A comparative study of classification methods: case of application to Asian cuisines ingredients

TSHIMANGA KAMANGA Celestin[1], TSHIMANGA KAPUBA Patrick[2], NTUMBA BADIBANGA Simon[3]

[1] Mathématiques et Informatique, Université de Kinshasa, Kinshasa, DRC

[2] Mathématiques et Informatique, Université de Kinshasa, Kinshasa, DRC

[3] Mathématiques et Informatique, Université de Kinshasa, Kinshasa, DRC

Abstract

In this study, we devise a novel classification approach using machine learning based on diet preferences of populations living in geographically remote areas. Previously, some researchers have performed different computational methods to discover the similarity between food in a common environment for all recipes. The correlation in our approach is based on the similarity between different combinations of ingredients belong to the particular geographic areas.

The food data for the related countries was downloaded from : www.bigoven.com. To be sure of the outcome of this study, three algorithms such as: logistic regression, support vector machine, and neural networks were applied to the same set of data but with different parameters. After performing several experiments and analyzing the results obtained from these three algorithms, we deduce that the support vector machine support vector machine with a normalized polynomial kernel worked better for this classification and prediction task.

Keywords: *Nutritional ingredients; Asian recipes; SVM; Machine learning.*

1. Introduction

As the population grows and the number of travelers increases around the world, the development of automatic food-based technologies becomes essential in order to cope with environmental change issues. Indeed, nowadays, there are thousands of recipes all over the world, which are freely available and published by the people from different cultures and with deep culinary skills that are accessible to all. This availability of information on many recipes from different cultures is very helpful because it facilitates the exchange of recipes between cultures and regions geographically distant from each other, but also presents a major challenge that it is difficult to find specific recipes among millions of other recipes. To do this, it is important to find ways to improve research on food recipes based on several parameters such as ingredients, cuisine, food category, nutritional values, costs, etc.[1]

Thus, the persistence of our study is to suggest a classification system based on the ingredients of Asian cuisines, which takes into account food recipe choices as well as can be recommended to a person living in a country other than his country. So, finding the similarity between cuisines is one of the important tasks that encompass a lot of problems because it sets the trend amid diverse cuisines that are built on the ingredients recycled in the cuisine, regardless of the cooking methods.

However, the choice of an appropriate machine learning algorithm to solve this problem seems to be very difficult due to the number of existing algorithms which are described in the literature. In addition, the development of this type of system using machine learning algorithms often has parameter setting problems that must be evaluated.

Thus, implementing the support vector machine model on the sample data, it has been created a simple prediction approach to illustrate a classifier. So far, the analysis of the data is still applicable in different directions, i.e. starting from the classification of the texts, the medical tests and the marketing focused on the customer until to the analysis of the food data. Therefore, the great problem of the best classifier depends on the data structure on which the analysis is performed and the constraints that measure the rate of false rejection and bad acceptance.

The SVM classifier used in this research has been compared to other two classifiers such as the logistic regression and neural network. After applying all these methods to our dataset, we considered that the SVM with a normalized polynomial kernel classifier worked well in the

food prediction task compared to other classifiers. The data used for our study have been downloaded online and are limited only to food data from Asian countries such as Vietnam, India, Thailand, Japan and China because we assume that these countries are geographically closed and their people eat almost the same diet.
Human existence is impossible without food. It gives energy to our body to live, work and performs our daily tasks. All in all, it helps to human race to exist. Besides the fact that It helps to in some celebrations like marriage and birthdays as well as it is sensual, entertaining and a manner of gaining experience of new beliefs and new landscapes [2]. However, with the advancement in technology, quality of food and dietary habits of the people have been changed that give researchers, nutritionist and food experts many chances to explore its different research areas. Due to these factors our study is related to this research field.

