

Using KNN Method for Educational and Vocational Guidance

Essaid EL HAJI

LIST lab, Abdelmalek Essaadi
University
Faculty of Science and
Technology, Tangier, Morocco
BP: 416, Ziaten, Tanger –
Maroc

Abdellah AZMANI

LIST lab, Abdelmalek Essaadi
University
Faculty of Science and
Technology, Tangier, Morocco
BP: 416, Ziaten, Tanger –
Maroc

Mohamed EL HARZLI

LIST lab, Abdelmalek Essaadi
University
Faculty of Science and
Technology, Tangier, Morocco
BP: 416, Ziaten, Tanger –
Maroc

ABSTRACT

This paper presents a decision support tool for educational and vocational guidance, based on the supervised classification method k-nearest neighbors (KNN). This method consists in determining, for each new observation to be classified, the list of nearest neighbors of the observations already classified. The use of the KNN method requires choosing a distance and the most classical one is the Euclidean distance. In the context of this work, two functions were tested to measure resemblance as far as similarity and dissimilarity are concerned.

General Terms

Educational and vocational guidance, k-nearest neighbors, similarity, dissimilarity

Keywords

Educational and vocational guidance, RIASEC, pairing, k-nearest neighbors, similarity, dissimilarity.

1. INTRODUCTION

The question of the educational and vocational guidance represents a major importance in any educational system, since of the latter depend, in a great extent, the effectiveness and the efficiency of schools. In fact, a bad educational or professional guidance may be at the origin of several educational and social problems: school failure, school dropout, lack of skills, difficulties of integration, unemployment, etc. Therefore, choosing a job is a major issue for any person in the sense that it determines his professional future. Such a choice has become a difficult business today. In fact, in a socio-economic context marked by great changes and especially fierce competition with regard to job (unemployment rate, plenty of graduates, ...), making bad choices should be avoided. This choice is often based on several criteria or parameters or factors whose weights (or coefficients) are of different natures. In fact, several studies have examined demographic and personal factors that can influence the vocational and educational choice. Among personal factors, general aptitude level [1], cultural capital or status culture [2], values and principles [3], self-esteem [4] self-efficacy [5], interests [5], and personality [6] are determining factors in students' choice of vocational and educational guidance. Among these demographic factors are parents' occupation [7], parents' educational level [8], ethnic origins [2], as well as students' socio-economic status, gender and age [1].

This undeniable importance of guidance has prompted specialists, each on his side, to give it a central place in the educational system resulting in a multitude of theoretical approaches and many computer systems.

With regard to IT interventions, they have experienced a great growth since their introduction in the mid-1960s due to the development of information technology and communication (ICT) [9], for their many benefits: reduction of costs of psychological consultations, great ease in the updating and revision of data, improvement of the interactive dialogue, large process control, high speed, precision, availability of results, increase in the motivation and autonomy of the user, and ease at counting and analyzing tests' scores, Individualization of labor which allows everyone to progress at one's own pace [10] [11] [12].

However, this progress has not led to implement a fairly complete system which, both, considers all the factors involved in the guidance process and is based on a solid theoretical and technological base.

In this paper, the classification method k-nearest neighbors (KNN) is used to determine to which class of trades belongs an individual according to his profile.

2. KNN AND wKNN

2.1.K-Nearest-Neighbor Techniques (KNN)

KNN is a basic method of classification, it's nonparametric, simple and direct, and it is applied to the bases of the labeled data and does not require learning but simply storing the training data. Its principle is the following: In the context of classification of a new observation x , the principle of KNN is to find the nearest neighbors of this observation. The x class is determined by the majority classes among the k-nearest neighbors of the observation x [13] [14] [15]. Formally this simple fact can be described as follows. Let [16] [17] [18] [19].

$$L = \{(y_i, x_i), i=1, \dots, n_L\}$$

Where L denotes the set of learning data, $y_i \in \{1, \dots, c\}$ denotes class membership and the vector $x_i = (x_{i1}, \dots, x_{ip})$ represents the predictor values.

