IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 4, Issue 6, June 2024

Impact Factor: 5.731

Health Analysis and Recommendation Based on Food Using Data Mining

Dr. Krithika. D. R[1], Dr. A. Poongodi[2], T. Swathi[3]

Assistant Professor, Department of Computer Applications[1,2]

PG Student, Department of Computer Applications[3]

Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India

krithika.scs@velsuniv.ac.in and apoongodi.scs@velsuniv.ac.in

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 project, we profoundly analyzed the relationship between nutritional ingredients and diseases by using data mining methods. First, more than n number of diseases was obtained, and we collected the recommended food and taboo food for each disease. The experiments on reallife data show that our method based on data mining improves the performance compared with the traditional statistical approach. 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. The dataset uploaded will be pre-processed, Feature Extraction, noisy data will be removed, and classification of dataset will take places using random forest algorithm based on this analysis the diseases prediction takes places for the food intake by the individual*

Keywords: Data mining, non-communicable diseases NCD, Random Forest, Machine learning

I. INTRODUCTION

Disease has become the primary reason of death in developed countries. The most effective way to reduce liver disease deaths is to detect it earlier. Early diagnosis needs an accurate and reliable diagnosis procedure that can be used by physicians to distinguish. In this project data mining methods that have been applied in disease data set to know people who are affected by disease and problems are surveyed. We will represent the proper analysis of data mining techniques applied on data sets. i.e., accuracy of those data set. With this project, we will implement data mining algorithms to classify the given data to determine which algorithm is best suitable to classify it and to find out whether the person is having the disease or not. Data Mining also inspires many practical solutions that maximize resource utilization and prolong the lifespan of the system.

II. LITERATURE REVIEW

Tapas Ranjan Baitharu, delivered the healthcare industry collects a huge amount of data which is not properly mined and not put to the optimum use. Discovery of these hidden patterns and relationships often goes unexploited. Our research focuses on this aspect of Medical diagnosis by learning pattern through the collected data of Liver disorder to develop intelligent medical decision support systems to help the physicians. In this paper, we propose the use decision trees J48, Naive Bayes, ANN, ZeroR, 1BK and VFI algorithm to classify these diseases and compare the effectiveness, correction rate among them. Detection of Liver disease in its early stage is the key of its cure. It leads to better performance of the classification models in terms of their predictive or descriptive accuracy, diminishing of computing time needed to build models as they learn faster, and better understanding of the models. In this paper, a comparative analysis of data classification accuracy using Liver disorder data in different scenarios is presented. The predictive performances of popular classifiers are compared quantitatively[1]. Matthew Herland discussed about