2. RELATED WORK

Recently, it was difficult to establish a logical connection between the recipe contents of different cuisines, that is, people were limited to eat only what they knew. Therefore, for lack of this connection, it was sometimes impossible to automatically produce combinations of ingredients in order to obtain a new recipe never known in advance and also the automatic labeling of recipes based on the ingredients seemed to be an almost impossible or uncertain task. Due to these difficulties, recent research on the systems automation of cuisines based on their contents have already been approached in several ways such as: Data-driven recipe completion using machine learning approaches", which can be used to figure simulations to complete automatically a recipe[3]; A list of options generation system, that uses a recipe dataset and notes to suggest a menus according to consumer's predilections [4]; Automatic Recipe Cuisine Classification by Ingredients, which investigate the underlying correlation between recipes cuisine and ingredients[5]; Purposeful diets improvement (trends and skills), which enumerates an indication of changing skills to develop targeted foods [6]; exposure evaluation in the a full regime learning.: A judgment of the usage of the Pan-European grouping structure, which defines whether it is thinkable to usage a pan-European matched food grouping structure[7]; A user-focused on ideal for analyzing the fitness of food grouping structures, which analyzes the relevance or validity of some food grouping system as of the consumer's point of opinion[8]; Customer's getting and predilections for diet changed and purposeful dairy foods, which propose an efficient works analysis amasses and reviews study on customer receipt and predilections for food-changed and efficient dairy foods, to reunite, and increase upon, the results of preceding lessons[9]; A novel method for showing customers favorites is suggested in order to denote the taste behavior and the indecision of the customers, and to clarify the part that the users' covariates such as sex, profession, etc. Show in the valuation of the favorites[10]; Plan, care and Campaign for Nutrition correlated holiday business Creativities, which examine the behaviors in which diet and holiday business are being strained together at a theoretic side by side by researchers, at a planned side by side by rule producers, and at an smeared level by designers and experts[11] and Growing a method resemblance measure for suggesting healthy mealtimes, which propose a food suggesting approach which excite healthful and various consumption, when the offered foods fit the choice of the consumer[12].

Therefore, in this thesis, we proposed ingredients based classification of Asian cuisines using machine learning methods. After analyzing the results of each machine learning method on the cuisines data from several Asian countries, we proposed the best method that can be used in this case and considered as a one of main objectives of our project. By establishing a link of our results to the other domains such as: cultural, people behavioral of these Asian countries, etc., we can say that we achieved our goal properly.

3. METHODOLOGY

3.1 Binary logistic regression

Binary logistic regression is a variety of regression appropriate when dependency is a contrast. In the situation of a binary logistic regression model, the values of the dependent variable have two sorts, specifically occurrence $(d = 1)$ and non-occurrence $(d = 0)$. Given x reflected as an event, typically the probability of occurrence can be dignified

$$P((d=1)|x) \text{, and}$$
$$P((d=1)=1-P(d=1)|x); \qquad (1)$$

Which represents the alteration of the logistic function of $P((d=1)|x)$ which is known as the logit alteration below:

$$\vartheta(P((d=1)|x)) = \log it(P((d=1)|x)) = In\left(\frac{P((d=1)|x)}{1-P((d=1)|x)}\right), \quad (2)$$

Where $\vartheta(P((d=1)|x))$ which can be applied by a linear function of independent variables $a_1, a_2, ..., a_k$, and then, (3.18) can be revision as:

$$\vartheta(P([d=1])|x) = n\left(\frac{P((d=1)|x)}{1-P((d=1)|x)}\right) \quad (3)$$

$$= n_0 + n_1 a_1 + n_2 a_2 + ... + n_k a_k \quad (4)$$

Which represents the intercept, and $n_1, n_2, ..., n_k$, represents the regression coefficients of $a_1, a_2, ..., a_k$, respectively. So, the above probability of incidence of event $P((d=1)|x)$ can be dignified as:

$$P((d=1)|x) = \frac{e^{\vartheta(P((d=1)|x))}}{1+e^{\vartheta((d=s)|x)}} \quad (5)$$

$$\frac{e^{n_0+n_1a_1+n_2a_2+...+n_ka_k}}{1+e^{n_0+n_1a_1+n_2a_2+...+n_ka_k}} \quad (6)$$