The determination of k-nearest neighbor is based on the calculation of similarity (or dissimilarity) between the new observation and all data of learning. The calculation of similarity is based on a distance function; the most used are the Euclidean distance, Manhattan distance, the Minkowski distance and the Hamming distance.

- Euclidean distance, which is the classic one, is given by:

$$d(x_i, y_j) = \sqrt{\sum_{s=1}^p (x_{is} - y_{js})^2} \quad (1)$$

- Minkowski distance

The Minkowski distance is more general than the Euclidean distance, it is given by:

$$d(x_i, y_j) = \left(\sum_{s=1}^p |x_{is} - y_{js}|^q \right)^{1/q} \quad (2)$$

Where q denote the constant of Minkowski, which takes the value 2 in the case of Euclidean distance and 1 in the case of the Manhattan distance.

- Manhattan distance

$$d(x_i, y_j) = \sum_{s=1}^p |x_{is} - y_{js}| \quad (3)$$

- Hamming distance

$$D_H = d(x_i, y_j) = \sum_{s=1}^p (x_{is} \oplus y_{js}) \quad (4)$$

$$D_H = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

For example, consider the following binary sequences:

$x = (0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1)$ et $y = (1 \ 1 \ 0 \ 1 \ 0 \ 1 \ 1)$ So
 $D_H = 1+1+0+0+1+0+0$. The distance between x and y is equal to 3 because 3 bits differ.

2.2. Weighted k-Nearest-Neighbors (wKNN)

The distance-weighted k-nearest neighbor rule (WKNN) was introduced by DUBANI [20]. According wKNN, the closer neighbors are weighted more heavily than the farther ones, using the distance-weighted function. The weight w_i for i-th nearest neighbor of the query x_i is defined as follow [17], [21], [22], [23], [24], [25].

$$W_i = \begin{cases} \frac{d(x-x_k)-d(x-x_1)}{d(x-x_k)-d(x-x_1)}, & \text{if } d(x-x_k) \neq d(x-x_1) \\ 1, & \text{if } d(x-x_k) = d(x-x_1) \end{cases} \quad (5)$$

3. Using KNN in educational and vocational guidance

3.1. General principle

The principle is to assign a new individual to a class of trades by calculating the resemblance between her profile (acquired) and the different profiles (profiles required) of the trades. Is a pairing based on a set of criteria (preacher's variables) whose weight varies according to the trade. The result of this pairing is a trade group considered nearest to the profile of the individual expressed by a query.

3.2. Categorization of trades

The vocational choice is based on several factors acting together. These factors do not have the same weight and are not stable throughout the life of the person [26].

In this approach, the trades must undergo a categorization according to different criteria which may be involved in the process of orientation. Such a categorization may appear according to the diagram in Figure 2 where C_i denotes a criterion and w its weight.

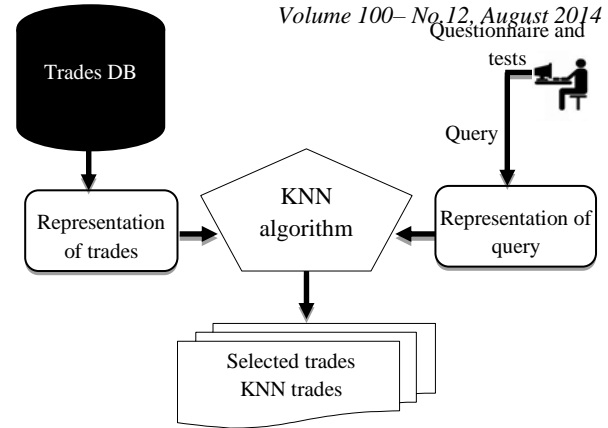


Fig 1: working principle

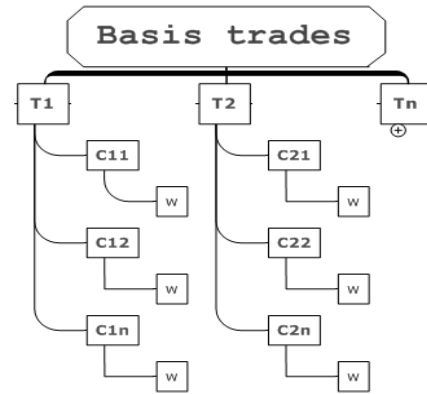


Fig 2: schematic description of trades' categorization according criteria and weight

To illustrate this categorization, we present below a non-exhaustive list of factors (criteria) that may be involved in the vocational choice is presented.

