International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

MINING OF NUTRITIONAL INGREDIENTS IN FOOD FOR DISEASE ANALYSIS

[1]B.Anoch , [2]Pandruvada Lakshmi Sravya, [3]Grandhi Yasaswini Nirmala, [4]Shaik Adavi Haleema,
[5] Dumpa Sai Divya
[1,2,3,4,,5]Assistant professors, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT

Suitable nutritional diets have been widely recognized as important measures to prevent and control non-communicable diseases (NCDs). However, there is little research on nutritional ingredients in food now, which are beneficial to the rehabilitation of NCDs. In this paper, we profoundly analyzed the relationship between nutritional ingredients and diseases by using data mining methods. First, more than 7,000 diseases were obtained and we collected the recommended food and taboo food for each disease. Then, referring to the China Food Nutrition, we used noise-intensity and information entropy to find out which nutritional ingredients can exert positive effects on diseases. Finally, we proposed an improved algorithm named CVNDA_Red based on rough sets to select the corresponding core ngredients from the positive nutritional ingredients. To the best of our knowledge, this is the first study to discuss the relationship between nutritional ingredients in food and diseases through data mining based on rough set theory in China. The experiments on real-life data show that our method based on data mining improves the performance compared with the traditional statistical approach, with the precision of 1.682. Additionally, for some common diseases such as Diabetes, Hypertension and Heart disease, our work is able to identify correctly the first two or three nutritional ingredients in food that can benefit the rehabilitation of those diseases. These experimental results demonstrate the effectiveness of applying data mining in selecting of nutritional ingredients in food for disease analysis.

1. INTRODUCTION

NCDS are chronic diseases, which are mainly caused by occupational and environmental factors, lifestyles and behaviors, including Obesity, Diabetes, Hypertension, Tumors and other diseases. According to the Global Status Report on Non-communicable Diseases issued by the WHO, the annual death toll from NCDs keeps adding up, which has caused serious economic burden to the world. About 40 million people died from NCDs each year, which is equivalent to 70% of the global death toll. Statistics of Chinese Resident's Chronic Disease and Nutrition shows that, the number of the patients suffering from NCDs in China is higher than the number in any other countries in the world, and the current prevalence rate has blown out. In addition, the population aged 60 or over in China has reached 230 million and about two-thirds of them are suffering from NCDs according to the official statistics. Therefore, relevant departments in each country, especially in China, such as medical colleges, hospitals and disease research centers all are concerned about NCDs. Suitable nutritional diets play an important role in maintaining health and preventing the occurrence of NCDs. With the gradual recognition of this concept, China has also repositioned the impact of food on health. However, research on nutritional ingredients in food via data mining, which are conducive to the rehabilitation of diseases is still rare in China. At present, China has just begun the IT (Information Technology) construction of smart health-care. Most studies on the relationship between nutritional ingredients in food and diseases are still through expensive precision instruments or long-term clinical trials. In addition, there are also many prevention reports, but they studied only one or several diseases. In China, studying the relationship between nutritional ingredients and diseases using data mining is immature. Most doctors only recommend the specific food to patients suffering from NCDs, without giving any relevant nutrition information, especially about nutritional ingredients in food. The solutions for NCDs require interdisciplinary knowledge. In the era of big data, data mining has become an essential way of discovering new knowledge in various fields, especially in disease prediction and accurate health-care (AHC). It has become a core support for preventive medicine, basic medicine and clinical medicine research. With respect to the disease analysis through the mining of nutritional ingredients in food, we mainly make the following contributions: (i) We extracted data related to Chinese diseases, corresponding recommended food and taboo food for each disease as many as possible from medical and official websites to create a valuable knowledge base that are available online; (ii) Applying noise-intensity and information entropy to find out which nutritional ingredients in food can exert positive effects to diseases; (iii) In this paper, the data is continuous and has no decision attributes. To address this problem, we proposed an improved algorithm named CVNDA_Red based on rough set theory, which can better select corresponding core ingredients from the positive nutritional ingredients in food. The structure of this paper is organized as follows: Section II reviews the related work in the field of disease analysis and data mining. Describes the specific data mining algorithms used in this paper, reasons why we select the algorithms, as well as two evaluation indexes. Elaborates the data, experimental results and analysis in detail. Presents discussions between methods. Some conclusions and potential future research directions are also discussed.

2. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from books or from websites. Before building the system the above considerations are taken into account for developing the proposed system. A Retrospective Analysis of Hypertension Screening at a Mass Gathering in India:

Implications for Non-communicable Disease Control Strategies Cardiovascular disease is the leading case of mortality from non-communicable diseases (NCD) in India. The government's National Programme for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases and Stroke seeks to increase capacity building, screening, referral and management of NCDs across India, and includes community-based outreach and screening programmers. The government in India routinely provides basic care at religious mass gatherings. However, in 2015, at the Kumbh Mela in Nashik and Trimbakeshwar, the state government extended its services to include a hypertension screening programmer. We examine here the value and implications of such opportunistic screening at mass gatherings. At the Kumbh, 5760 persons voluntarily opted for hypertension screening, and received a single blood pressure measurement. In all, 1783 (33.6%) screened positive, of whom, 1580 were previously unaware of their diagnosis. Of the 303 that had previously known hypertension, 240 (79%) were prescribed medications, and 160 were compliant (that is, 52.8% under treatment). Fifty-five (18%) had normal blood pressure readings (BP under control). The data also demonstrated higher prevalence (39%) of hypertension among tobacco users compared to non-users (28%) (P<0.001). Poor recording of phone numbers (0.01%) precluded any phone-based follow-up. The low rates of hypertension awareness, treatment and control underscore the ongoing challenge of both hypertension screening and management in India.

3. SYSTEM ANALYSIS

3.1 Existing system:

Existing in all recommended food (high stability), otherwise these recommended food have no reason to be recommended. Conversely, if these nutritional ingredients are not PNIs for that specified disease, they may or may not exist in different recommended food (poor stability). The SA just considers the level of nutritional ingredient values to determine simply whether they are PNIs or not.

Disadvantages:

1. Performance is less.
2. Most cost effective.

3.2 Proposed system:

Proposed that we can recommend food according to the body’s Creatinine values. However, the above studies are basically carried out through long-term clinical trials, which just recommend food for certain specific diseases and they seldom study the relationship between nutritional ingredients and diseases by data mining techniques.

Advantages:

1. Performance is improved.
2. Less cost effective

4. ALGORITHM

4.1 Linear Regression

Linear regression may be defined as the statistical model that analyzes the linear relationship between a dependent variable with given set of independent variables.

Linear relationship between variables means that when the value of one or more independent variables will change (increase or decrease), the value of dependent variable will also change accordingly (increase or decrease).

Mathematically the relationship can be represented with the help of following equation −

Y=mX+bY=mX+b

Here, Y is the dependent variable we are trying to predict.

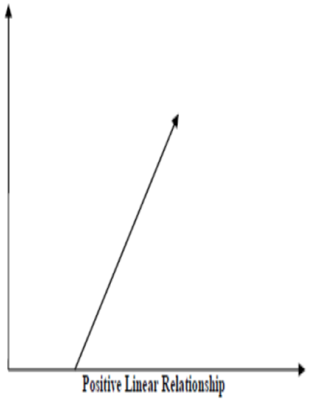
X is the dependent variable we are using to make predictions.

m is the slop of the regression line which represents the effect X has on Y

b is a constant, known as the *Y*Y-intercept. If X = 0,Y would be equal to *b*b.

Positive Linear Relationship

A linear relationship will be called positive if both independent and dependent variable increases. It can be understood with the help of following graph



Positive Linear Relationship

4.2 K-nearestneighbor’s algorithm (k-NN)

k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

● In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

● In k-NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. Both for classification and regression, a useful technique can be to assign weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/d, where d is the distance to the neighbor. The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required.

