P.Abinaya, 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Construct Food Safety Traceability System for People Health Using the Internet of Things and Big Data

P.Abinaya, Dr. S.Prasanna
Department of Computer Applications PG
Vels Institute of Science Technology and Advanced Studies

Abstract- It is well known that eating a healthy diet is crucial to preventing and managing non-communicable diseases (NCDs). The nutritional components of food that aid in the rehabilitation of NCDs, however, are currently the subject of scant research. Using data mining techniques, we conducted a thorough analysis of the connection between diseases and dietary components in this study. First, we gathered the banned and suggested foods for each of the more than n ailments that were obtained. Experiments on actual data demonstrate that our data mining-based methodology outperforms the conventional statistical method. As precisely as possible, we can help physicians and disease researchers identify beneficial dietary components that support the recovery of various illnesses. Some data are currently unavailable as they are still undergoing medical verification. The uploaded dataset will be pre-processed, feature extracted, noisy data eliminated, and classified using the random forest method. Based on this analysis, the individual's food intake will be used to predict their disease.

Keywords- non-communicable diseases, healthy diet, dietary components, data mining, disease prediction, nutritional analysis, food intake, Random Forest, feature extraction, data preprocessing, noise elimination, statistical comparison, medical verification.

I. INTRODUCTION

Diseases have been the leading cause of death for people in wealthy nations in recent years. Due to its sometimes silent course and the scarcity of effective treatments at advanced stages, liver disease has emerged as one of the most alarming of these. Early detection is one of the most effective ways to lower liver disease-related mortality since it can greatly enhance treatment results and survival rates. Early detection is still difficult, though, and calls for precise, effective, and scalable diagnostic instruments that doctors can use to distinguish between healthy people and those exhibiting early disease symptoms.

Data mining has become a potent strategy in the medical industry to address this issue. Analyzing vast and intricate information to find significant patterns, trends, and connections that conventional statistical techniques can miss is known as data mining. Our project's main goal is to investigate and assess different data mining approaches used on liver disease-related medical datasets. By doing this, we want to find people who are either already impacted by the disorder or at risk.

In order to evaluate the effectiveness and precision of various popular classification algorithms using actual liver disease datasets, our work involves a survey and implementation of these algorithms. Finding the optimal algorithm for disease categorization tasks is the aim. Performance measurements including accuracy, precision, recall, and F1-score will be used to test and compare methods like Random Forest, Support Vector

Machines (SVM), Decision Trees, and Naive Bayes.Data mining contributes more broadly to the enhancement of healthcare systems than only classification. Better resource utilization, data-driven decision-making, and the creation of intelligent systems that can help medical practitioners are all made possible by it. We hope to improve disease prediction and optimize healthcare services by utilizing these technologies, which will ultimately improve patient outcomes and prolong the life of diagnostic devices.

II. LITERATURE REVIEW

Title: Examining Data Mining Methods with Liver Disorder Dataset for Healthcare Decision Support Systems.
Tapas Ranjan Baitharu is the author.

Abstract:
As one of the main causes of death for both men and women, liver disease has emerged as one of the world's most significant health issues. Due to a dearth of efficient data mining methods, a large portion of the enormous volume of clinical and healthcare data that is produced daily is still underutilized. Early detection is essential when it comes to liver problems because prompt diagnosis can greatly increase the likelihood of recovery and successful treatment. Using a dataset of liver disorders, we test the effectiveness of several classification algorithms, such as Decision Trees (J48), Naive Bayes, Artificial Neural Networks (ANN), ZeroR, 1-Nearest Neighbor (1BK), and Voting Feature Intervals (VFI), in categorizing patients according to their medical characteristics. The predicted accuracy, correction rate, and computing efficiency of each method are evaluated. Our goal is to compare these models in order to ascertain which method is best for precisely detecting liver disease in its early stages. The research's conclusions not only advance healthcare diagnostics but also emphasize how crucial it is to mine massive medical data sets for useful insights. This study demonstrates how machine learning and data mining approaches can revolutionize healthcare decision-making, streamline clinical procedures, and eventually enhance patient outcomes.[1].

Title: An overview of large data data mining in health informatics.
Matthew Herland is the author.

Abstract:
The volume of data generated in the field of health informatics has increased significantly, and analyzing this Big Data offers almost endless opportunities for knowledge acquisition. Furthermore, this knowledge can raise the standard of care that patients get. But dealing with these enormous amounts of data raises a variety of problems, particularly with regard to trustworthy data analysis. The fundamental objective of health informatics is to improve our knowledge of medicine and medical practice by incorporating real-world medical data from all spheres of human existence. This article will discuss current studies that analyze health informatics data collected at many levels, such as the molecular, tissue, patient, and population levels, utilizing Big Data tools and methodologies. Human-scale biology, clinical-scale, and epidemic-scale concerns are addressed in addition to collecting data at many levels. We'll also look at potential future research in each of these categories and how integrating data from all levels could be the most effective way to learn as much as possible about health informatics.[2].