- Professional interests
- Personal values
- Personality Traits
- Education system
- Productive system
- Intelligence
- socio-economic environment
- Level of study
- Morphology and Health
- capacities
- sex
- Etc.

By leaving specialists to worry about criteria and their weights, we look at the implementation of a generic model that is easily adaptable and, to work properly, it requires a sufficient number of parameters.

3.3. Formalism of problem

3.3.1. Notation

In what follows, we use the following notations:

- m and c respectively denote the number of trades and the number of criteria
- B_{met} denotes the basis of trades
- B_{crt} denotes the basis of the criteria
- B_{crt} consists of trades transformed according to the terms of criteria

- Q represents the user query in natural language
- \bar{Q} Transformed Q in terms of criteria
- w_{ji} is the weight of j^{th} criterion for the i^{th} trade

3.3.2. Coding

Given a trade m_h , where $h \in \{1, \dots, m\}$ and a set of criteria c_j where $j= 1, \dots, c$ descriptor associated with m_h has the following form:

$$m_h = (w_{1h}, w_{2h}, \dots, w_{ch})$$

Where w_{jh} weight of criterion j in the trade m_h .

A descriptor of a business is the concatenation of the weights of all criteria.

We have, therefore, an individual

$$i = (m_1, m_2, m_3, \dots, m_m) \\ = (w_{11}, \dots, w_{c1}, w_{12}, \dots, w_{c2}, w_{1h}, \dots, w_{ch})$$

3.3.3. Similarity and dissimilarity

Definitions:

The notion of similarity or dissimilarity is introduced to measure the homogeneity of a group of observations by measuring the resemblances.

3.3.3.1. Dissimilarity

Dissimilarity is a function d which in any couple (x_1, x_2) associates a value in R^+ , and such as

- $d(x_1; x_2) = d(x_2; x_1) \geq 0$,
- $d(x_1; x_2) = 0 \rightarrow x_1 = x_2$

Two conclusions can be drawn from this definition:

- Distance is a dissimilarity because any distance has the previous two properties
- The less the units x_1 and x_2 resemble themselves, the more the score is high.

3.3.3.2. Similarity

Similarity is a function s which in any couple (x_1, x_2) associates a value in R^+ , and such as

- $s(x_1; x_2) = s(x_2; x_1) \geq 0$,
- $s(x_1; x_1) \geq s(x_1; x_2)$. [27]

Contrary to dissimilarity, the more the units x_1 and x_2 resemble the more the score is high.

In the context of this paper, we use two functions to calculate the resemblance between the profile of an individual and the various trades.

- Similarity function

The proposed similarity function is inspired by the field of information search in an information system [28]. It measures the degree of similarity between a user request and the trades. This correspondence function is defined by:

$$\text{simis: } \bar{B}_{\text{crt}} \times \{ \bar{Q} \} \rightarrow R_+ \\ (\bar{m}_i, \bar{Q}) = \frac{\sum_{c_j \in \bar{m}_i \cap \bar{Q}} w_{ji}}{1 + \sum_{c_j \in \bar{m}_i \cap \bar{Q}} w_{ji}} \quad (6)$$

Where w_{ji} weight of criterion j in the trade i

This function is an alternative of the cosine vector similarity used field of information search in an information system [28].

The advantage of the simis function is that uses only the weights of the criteria (w_{ji}).

- Dissimilarity using Euclidean distance

$$d(\bar{m}_i, \bar{Q}_j) = \sqrt{\sum_{s=1}^p (\bar{Q}_{js} - \bar{m}_{is})^2} \quad (7)$$

3.3.4. Case study

In this case, we propose to make a pairing between the profile of a user and the trades corresponding, based on professional interests of Holland (the RIASEC code), business sectors and the level of study.