4.3 Working of KNN Algorithm

K-nearest neighbors (KNN) algorithm uses ‘feature similarity’ to predict the values of new datapoints which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. We can understand its working with the help of following steps –

Step 1 − For implementing any algorithm, we need dataset. So during the first step of KNN,we must load the training as well as test data.

Step 2 − Next, we need to choose the value of K i.e. the nearest data points. K can be any integer.

Step 3 − For each point in the test data do the following −

● 3.1 − Calculate the distance between test data and each row of training data with the help of any of the method namely: Euclidean, Manhattan or Hamming distance. The most commonly used method to calculate distance is Euclidean.

● 3.2 − Now, based on the distance value, sort them in ascending order.

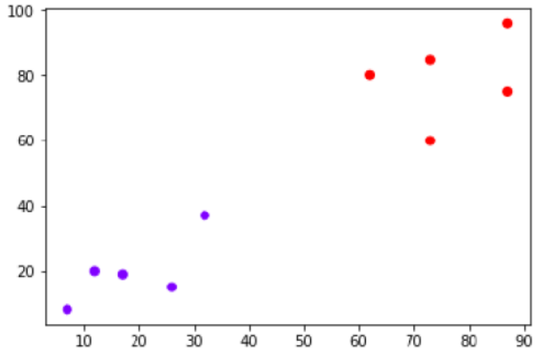
● 3.3 − Next, it will choose the top K rows from the sorted array.

● 3.4 − Now, it will assign a class to the test point based on most frequent class of these rows.

Step 4 − End

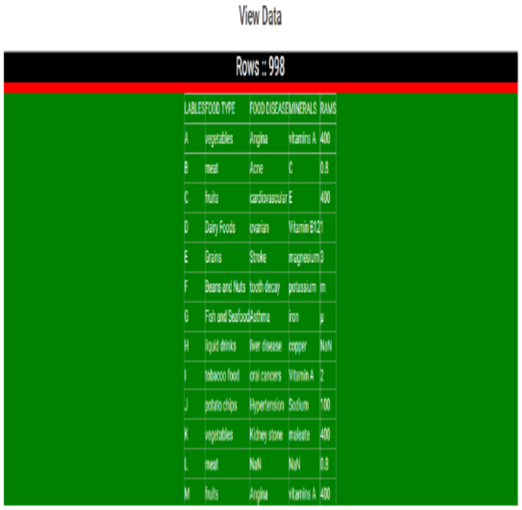
Example

The following is an example to understand the concept of K and working of KNN algorithm Suppose we have a dataset which can be plotted as follows –



Now, we need to classify new data point with black dot (at point 60,60) into blue or red class. We are assuming K = 3 i.e. it would find three nearest data points. It is shown in the

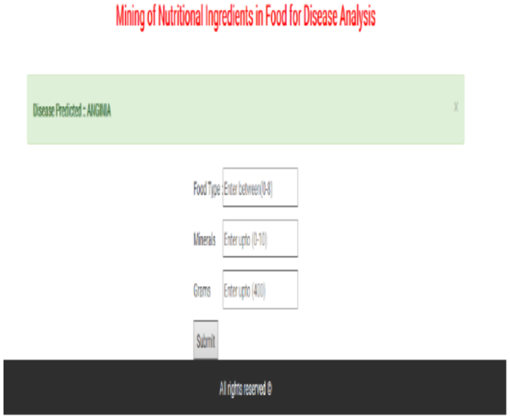
1. Results:



View Data

Rows :: 998

LABLES	FOOD TYPE	FOOD DISEASE	MINERALS	RAMS
A	vegetables	Angina	vitamins A	400
B	meat	Acne	C	0.8
C	fruits	cardiovascular	E	400
D	Dairy Foods	ovarian	Vitamin B12	1
E	Grains	Stroke	magnesium	3
F	Beans and Nuts	tooth decay	potassium	m
G	Fish and Seafood	Asthma	iron	µ
H	liquid drinks	liver disease	copper	NaN
I	tobacco food	oral cancers	Vitamin A	2
J	potato chips	Hypertension	Sodium	100
K	vegetables	Kidney stone	maleate	400
L	meat	NaN	NaN	0.8
M	fruits	Angina	vitamins A	400