Finally, the discriminant rules can be engendered as follows:

$$d = \begin{cases} 1, P((d=1)|\, x > 0.5) \\ 0, P((d=1)|\, x) \le 0.5 \end{cases} \quad (7)$$

3.2 Multinomial logistic regression

Multi-nominal logistic regression is significant once the reliant on variable is nominal, i.e. comprises more than two classes. Suppose χ is an event and the dependent variable has m classes $(m>2)$. The alteration of the logistic function of $P((d=j)|x)((d=j)|x)(1 \le j \le m-1)$ is altered like this:

$$\vartheta(P((d=1)|x)) = In\left(\frac{(P((d=1)|x))}{(P((d=m)|x))}\right) \quad (8)$$

$$= n_0^1 a_1 + n_2^1 a_2 + ... + n_k^1 a_k \quad (9)$$

$$\vartheta(P((d=2)|x)) = In\left(\frac{(P((d=2)|x))}{(P((d=m)|x))}\right) \quad (10)$$

m is taken is this case as the reference class, n_0^J that indicates the intercept, and $n_1^j, n_2^j, ..., n_k^j$ represents the regression coefficients of $a_1, a_2, ..., a_k$ in the j^{th} category, respectively $(1 \le j \le m-1)$. From the above exploration, the probability of incident of event can be define as:

$$P((d=j)|x) = \frac{e^{\vartheta(P((d=j)|x))}}{1+\sum_{S=1}^{m-1}\left(e^{\vartheta(P((d=s)|x))}\right)} \quad (11)$$

$$= \frac{e^{n_0^s+n_1^s a_1+...+n_k^s a_k}}{1+\sum_{s=1}^{m-1} e^{n_0^s+n_1^s a_1+...+n_k^s a_k}} \quad (12)$$

For the probability of incident of event $P((d=m)|x)$ can be defined as follows:

$$P((d=m)|x) = \frac{1}{1+\sum_{s=1}^{m-1}\left(e^{\vartheta(P(d=s)|x)}\right)} \quad (13)$$

$$= \frac{1}{1+\sum_{s=1}^{m-1} e^{n_0^s+n_1^s a_1+...+n_k^s a_k}} \quad (14)$$

IJCSI International Journal of Computer Science Issues, Volume 17, Issue 3, May 2020
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.5281/zenodo.3987151

Finally, the discriminant procedures can be engendered as follows:

$$\arg\max x_{t-1}^{m} P((d=t)|x) \quad (15)$$

Accordingly, we come to discriminate the maximum probability of occurrence of the event as a class. However, this way of discriminating can be realistic to both the binary scenario and the multinomial scenario[14, 15].

3.3 Support Vector Machine

3.3.1 Polynomial kernel

Naturally, the polynomial kernel forms not only the features of input examples to decide their resemblance, but also mixtures of these. In the circumstance of regression analysis, such mixtures are known as interaction features. The feature space of a polynomial kernel is equal to that of polynomial regression, but deprived of the combinatorial expandable in the number of parameters to be learned. When the input features are binary-valued, then the features resemble to logical combinations of input features. For degree d polynomials, the kernel is defined as:

$$K(x,y) = (x^T y + c)^d \quad (16)$$

x and y are considered as vectors in the input space. So, features vectors calculated between training or test examples and $c \geq 0$ is a permitted parameter exchange off the effect of higher-order against lower-order relations in the polynomial. When $c = 0$, the kernel is named homogeneous. A further generalized poly kernel divides $x^T y$ by a user-specified scalar parameter α.

3.3.2 Normalized polynomial kernel

In this case, the polynomial kernel is followed by certain normalization, such that the normalized kernel is not ever external of $(-1,1)$ for odd exponent and $(0,1)$ for even exponents. It's not easy to say why that achieves better deprived of knowing more about what type of features you are used to denote your data.