3.3.4.1. RIASEC code

Holland (1966) proposed a theory of "vocational choice", distinguishing six categories of professional interests (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional), corresponding to different personality profiles. This classification is used to describe people, environments and their interactions; it also serves to establish a typology of "Vocation choice" which explains "the vocational choice" of an individual. Holland has shown this typology with a hexagonal pattern defining the psychological resemblances and interactions between personality types and environments [29], [30], [31].

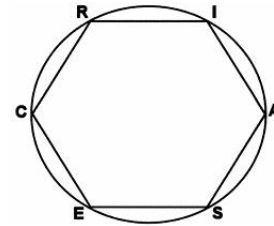


Fig 3: Circular representation of Holland RIASEC model

In this case, we take the first three dimensions for each individual with weight as 1.5, 1.0 and 0.5. The last three positions are assumed to be zero.

3.3.4.2. The business sectors

In this sample, we used the following sectors: education (Ed), health (Sn), social sectors (Sc), economy (Ec), administration (management) (Ad), technic (Tn), media (Md), culture, agriculture (Ag), nature (Na), environment (En), research (Rc), art and architecture (Ar) (ie 14 business sectors).

3.3.4.3. The level of study

At the moment, we classified these levels into three categories: alternating training, technician training and graduate studies (ie 3 level of study).

3.3.4.4. Choice of criteria weights

The choice of criteria and their weight is a critical operation and requires the intervention of domain experts. So by leaving specialists to worry about criteria and their weights, we look at the implementation of a generic model that is easily adaptable and, to work properly, it requires a sufficient number of parameters.

In this example, the user has to choose:

- His code RIASEC noting the three priority aspects (coded respectively 1.5, 1 and 0.5)
- His three favorite business sectors coded respectively 1.5, 1, and 0.5.

- His level of study

So a user whose RIASEC code is RCS (Realistic, Social, and Conventional), the level of study (LS) is N3 (graduate studies) and favorite business sectors are Ed (Education), Ar (Art) and Ad (Administration) is represented by a vector of 23 elements and is encoded as follows:

1.5;0;0;0.5;0;1 0;0;0;1;1.5;0;0; 0;0.5;0;0;0;0 0;0;1

3.3.4.5. Sample of the trades

The table below shows an extract of the trades' data base

RIASEC code	Business sectors	LS
-------------	------------------	----

Table 1: a sample of trades

Trade	RIASEC CODE			Business activity			Level of study
	Choice 1	choice 2	Choice 3	Choice 1	Choice 2	Choice 3	
Farmer	R	S	E	Agriculture	Nature	Environment	alternating training
Assistant Mechanical automobile	R	I	C	Technic	Economy	Administration	alternating training
assistant horticulturist production	R	I	C	Agriculture	Nature	Environment	alternating training
architectural draftsman	R	C	E	Architecture			alternating training
Technician Electrical division, energy section	R	I	E	Agriculture	Nature	Environment	technician training
Chief editor film	R	A	C	Technic			graduate studies
Educator	S	E	A	social sectors	Education	Health	graduate studies
male nurse	S	A	I	social sectors	Health	Education	technician training
schoolteacher	S	A	C	social sectors	Education	Health	graduate studies
architect engineers	A	E	R	Architecture			graduate studies

4. RESULTS AND DISCUSSION

The table below shows an extract of the results based on a sample of 100 trades.