amount of data produced within Health Informatics has grown to be quite vast, and analysis of this Big Data grants potentially limitless possibilities for knowledge to be gained. In addition, this information can improve the quality of healthcare offered to patients. However, there are a number of issues that arise when dealing with these vast quantities of data, especially how to analyze this data in a reliable manner. The basic goal of Health Informatics is to take in real world medical data from all levels of human existence to help advance our understanding of medicine and medical practice. This paper will present recent research using Big Data tools and approaches for the analysis of Health Informatics data gathered at multiple levels, including the molecular, tissue, patient, and population levels. In addition to gathering data at multiple levels, multiple levels of questions are addressed: human-scale biology, clinical-scale, and epidemic-scale. We will also analyze and examine possible future work for each of these areas, as well as how combining data from each level may provide the most promising approach to gain the most knowledge in Health Informatics[2]. Panel Jayasri N.P discusses about Big data is the compilation of enormous data that arrives from diverse sources for instance online transaction details, social media, sensor data, etc. Such assortment of enormous data developed is tough to evaluate by conventional processing relevance's. By the development and upcoming latent in the healthcare business field, it is essential to analyze an enormous noisy data to get significant information. In healthcare system, the aim of this work is to evaluate the medical database of diabetes patients by a mixture of innovative hierarchical decision attention network, association rules (AR) and multiclass outlier classification with MapReduce framework. The association rule apriori algorithm in a MapReduce framework considers health data to create regulations. This is employed to discover the association among disease and their signs. This examination is made by means of UCI machine learning datasets of diabetes containing 50 attributes. The results of the proposed algorithm are offered by parameters for instance precision, accuracy, recall, and F-score[3]. : Matthew Herland discuss about this information can improve the quality of healthcare offered to patients. However, there are a number of issues that arise when dealing with these vast quntities of data, especially how to analyze this data in a reliable manner. The basic goal of Health Informatics is to take in real world medical data from all levels of human existence to help advance our understanding of medicine and medical practice. This paper will present recent research using Big Data tools and approache s for the analysis of Health Informatics data gathered at multiple levels, including the molecular,tissue, patient, and population levels. In addition to gathering data at multiple levels, multiple levels of questions are addressed: human-scale biology, clinical-scale, and epidemic-scale. We will also analyze and examine possible future work for each of these areas, as well as how combining data from each level may provide the most promising approach to gain the most knowledge in Health Informatics [4]. Maikel Luis Kolling discussed about a bibliometric performance and network analysis (BPNA) was conducted. For this purpose, 6138 articles were sourced from the Web of Science covering the period from 1995 to July 2020 and the SciMAT software was used. Our results present a strategic diagram composed of 19 themesare depicted in a thematic network. An in-depth analysis was carried out in order to find hidden patterns and to provide a general perspective of the field. The thematic network structure is arranged thusly that its subjects are organized into two different areas, (i) practices and techniques related to data mining in healthcare, and (ii) health concepts and disease supp orted by data mining, embodying, respectively, the hotspots related to the data mining and medical scopes, hence demonstrating the field's evolution over time. Such results make it possible to form the basis for future research and facilitate decision -making by researchers and practitioners, institutions, and governments interested in data mining in mining in healthcare [5].

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 4, Issue 6, June 2024
Impact Factor: 5.731

III. METHODOLOGY

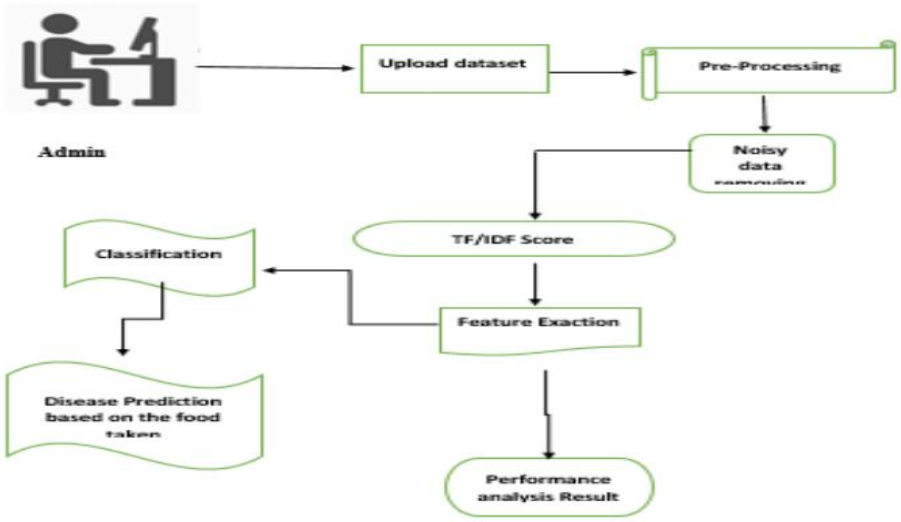

Fig 1 Architecture

This paper to develop a system that can analyse individual's health based on their food intake and provide recommendations for a healthier diet plan. The system will use data mining techniques to analyses the nutritional value of food consumed by the user and identify patterns that may lead to unhealthy eating habits. The system will then provide personalized recommendations based on the user's dietary requirements and health goals.