4.3.3 Pearson VII universal kernel (PUK)

Is one of the functions that can be used SVM. In the case of multi-dimensional it's given by:

$$K(x_i, x_j) = 1 / \left[1 + \left(2\sqrt{x_i, x_j}\sqrt{2^{\frac{1}{W}} - 1} / \alpha \right) \right]^2 \quad (17)$$

Where (x_i, x_j) represent the independents variables; w and α check the half-width and tailing factor of the peak. The principal reason to use the PUK, it is flexibility to change by varying the parameters w and σ.

4.3.4 The Radial Basis Function Kernel (RBF)

The Radial basis function kernel, also called the RBF kernel, or Gaussian kernel, is a kernel that is in the usage of a radial basis function (more specifically, a Gaussian function). The RBF kernel is defined as

$$K_{RBF}(x_i, x_j) = \exp\left[-\gamma \left\| x_i - x_j \right\|^2\right] \quad (18)$$

Where γ is a parameter that set the spread kernel.

4.4 Multi-Layer Perceptron

There are several categories of artificial neural networks and each category is related to precise task. For the classification case it is the Perceptron Multi-Layer (MLP) which is the most used. This variant maps an input layer otherwise called a prediction attribute to an output layer known as target attributes based on strongly connected hidden layers.

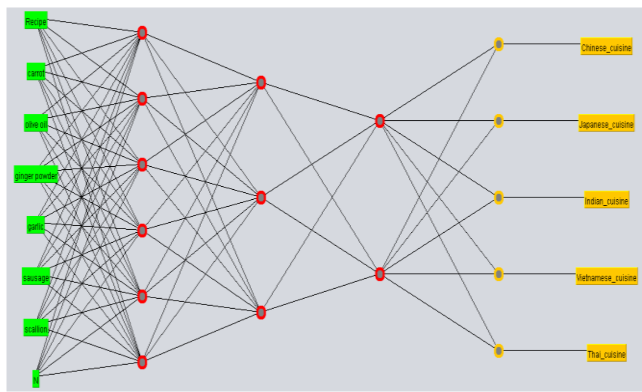


Fig. 1 Sample structure of our multi-layer perceptron

4. DATA COLLECTION AND FORMATING

The data has been taken from https://www.bigoven.com/. This link is composed a set of data covering cuisines for each country, the recipes for each cuisine and the ingredients as atomic elements of each recipe considered here as being the respective country. This web site contains more than 350000 recipes and at least 300 ingredients from several cuisines. For much more detail, BigOven is a platform that features organized and inspired cooks in the cuisine and on the go. With more than 6 million downloads of its cookbook apps, BigOven assistances people strategy great meals. In fact, BigOven platform is a menu designer, grocery list manager, recipe manager, nutrition facts, reviews and tagging[13]. Of our interest, we have focused on five cuisines such as:

Chinese's cuisine, Japanese's cuisine, Vietnamese's cuisine, Indian's cuisine and Thai's cuisine. Thus, our file has been transformed into a binary matrix $L(n\times m)$, consisting of all the cuisines, recipes and ingredients under this format:

$$L = \begin{cases} 1, & \textit{if recipe j contain ingredient k} \\ 0, & \textit{otherwise.} \end{cases}$$

After this step in matrix L we added another cuisine column linking each recipe to its corresponding cuisine. As we know that, most of classifiers share a common characteristic, which is this one of definition's construction of a function that associates to a given element a value 0 or 1, thus representing its membership or not to the category or class. The matrix L can be represented in the following way:

- Each cuisine c is composed at least one recipe;
- Each recipe r is composed at least one ingredient, i.e. If, for example, the ingredient i represents to one of the recipe's elements r during the way cooking, then the relation between recipe r and ingredient i is represented by value 1, otherwise 0;
- Finally, in each recipe r, an ingredient r can appear only once.