Table 2: A sample of results

- K=1

Similarity function		Euclidean distance		Same result?
function value	KNN Trades selected	distance value	KNN Trades selected	
0,8750	Trade 1	0,0000	Trade 1	YES

- K=2

Similarity function		Euclidean distance		Same result?
function value	KNN Trades selected	distance value	KNN Trades selected	
0,8750	Trade 1	0,0000	Trade 1	YES
0,8462	Trade 4	1,5811	Trade 4	YES

- K= 3

Similarity function		Euclidean distance		Same result?
function value	KNN Trades selected	Value distance	KNN Trades selected	
0,8750	Trade 1	0,0000	Trade 1	YES
0,8462	Trade 4	1,5811	Trade 4	YES
0,8333	Trade 13	2,0000	Trade 13	YES

- K= 4

Similarity function		Euclidean distance		Same result?
---------------------	--	--------------------	--	--------------

function value	KNN Trades selected	Value distance	KNN Trades selected	
0,8750	Trade 1	0,0000	Trade 1	YES
0,8462	Trade 4	1,5811	Trade 4	YES
0,8333	Trade 13	2,0000	Trade 13	YES
0,7778	Trade 7	2,2361	Trade 39	NO

- K=10

Similarity function		Euclidean distance		Same result?
function value	KNN Trades selected	Value distance	KNN Trades selected	
0,8750	Trade 1	0,0000	Trade 1	YES
0,8462	Trade 4	1,5811	Trade 4	YES
0,8333	Trade 13	2,0000	Trade 13	YES
0,7778	Trade 7	2,2361	Trade 39	NO
0,7778	Trade 25	2,3452	Trade 25	YES
0,7500	Trade 2	2,6458	Trade 16	NO
0,7500	Trade 5	2,7386	Trade 7	NO
0,7500	Trade 9	2,7386	Trade 22	NO
0,7500	Trade 10	2,7386	Trade 23	NO
0,7500	Trade 39	2,7839	Trade 5	NO

The obtained results allow us to make a comment according to two axes:

Firstly, these results show that the two functions, similarity or dissimilarity arrive at finding the best near neighbors. However, when k exceeds a certain threshold (3 in table II), we do not get the same nearest neighbors. In the previous example, when using the similarity function simis, the 4th nearest neighbor is Trade7, but it is the Trade39 when using dissimilarity (Euclidean distance). This has led us to several tests to determine the function that gives the best results: similarity or dissimilarity, and we concluded that the similarity function simis can often return the best nearest neighbor. In

fact, the graph (figure 5) allows a comparison between a new

observation (NO) and two trades: Trade7 and Trade39.

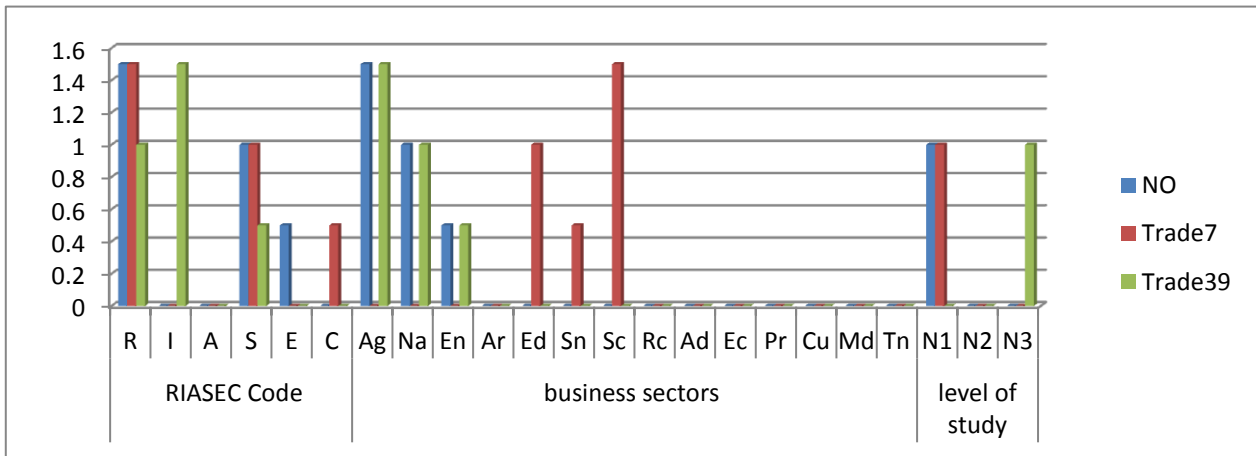


Fig 5: comparison of 4th NN Trade7 calculated using simis function and Trade39 calculated using Euclidean distance [32]

The table below summarizes the results of the previous comparison.

