata.com/articles/foods-highest-in-carbohydrates.php (visited on 01/11/2021).

[29] Whitbread, D. "Top 10 Foods Highest in Iron". USDA Nutrition Data. URL: https://www.myfooddata.com/articles/food-sources-of-iron.php (visited on 01/11/2021).

[30] T. H. Harvard, Chan, "School of public health". Food and Diet. Obesity Prevention Source. Oct. 21, 2012. URL: https://www.hsph.harvard.edu/obesity-prevention-source/obesity-causes/diet-and-weight/(visited on 01/11/2021)

[31] B.A. Swinburn, I. Caterson, J.C. Seidell,"Diet, nutrition and the prevention of excess weight gain and obesity." In: Public Health Nutr.7.1 Feb. 2004, pp. 123–146, Doi:10.1079/PHN2003585.

[32] A. M. Prentice, H. Ghattas, C. Doherty, S.E. Cox, "Iron Metabolism and Malaria." In: Food Nutr Bull 28.4 Dec. 1, 2007. Publisher: SAGE Publications Inc, Doi:10.1177/15648265070284S406.

[33] C.P. Kouebou, M. Achu, S. Nzali, M. Chelea, J. Bonglaisin, A. Kamda, P. Djiele, G. Yadang, R. Ponka, G. Ngoh Newilah, G. Nkouam, C.Teugwa, M. M. Kana Sop, "A review of composition studies of Cameroon traditional dishes: Macronutrients and minerals." In: Food Chemistry. 9th International Food Data Conference: Food Composition and Sustainable Diets 140.3 Oct. 1, 2013, pp. 483–494, Doi:10.1016/j.foodchem.2013.01.003.

Authors' Profiles

Paul Dayang received the Ph.D. degree in computer science in 2014 from the University of Bremen, Germany. He is a Senior Lecturer with the University of Ngaoundéré, Cameroon. His research interests include the area of information systems, intelligent systems, and information retrieval techniques for natural language processing with focus on resource-scarce African languages.

Cyrille Sepele Petsou received a Master degree in Software and Systems in Distributed Environment from the University of Ngaoundéré, Cameroon in 2021. His research interest includes fuzzy logic, k-nearest neigbor algorithms.

Damien Wohwe Sambo received the B.Sc. degree in mathematics and computer science and the M.Sc. degree in computer engineering from the Faculty of Science, University of Ngaoundéré, Cameroon, in 2012 and 2016, respectively. He is enrolled in the co-directed Ph.D. thesis among the University of Ngaoundéré and the University of Bremen, Germany. In 2019, he made a Ph.D. Internship with University Cheikh Anta Diop, Dakar, Senegal. His research interest concerns the energy efficiency and reliability of the Internet of Things applications, and especially for wireless underground sensor networks.