Note: each cuisine c represents the set of all foods recipe belonging to a given country.

Using the machine learning representation, we can otherwise define our transformation as follows:

Let us consider $S=\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\ldots,(x^{(n)},y^{(n)})\}$ defined as a subset of all recipes; where x^{i} defines the combination of certain ingredients for any given recipe. Therefore, $x_j^i=1$, that means an ingredient j appears in the combination of example i and 0 otherwise. With $j=1,2,3,\ldots,l$ and all cuisines are defined by $y^{(i)}\in\{1,2,3,\ldots,k\}$, where k is the number of cuisines in our training set. To facilitate the processing of our learning set, we use the frequency of the i^{th} feature. Thus, $tf(x_v^w)$ stands for term frequency of ingredient v in recipe w. The frequency term is very similar to the word bag that means that every word will be over-weighted by the way it occurs in a document, it is just in a bag of words. For example, if we have word that occurs t times, it's going to have t times as much weight as a word that occurs only once. So, $tf=1$ if a recipe w contain ingredient v and 0 otherwise. $idf(v_i,S)$, that characterizes the inverse document frequency to measure the number of times an ingredient repeats in the dataset and is measured by:

$$idf(v_i,S)=\log\frac{N}{|v\in S: v_i=1|} \tag{19}$$

where , is the total number of recipes in our training set an $|v\in S: v_i=1|$ which defines the total number of ingredient recipes that meet the criteria. Finally, we can calculate the frequency by:

$$tf-idf=(x_v^w,S)=tf(x_v^w)\times idf(v_i,S) \tag{20}$$

5. EXPERIMENTAL RESULTS

In this section, we first described the implementation steps for each algorithm mentioned in the general introduction. Second, we used more parameters to test each algorithm in order to keep the parameter that gives the best accuracy. Third we compared the optimal parameters of the three algorithms and selected parameter that gives a greater precision compared to the others and finally we used this parameter to realize the classification or the prediction of the cuisines.

5.1 Computation steps of each algorithm

5.1.1. Support Vector Machine Implementation

We have implemented the sequential minimal optimization which is an algorithm that solves the quadratic programming problem that occurs during learning of the support vector machine. To do this, we have succeeded to use the penalty parameter C which is tuned to improve the generalization performance of the model. After several experiments, we conclude that the best accuracy for the types of kernel function applied on our dataset is getting when the penalty parameter C is belonging to the gap of 0.5 to 2.0, that means if the penalty is less than 0.5 or great than 2.0, the accuracy is going to decrease. In order to evaluate the best kernel function that optimize this problem correctly, we have set the penalty parameter C to 2.0 for each kernel function and we got the result mentioned below:

Table 1: Assessment of learning accuracy from kernels (SVM)

Types of kernels function	Learning Accuracy (%)
Polynomial	71.3
Normalized Polynomial	75.6
PUK	69.9
RBF	68.3

A kernel function of SVM based on the Normalized Polynomial has been applied and compared with the generally applied kernel functions, i.e. the RBF, PUK and polynomial to classify and predict the cuisines. From the table above, it is found that the SVM model based on Normalized Polynomial shows the good accuracy ($\approx 76\%$) in classification and prediction of cuisines than SVM based RBF kernel, polynomial and PUK.

5.1.2 Neuronal network Implementation

The structure of ANN implemented in this thesis is composed of N inputs which represents the all features (recipes of ingredients) and five output that denote the five cuisines, that means it's composed of N neurons in the input layer and five neurons in the output layers. During the implementation processes the hidden layers of this study are activated by a simple group of function $\{h_{(k)}(x), k = 1,2,3,..., k\}$. So, k represents the number of the hidden layers neurons.
In order to calculate the number of hidden neurons, we used a win-taker method to specify which class the input recipes belong to. To do this, the neurons of the output layer entered the competition stage where the neurons are set to 0. The neuron with 1 decides the cuisine to which the input recipe belongs.
To certify statistically a best accuracy with artificial network neural for solving our problem, 45 executions were performed and for each execution, the learning rate was setting to 0.3 and momentum to 0.2. In this context, given any number of neurons for each layer, the conclusion is given bellow:

Table 2: Assessment of learning accuracy from ANN

Types of Hidden layer	Learning Accuracy (%)
Single Hidden layer	< 60
Double Hidden layers	≈ 68
Multiple Hidden layer	< 25

From Table 2, we noticed that during the execution process when the number of hidden layers exceeded 2, the accuracy always remained low and decreased with each addition of a supplementary hidden layer. In the case of a single hidden layer, for any number of neurons assigned to the hidden layer, the accuracy is always in the range of 50%. Finally, we showed that double hidden layers, is the best structure ANN for task, because, even given any number of neurons for two layers the accuracy was maintaining to around 68%. Finally, we have shown that hidden double layers are the best ANN structure for this task, because, for any number of neurons assigned to these two hidden layers, the accuracy was always around 68%.

5.1.3 Logistic regression

We applied the multi-class classification logistic to our dataset tuning the parameters: inverse of regularization and penalty, penalty. So to test accuracy for each parameter, we fixed batch size to 1000 instances and we have gotten the result below:

Table 3: Assessment of learning accuracy from multi-class logistic regression

Parameters	Learning Accuracy (%)
Inverse of regularization	69.5
Penalty	50.5

From the table 3, it's clearly demonstrated that the best accuracy obtained from the logistic regression of the multi-class classification is that which is configured with a reverse regularization parameter, because, its accuracy is higher than this one of penalty.

5.2 Discussion

After a thorough analysis of the previous results, we retain the technique of each method which gives a better accuracy, which is to say for:
SVM: We retain Normalized Polynomial Kernel technique (75.6% accuracy)
Logistic regression: We retain Inverse of regularization technique (69.5% accuracy)
ANN: We retain double hidden layers technique ($\approx$ 68 % accuracy)

However, before making an optimal decision on the best technique among the three techniques retained so far, we also thought of other evaluation parameters such as:

IJCSI International Journal of Computer Science Issues, Volume 17, Issue 3, May 2020
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.5281/zenodo.3987151

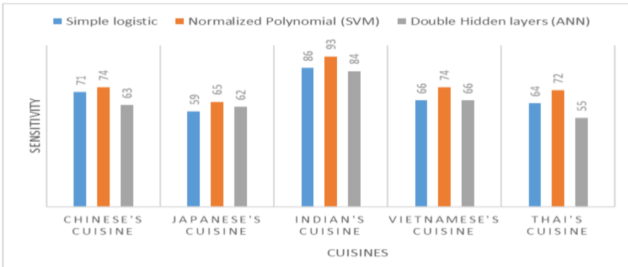
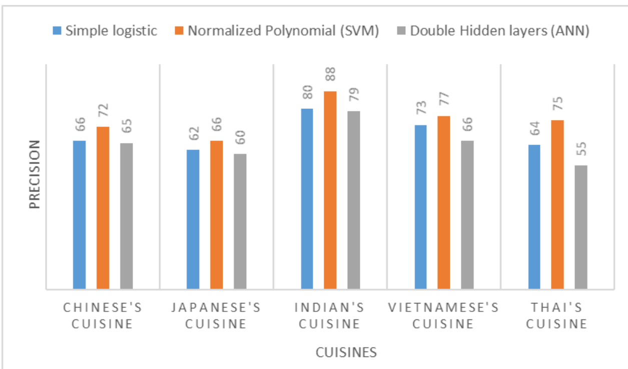