Table 3: An example of resemblance degree

Criteria	resemblance degree	
	similarity	Dissimilarity
RIASEC code	83.3%	0%
Business sectors	0%	100%
level of study	100%	0%

After processing 7 examples randomly selected and are based on the previous example, we concluded that the use of the simis function (similarity) defined above, allows to return the best results compared to the dissimilarity function processed by the Euclidean distance.

Secondly, the results showed that we obtain even values for different trades what makes their classification a difficult task;

this is true for the function simis and the Euclidean distance. For example, the table II shows that the simis function returns

the value 0.7778 for the two trades Trade7 and Trade25 and the value 0.7500 for the five trades Trade2, Trade5, Trade9, Trade10 and Trade39. To overcome this problem of resemblance of the closest neighbors, we propose two approaches:

- weighting of the criteria of guidance

Generally speaking, a criterion does not represent same importance for all the trades. Let us take, for example, the criterion level of study, it is very important for certain trades, but less important for some others. And so on for the other criteria.

To clarify this, we compare the two trades Trade7 and Trade25 whose corresponding value calculated by the simis function is 0.7778.

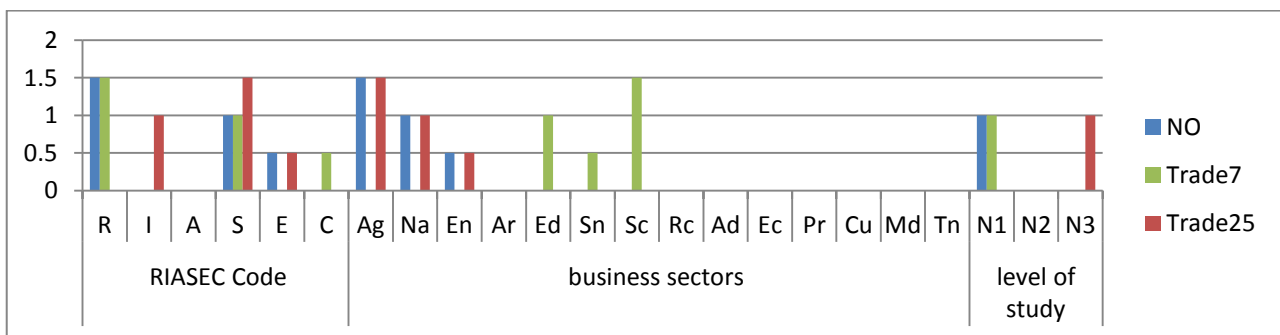


Figure 6: Comparison of degrees similarity of the trade7 and trade25 compared to the new observation (NO)

The figure 6 shows that the new observation is very similar to the Trade7 with regard to the code RIASEC and the level of studies and totally different with regard to the business sector. On the other hand, the similarity between the new observation and the Trade 25 is perfect with regard to the business sector and zero with regard to the other criteria. Therefore, to differentiate and classify these two Trades (Trade7 and Trade25), the weighting of criteria can be used.

- using a Meta classifier KNN

In this view, a Meta KNN is a KNN applied only to the closest neighbor already selected. So we have to use the principle of KNN twice:

- First, the principle of KNN is applied to all of the data to determine the k-nearest neighbor using simis function.

- Secondly, to overcome the problem of resemblance of nearest neighbors, the principle of KNN is only applied on k-nearest neighbors using the Euclidean distance.