Mining of Nutritional Ingredients in Food for Disease Analysis

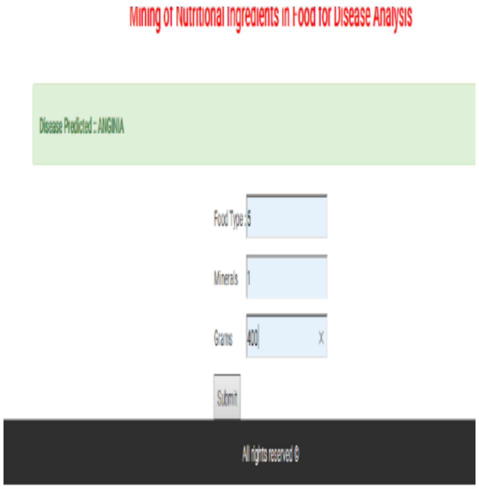
Disease Predicted :: ANGINIA

Food Type : Enter between(0-8)

Minerals Enter upto (0-10)

Grams Enter upto (400)

All rights reserved ©



Mining of Nutritional Ingredients in Food for Disease Analysis

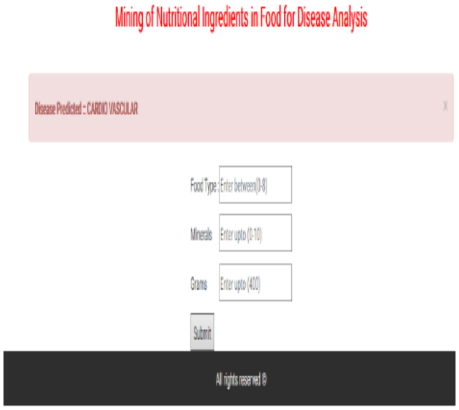
Disease Predicted :: ANGINIA

Food Type : 5

Minerals 1

Grams 400

All rights reserved ©



Mining of Nutritional Ingredients in Food for Disease Analysis

Disease Predicted :: CARDIO VASCULAR

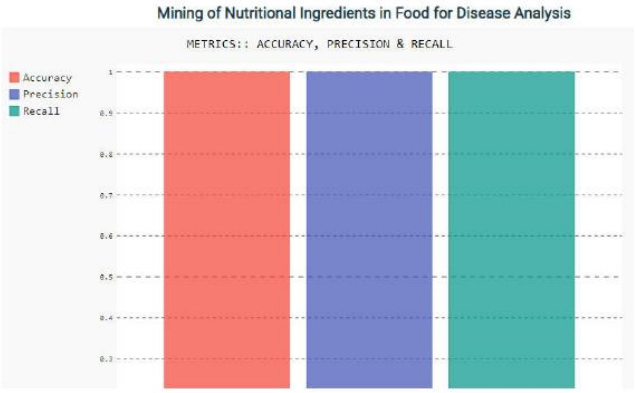
Food Type : Enter between(0-8)

Minerals Enter upto (0-10)

Grams Enter upto (400)

All rights reserved ©

Graph:



5. CONCLUSION

In conclusion, the project concentrates on the development of a system for student performance nalysis. A data mining technique, classification algorithm is applied in this project to ensure the prediction of the student performance in course "TMC1013 System Analysis and Design" is possible. The main contribution of the SPAS is that it assists the lecturers in conducting student performance analysis. The system assists lecturers in identifying the students' that are predicted to fail in the course TMC1013 System Analysis and Design". Other than that, SPAS assists lecturers' to retrieve information of their students' performance throughout the semesters. The main work of this paper can be divided into two parts: firstly, we obtained and sorted out more than seven thousand Chinese diseases and orresponding recommended and taboo food from medical and official websites; secondly, we discussed the relationship between nutritional ingredients and diseases, which mainly aims to find out which ingredients play a positive role in the rehabilitation of diseases. To the best of our knowledge, this is the first study in China, which mines the relationship between nutritional ingredients in food and diseases by using data mining technology. Experimental results showed that although we could not completely find all the positive nutritional ingredients for diseases by data mining methods, the first two or three ones were selected accurately. In addition, if our perspective can be combined with taboo food, the results would be likely to be better and in line with reality, which will be our future work direction. There are two main benefits of this work: 1) You can access to our website2 to get diseases, ecommended food, taboo food and corresponding nutrition information involved in this paper; 2) We can assist doctors and disease researchers to find out positive nutritional ingredients that are conducive to the rehabilitation of the diseases as accurately as possible. At present, some data is not available, because they are still in the medical verification. Besides, our knowledge base is still gradually mproving, if researchers find out something incorrect in our work, we hope to contact us and make our research improved.

REFERENCES

[1] CNS, “2016 Global Nutrition Report,” in Chinese Nutrition Society, 2016.

[2] WHO, “Global Status Report on Noncommunicable Diseases (2014),” in World Health Organization, 2014.

[3] S. Balsari, P. Vemulapalli, M. Gofine et al., “A Retrospective Analysis of Hypertension Screening at a Mass Gathering in India: Implications for Non-communicable Disease Control Strategies,” Journal of Human Hypertension, vol. 31, no. 11, pp. 750–753, 2017.

[4] DNHFPC of PRC, “Chinese ResidentâA˘ Zs Chronic Disease and Nutrition ´(2015),” in National Health and Family Planning Commission of the People’s Republic of China, 2015.

[5] S. Tellier, A. KiabyLars, P. Nissen et al., “Basic Concepts and Current Challenges of Public Health in Humanitarian Action,” International Humanitarian Action, pp. 229–317, 2017.

[6] F. Ara1, F. Saleh, S. J. Mumu, F. Afnan and L. Ali, “Awareness Among Bangladeshi Type 2 Diabetic Subjects Regarding Diabetes and Risk Factors of Non-communicable Diseases,” Diabetologia, pp. S379, 2011. DOI:10.1007/s00125-011-2276-4.

[7] QIANZHAN, “Report of Market Prospective and Investment Strategy Planning on China Intelligent medical construction industry (2017-2022),” in Qianzhan Intelligence CO.LTD, 2017.

[8] W. H. Ling, “Progress of Nutritional Prevention and Control on Noncommunicable Chronic Diseases in China,” China J Dis Control Prev, vol. 21, no. 3, pp. 215–218, 2017.

[9] M. B. Margaret, B. K. Barbara and D. Colette, “Developing Health Promotion Workforce Capacity for Addressing Non-communicable Diseases Globally,” Global Handbook on Noncommunicable Diseases and Health Promotion, pp. 417–439, 2013.

[10] M. Williams and H. Moore, “Lumping Versus Splitting: the Need for Biological Data Mining in Precision Medicine,” BioData Mining, vol. 8, no. 16, pp. 1–3, 2015.

[11] G. M. Oppenheimer, “Framingham Heart Study: The First 20 Years,”Progress in Cardiovascular Diseases, vol. 53, no. 1, pp. 55–61, 2010.

[12] W. Y. Jiao, Y. Xue, T. C. He, Y. M. Zhang and P. Y. Wang, “Association Between South Korean Dietary Pattern and Health,” Food and Nutrition in China, vol. 23, no. 5, pp. 81–84, 2017.

[13] K. W. Lee and M. S. Cho, “The Traditional Korean Dietary Pattern Is Associated with Decreased Risk of Metabolic Syndrome:Findings from the Korean National Health and Nutrition Examination Survey 1998-2009,” Journal of Medicinal Food, vol. 17, no. 1, pp. 43–56, 2014.