Fig. 2 Comparison between sensitivity of cuisines

Fig. 3 Comparison between precision of cuisines



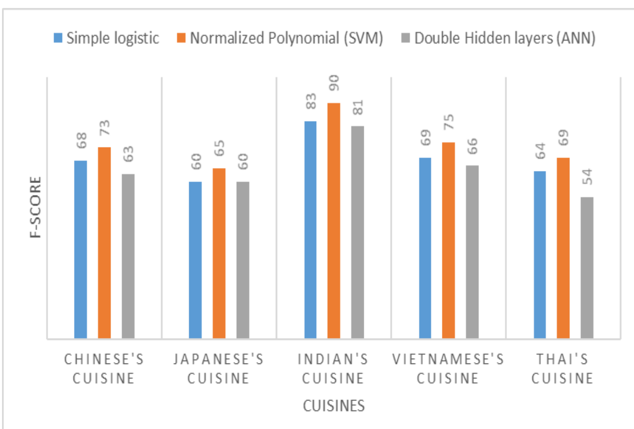


Fig. 4 Comparison between F-Score of cuisines

Table 4: Classification result in the form of a confusion matrix

	C	J	I	V	T
C	113	21	6	4	8
J	31	97	3	13	6
I	3	2	141	1	4
V	3	14	4	112	18
T	8	12	7	16	108

C: *Chinese_Cuisine*. J: *Japanese_CuisineI*: *Indian_Cuisine*. V: *Vietnamese_Cuisine* T: *Thai_Cuisine*

First, according to the accuracy of every method, we have seen that, only SVM with Normalized Polynomial Kernel do noticeably better for this task, while some of the other method such as ANN with Double Hidden layers and Logistic Regression with Inverse of regulation do marginally better. Second, from the sensitivity, precision and f-score, we have plotted the results of all classification methods, by showing of course how all the methods have classified each cuisine point for the same set of input data, and we have seen that SVM with Normalized Polynomial Kernel is always a best classifier for the concerned problem. So to be completely sure of the choice of the best classifier to perfectly achieve the objectives pursued in this study, we also thought about the evaluation of the classification error shown in the above Table 4.6. From that table, we analyzed how the methods misclassified each cuisine point for the set of entries. As a result, all the cuisines have been classified by all methods. This implies that even though, we somehow choice SVM as a better classifier for this task, but we can't neglect its misclassification rate.
Thus, by using Normalized Polynomial Kernel (SVM), we presented the result of our experiments in the form of a confusion matrix where the element of row i and column j shows how the recipes of cuisine i are classified as being recipes of cuisine j.

Table .5: Evaluating classification error

Methods	**M**	**R**
Inverse of regulation (Logistic regression)	0.25	0.37
Normalized Polynomial Kernel (SVM)	0.15	0.34
Double Hidden layers ANN	0.27	0.50

M: Mean absolute error
R: Root means squared error

5. Conclusions

With the growing evolution of the food information sharing websites, the Internet has become one of the sources of access to food information from almost all countries. One of the major difficulties in using food data uploaded online is which of the cuisine recipes format; it varies proportionally from one source to another. For this specific study, a large number of cuisine recipe data were downloaded in various formats from a public website which discusses food data and cooking methods, and it has been analyzed the content of this data to obtain a uniform structure. Afterwards, it has been designed a classification model of collected data to obtain data format that allows the easy use of automatic classification methods.

This study compared three methods of machine learning that can be used to construct mathematical models to measure the similarity between different Asian cuisines based on their ingredients. To do this, it has been considered data on ingredients, recipes and cuisines from Asian countries, and it has been used logistic regression, support vector machine and neural network to evaluate the quality of our methods. In addition, our dataset was composed of a subset of the ingredients most commonly used by a subset of recipes in order to generate a simple description of similarities and differences between recipes belonging to different Asian cuisines.