The algorithm is as follows

Algorithm: The Meta KNN

Step 1: Compute the similarity of nearest neighbors of a query Q using simis function

for $i = 1$ to N do

$$\text{simis}(\bar{m}_i, \bar{Q}) = \frac{\sum_{c_j \in \bar{m}_i \cap \bar{Q}} w_{ji}}{1 + \sum_{c_j \in \bar{m}_i \cap \bar{Q}} w_{ji}}$$

end for

Step 2: Determine the k closest neighbors

for $i = 1$ to N do

Sort out in a decreasing way the nearest neighbors

Take the k closest neighbors

end for

Step 3: Overcome the problem of resemblance of the closest neighbors

for $i = 1$ to K do

% Compute the distance Euclidean between the new observation x'

% and the K nearest neighbors having the same value using Euclidean distance

$$d(x_k, x') = \sqrt{\sum (x' - x_k)^2}$$

end for

Step 4: Give the k- nearest neighbors

So, according this algorithm, the result of the table 2 become as follows:

Table 4: The results obtained by using the Meta KNN

Similarity function	Euclidean distance between NO and KNN	Final results
function KNN Trades value selected	Value distance	
0,8750 Trade 1	-	Trade 1
0,8462 Trade 4		Trade 4
0,8333 Trade 13	-	Trade 13
0,7778 Trade 7	2.7386	Trade 25
0,7778 Trade 25	2.3452	Trade 7
0,7500 Trade 2	3.0000	Trade 39
0,7500 Trade 5	2.7839	Trade 5/9
0,7500 Trade 9	2.7839	Trade 9/5
0,7500 Trade 10	2.9580	Trade 10
0,7500 Trade 39	2.2361	Trade 2

5. CONCLUSION AND PERSPECTIVES

In this article, the KNN method is used to determine the closest trades to an individual according to his profile calculated by using three criteria: professional interests of Holland (the RIASEC model), business sectors and the level of study. The method was used with two resemblance functions: a similarity function (simis function) and a dissimilarity function using Euclidean distance. The results showed, first, that the process restores the nearest neighbors and, on the other hand, from some k we do not get the same nearest neighbors with the two functions. The tests allowed us

to conclude that the similarity function simis often gives the best results.

To overcome the problem of resemblance of nearest neighbors, we used a Meta KNN applied only to the closest neighbors.

Subsequent work in the framework of this approach will be focused on two main points: using wKNN (weighted KNN) on all of the criteria involved in the process of guidance and confront this method with the other methods to compare and choose the best.

6. REFERENCES

- [1] Davies, S., & Guppy, N. (1997). Fields of study, college selectivity, and student inequalities in higher education. *Social Forces*, 75 (4), 1417–1438.
- [2] Simpson, J. C. (2001). Segregated by subject: racial differences in the factors influencing academic major between European Americans, Asian Americans, and African, Hispanic, and native Americans. *The Jthinsal of Higher Education*, 72(1), 63–100.
- [3] Galotti, K. M. (1999). Making a “Major” real-life decision: college students choosing an academic major. *Jthinsal of Educational Psychology*, 91(2), 379–387.
- [4] Coperthwaite, C. A. (1994). The effect of self-esteem, locus of control, and background factors on college students’ choice of an academic major. Unpublished doctoral dissertation, The University of Connecticut, Connecticut, USA.
- [5] Betz, N. E., & Rottinghaus, P. J. (2006). Current research on parallel measures of interests and confidence for basic dimensions of vocational activity. *Jthinsal of Career Assessment*, 14(1), 56–76.
- [6] Lounsbury, J. W., Smith, R. M., Levy, J. J., Leon, F. T., & Gibson, L. W. (2009). Personality characteristics of business majors as de fined by the Big Five and narrow personality traits. *Jthinsal of Education for Business*, 84(4), 200–206.
- [7] Leppel, K. (2001). Race, Hispanic ethnicity, and the future of the college business major in the United Sates. *Jthinsal of Education for Business*, 76(4), 209–215.
- [8] Ware, N. C., Steckler, N. A., & Leserman, J. (1985). Undergraduate woman: who chooses a science major? *Jthinsal of Higher Education*, 56(1), 73–84.
- [9] ELHAJIE, AZMANI.A, ELHARZLI.M, Expert system design form educational and vocational guidance, using a multi-agent system, proceeding of the 4th IEEE conference ICMCS'14, 14-16 April 2014 –Marrakech-Morocco.
- [10] Francisco Rivas, Tecnología informática en asesoramiento vocacional, Psicothema, 2005
- [11] S.A.I.O. – Rectorat de Versailles, Les logiciels d’aide à l’orientation, 2007.
- [12] Roger Chappat, “L’informatique comme aide à l’orientation”, LE BULLETIN DE L’EPI N° 65.
- [13] K. Ming Leung, k-Nearest Neighbor Algorithm for Classification, 2007
- [14] Jérôme Azé, K-plus proches voisins, 2007