Objectives:

- To develop a system that can analyse individual's health based on their food intake.
- To identify patterns in the user's dietary habits that may lead to unhealthy eating habits.
- To provide personalized recommendations based on the user's dietary requirements and health goals.
- To improve the overall health and wellbeing of the users by promoting healthier eating habits

IV. DATA AND IMPLIMENTATION

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

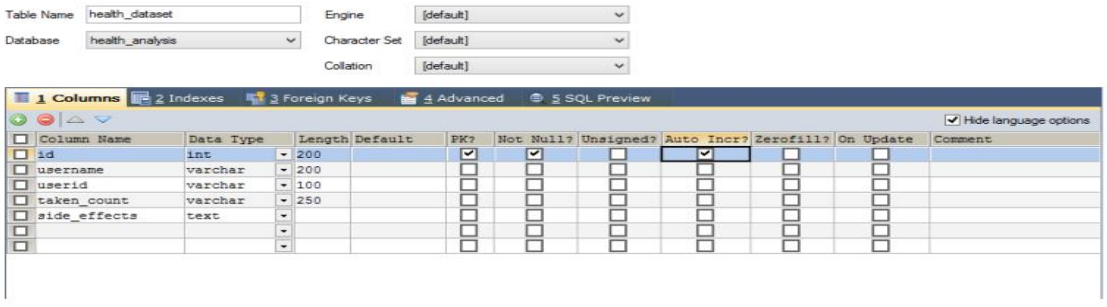

Fig.2 Data set

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 4, Issue 6, June 2024
Impact Factor: 5.731

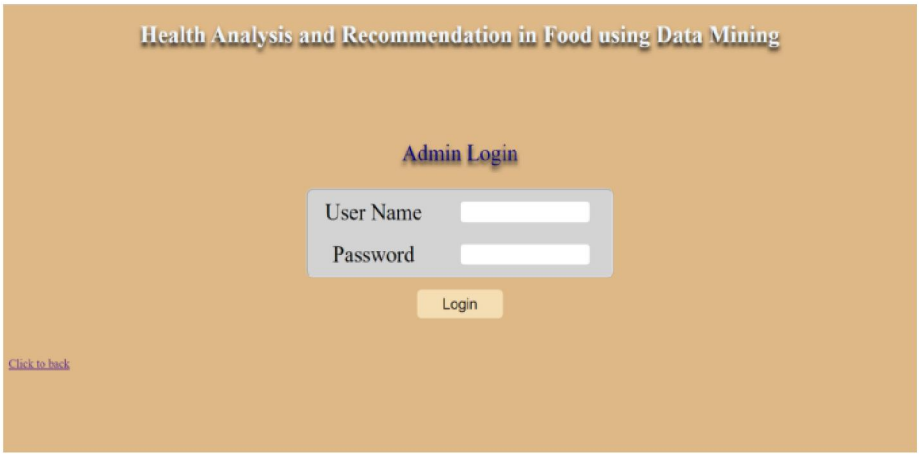

Fig 3 Login page

Fig 4 Upload page

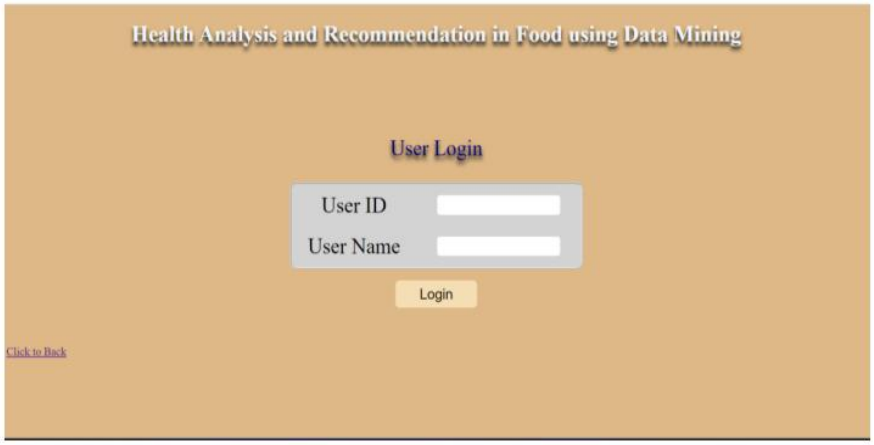

Fig. 5 User registration page

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 4, Issue 6, June 2024
Impact Factor: 5.731