Thus, after many experiments, it has been found that support vector machine with a normalized polynomial kernel is the best classifier because it gives a simple approach with clear semantics to describe, use and learn well-defined a boundary knowledge. Therefore, by applying this method to dataset, it has been achieved impressive results that help to determine a close relationship between the cuisines of Asian countries. For this, it has been discovered a classification with a simple support vector machine model by creating a simple prediction approach that perfectly illustrates a classifier. It has been justified the choice of this method by in-depth analysis of the evaluation parameters such as accuracy, precision, recall, f-score, etc. On this point, it should be recalled that the dataset was balanced and hence, it was not important for us to justify the choice of our classification method by Roc Curve. Finally, it has been used the confusion matrix to illustrate the sample of our results as well as the classification errors.

We hope that with the results of this paper, we have responded to various concerns about the relationship between cuisines from different countries, and as this study was limited in time and space, we considered only the case of five Asian countries. Nevertheless, this project can be extended to a large dataset taking into account, for example, information on foods from all countries.

References

1. Bähler, A., Smart Recipes. 2015.
2. Mitchell, R. and C.M. Hall, Consuming tourists: Food tourism consumer behaviour. Food tourism around the world: Development, management and markets, 2003: p. 60-80.
3. De Clercq, M., et al., Data-driven recipe completion using machine learning methods. Trends in Food Science & Technology, 2016. 49: p. 1-13.
4. Bianchini, D., V. De Antonellis, and M. Melchiori. A food recommendation system based on semantic annotations and reference prescriptions. in International Conference on Conceptual Modeling. 2015. Springer.
5. Su, H., et al. Automatic recipe cuisine classification by ingredients. in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. 2014. ACM.
6. Betoret, E., et al., Functional foods development: trends and technologies. Trends in Food Science & Technology, 2011. 22(9): p. 498-508.
7. Akhandaf, Y., et al., Exposure assessment within a Total Diet Study: A comparison of the use of the pan-European classification system FoodEx-1 with national food classification systems. Food and Chemical Toxicology, 2015. 78: p. 221-229.
8. Marano-Marcolini, C. and F.J. Torres-Ruiz, A consumer-oriented model for analysing the suitability of food classification systems. Food Policy, 2017. 69: p. 176-189.
9. Bimbo, F., et al., Consumers' acceptance and preferences for nutrition-modified and functional dairy products: A systematic review. Appetite, 2017.
10. Piccolo, D. and A. D'Elia, A new approach for modelling consumers' preferences. Food Quality and Preference, 2008. 19(3): p. 247-259.
11. Boyne, S., D. Hall, and F. Williams, Policy, support and promotion for food-related tourism initiatives: A marketing approach to regional development. Journal of Travel & Tourism Marketing, 2003. 14(3-4): p. 131-154.
12. van Pinxteren, Y., G. Geleijnse, and P. Kamsteeg. Deriving a recipe similarity measure for recommending healthful meals. in Proceedings of the 16th international conference on Intelligent user interfaces. 2011. ACM.
13. bigoven platform. Available from: https://www.crunchbase.com/organization/bigoven.
14. Macmillan, F. and M. Hons, Risk, uncertainty and investment decision-making in the upstream oil and gas industry. 2000, Aberdeen University.
15. Witten, I.H., et al., Data Mining: Practical machine learning tools and techniques. 2016: Morgan Kaufmann.

The first author is Assistant and Ph. Candidate in Computer Sciences at University of Kinshasa.

The second author is Assistant and Ph. Candidate in Computer Sciences at University of Kinshasa. With a publication: "Decision-Making System by Neuronal Networks for Taking into Care Children Suffering from Serious Paludism", IJCSI, volume 11, Issue 5, september 2014,

The third author is professor and head of Mathematic and informatics department of the University of Kinshasa. As publications, Author of many publications, such as: "Enhanced Parallel Skyline on multi-core architecture with lax Memory space Cost", IJCSI, volume 13, Issue 5, September 2016, Poster et the 2nd International conference on Big Data Analysis and Data Mining, San Antonio, USA, 30 November -30 December 01 2015; "Data Mart Approach for Stock Management Model with a calendar Uner Budgetary contraint" IJCSI, volume 15, Issue 5, September 2016,