- [15] Eve MATHIEU-DUPAS, Algorithme des K plus proches voisins pondérés (WKNN) et Application en diagnostic, nria-00494814, version 1 - 24 Jun 2010
- [16] Klaus Hechenbichler, and Klaus Schliep, "Weighted k-Nearest-Neighbor Techniques and Ordinal Classification", Sonderforschungsbereich 386, Paper 399 (2004).
- [17] Jianping Gou and all, "A New Distance-weighted k-nearest Neighbor Classifier", *Journal of Information & Computational Science* 9: 6 (2012) 14291436
- [18] J. M. Keller, M. R. Gray, and J. A. Givens. A fuzzy k-nn neighbor algorithm. *IEEE Trans. Syst. Man Cybern.*, SMC-15(4):580–585, 1985.
- [19] COVER T.M., HART P.E., Nearest neighbor pattern classification, *IEEE Trans. On Inform. Theory*, Vol 13(1) pp. 21-27, 1967.
- [20] S. A. Dudani, The Distance-weighted k-Nearest Neighbor Rule. *IEEE Trans. Syst. Man Cybern.*, 6 (4) (1976), 325 – 327.
- [21] Jianping GOU and all, " Weighted K-nearest Centroid Neighbor Classification", *Journal of Computational Information Systems* 8: 2 (2012) 851–860.
- [22] Yu Takigawa and all, "Pattern classification using weighted average patterns of categorical K-Nearest Neighbors", Department of computer and Information Sciences, Nagasaki University.
- [23] DONOEUX T, A k-nearest neighbor classification rule based on Dempster-Shaffer theory, *IEEE Trans. on Systems, Man and Cybernetics*, Vol 25(5), pp. 804-813, 1995.
- [24] Y. Zeng, Y. Yang, L. Zhao, Pseudo nearest neighbor rule for pattern classification, *Expert Systems with Applications*, 36 (2009), 3587-3595
- [25] Bruno Taconet and all, Classification des k-ppv par sous-voisinages emboîtés, Equipe GED – Université du Havre.
- [26] María Luisa López González, "la toma de decisiones en los sistemas de autoayuda y asesoramiento vocacional (sav-r y savi-2000): propuesta y validación de un modelo de decisión vocacional", memoria para optar al grado de doctor , Universidad Complutense de Madrid, Facultad de Psicología, Departamento de Psicología Evolutiva y de la Educación
- [27] E. Lebarbier, T. Mary-Huard, Classification non supervisée.
- [28] Fatou Kamara-Sangaré, Contribution à la recherche d'information : Une fonction de correspondance, Actes des 1ères Rencontres Jeunes Chercheurs en Recherche d'Information.
- [29] Dina Guglielmi, Franco Fraccaroli and Maria Luisa Pombeni, The Professional Interests According to Holland's Hexagonal Model Structures and Gender Difference", *OSP*, 33/3|2004, p. 409-427.
- [30] Biljana Stevanovic and Nicole Mosconi, «Les représentations des métiers des adolescent(e-s) scolarisé(e-s) dans l'enseignement secondaire », *Revue française de pédagogie [En ligne]*, 161 | octobre-décembre 2007.
- [31] Lisa M. Larson, Patrick J. Rottinghaus, and Fred H. Borgen, Meta-analyses of Big Six Interests and Big Five Personality Factors, *Journal of Vocational Behavior* 61, 217–239 (2002).
- [32] ELHAJJE, AZMANI.A, ELHARZLI.M, A pairing individual-trades system, using KNN method: the educational and vocational guidance as a case study, *IEEE Conference CIST'14*, 20-22 October 2014, Tetuan, Morocco (Paper accepted).