ADMIN PROCESS STEPS:

1. The admin has to Login into the account with the correct username and the password.
2. Upload Data set:
3. In this project we upload two data sets. One is contains the food related fields like food type, food name, Number of counts taken, Severity (%). Second one is disease related which contains disease data (disease name, type of affect, severity).
4. View Data set
5. Pre-processing:
6. After view all data click for the pre-processing in that we have remove the all noises and the unwanted things.
7. Feature Extraction.
8. In that feature extraction we have extract the all information in the dataset with the % value
9. TF/IDF Score
10. We have view the all foods based on the diseases find the TF/IDF Calculation
11. Classification
12. In that final Classification admin can recommend to the users that is view in the users dashboard
13. Performance Measure By using Confusion Matrix.
14. Performance analysis is based on the confusion matrix
15. Graph ` 154 54
16. The overall result is shown here
17. Log-out

USER PROCESS STEPS:

Login the account with the correct user name and password.

Dashboard:

2. View all the admin Recommendation by the admin
3. Logout

V. DISCUSSION

Existing Method: Generally, all the health care organizations across the world stored the health care data in electronic format. Health care data mainly contains all the information regarding patients as well as the parties involved in health care industries. The storage of such type of data is increased at a very rapidly rate. Due to continuous increasing the size of electronic health care data a type of complexity is exist in it. In other words, we can say that health care data becomes very complex. By using the traditional methods it becomes very difficult in order to extract the meaningful information from it. But due to advancement in field of statistics, mathematics and very other disciplines it is now possible to extract the meaningful patterns from it.

Proposed Method : The health care data is very useful in order to extract the meaningful information from it for improving the health care services for the patients. • The quality of data depends on various factors such as removal of noisy data, free from missing of data etc. • We measure the performance measure by using the confusion matrix. For any algorithm its accuracy and performance are of greater importance. • But due to presence of some factors any algorithm can greatly lost the abovementioned property of accuracy and performance and recommended the food to users.

ADVANTAGES OF PROPOSED SYSTEM: User know about the what types food they taken. Easily find out the diseases of the food. Its user friendly. The performance and the accuracy of the system is high.

VI. CONCLUSION

Good nutritional habits and a balanced diet aren't developed in one day, nor are they destroyed in one unbalanced meal. Eating nutritious and healthful food while maintaining your proper body weight will contribute to a better performance in the classroom, in the gym.

REFERENCES

[1]. .http://www.mayoclinic.org/diseases-conditions/heart disease/basics/de nition/con20034056.
[2]. http://www.idf.org/aboutdiabetes.
[3]. http://www.cancer.org/cancer/breastcancer/detail edguide/breast-cancer-what-is-breast-cancer.Kumara, M., Vohra, R., Arora, A. (2014).
[4]. Prediction of diabetes using Bayesian network. International Journal of Computer ` 104 104 Science and Information Technologies, 5(4), 5174-5178.Thirumal, P. C., & Nagarajan, N (2015).
[5]. Utilization of data mining techniques for diagnosis of diabetes mel-litus - A case study. ARPN Journal of Engineering and Applied Sciences, January, 10(1), 8-13. Gomathi, K. (2012).
[6]. An empirical study on breast can-cer using data mining techniques. International Journal of Research in Computer Application & Management, July, 2(7), 97-102.Witten, H. I., & Frank, E. (2005).
[7]. Data Mining: Practical Machine Learning Tools and Techniques, (2nd Ed), Morgan Kaufmann Publishers. Witten, I. H. & Frank, E. (2005).