Title: Using data mining and classification approaches for big data analytics in healthcare.
The panel's author Jayasri, N.P.

Abstract:
Big data is the accumulation of vast amounts of data from various sources, such as social media, sensor data, online transaction information, etc. It is difficult to assess the relevance of such a vast array of data by traditional processing methods. Given the current and future developments in the healthcare industry, it is imperative to evaluate

massive amounts of noisy data in order to extract meaningful information. The goal of this work is to assess the medical database of diabetic patients in the healthcare system using a combination of multiclass outlier classification using the MapReduce framework, association rules (AR), and a unique hierarchical decision attention network. In a MapReduce framework, the association rule apriori algorithm takes health data into account when generating rules. This is used to determine how diseases and their symptoms are related. UCI machine learning datasets of diabetes with 50 attributes are used for this analysis. Parameters like as precision, accuracy, recall, and F-score provide the outcomes of the suggested approach.[3].

Title: Clinical big data data mining: commonly used databases, procedures, and methodological models. Wen-Tao Wu is the author.

Abstract:
Public databases like the Surveillance, Epidemiology, and End Results (SEER), National Health and Nutrition Examination Survey (NHANES), The Cancer Genome Atlas (TCGA), and Medical Information Mart for Intensive Care (MIMIC) have produced a large number of high-quality studies. However, these data are frequently characterized by a high degree of dimensional heterogeneity, timeliness, scarcity, irregularity, and other characteristics, which prevents the full value of these data from being fully utilized. Since data-mining technology performs very well in assessing patient risks and supporting clinical decision-making in the development of disease-prediction models, it has been a cutting-edge area in medical research. As a result, data mining offers distinct benefits for big data research, particularly in extensive public medical databases. In addition to introducing the primary medical public database, this paper provided a straightforward explanation of the procedures, assignments, and models involved in data mining. We also discussed data-mining techniques and real-world uses for them. In order to facilitate the generation of research findings that benefit physicians and patients, our work aimed to help clinical researchers develop a clear and intuitive knowledge of the application of data-mining technology on clinical big-data.[4].
TITLE: Using Strategic Intelligence Techniques to Illustrate 25 Years of Research Development in Healthcare through Data Mining Maikel Luis Kolling is the author.

Abstract:
In this paper, a bibliometric performance and network analysis (BPNA) was performed to determine the strategic issues and the thematic progression structure of data mining applied to healthcare. Using the SciMAT the software, 6138 publications from the Web of Science spanning the years 1995–July 2020 were sourced for this purpose. The eight motor themes—"NEURAL-NETWORKS," "CANCER," "ELECTRONIC-HEALTH-RECORDS," "DIABETES-MELLITUS," "ALZHEIMER'S DISEASE," "BREASTCANCER," "DEPRESSION," and "RANDOM-FOREST"—are represented in a thematic network of the strategic diagram of 19 themes that our results present. To uncover hidden trends and offer a broad viewpoint, a thorough investigation was conducted. To uncover hidden patterns and give a broad overview of the topic, a thorough investigation was conducted. The thematic network structure is set up so that its subjects are divided into two categories: (i) healthcare data mining practices and techniques, and (ii) data mining-supported health concepts and diseases. These categories represent the hotspots associated with the medical and data mining domains, respectively, and show how the field has changed over time. These findings enable academics, practitioners, organizations, and governments interested in data mining in healthcare to make decisions more easily and serve as the foundation for future studies.[5].

III. IMPLEMENTATION

The Admin Module is essential to the management and oversight of the entire "Construct Food Safety Traceability System for People Health Using the Internet of Things (IoT) and Big Data" project. The purpose of this module is to give the administrator complete control over the traceability processes. The admin dashboard has a number of useful options, including Home, Upload Dataset, View Dataset, Retrieve Dataset Type, T-View, Classification, Performance Analysis, User Feedback, Graph, and Logout, as can be seen in the UI.

Fig 1: Admin Login

With the help of these options, the administrator can upload and view food-related datasets, retrieve and analyze particular data types, run classification algorithms to find anomalies or guarantee quality standards, and evaluate system performance metrics using graphical representations and analytical tools. Furthermore, user feedback can be examined to enhance the accuracy and usability of the system. The "Welcome Admin" greeting ensures a secure and customized administrative environment by confirming successful login and system access. Food safety and public health depend on the module's provision for real-time monitoring and data processing.

Fig 2: Admin-Based Food Recommendation and Risk Monitoring System

The Construct Food Safety Traceability System for People Health Using the Internet of Things (IoT) and Big Data project's"View Admin Recommendation Food" module is made to offer individualized dietary advice based on user health information and food safety analysis. The administrator has access to individual user records in this module, which include information about the user's identity, risk level (e.g., Medium), severity or impact on the body (e.g., a numerical risk score), and dietary recommendations. The module advises foods that are excellent for you (e.g., "Broccoli provides a good amount of fiber, potassium, folate, and phytochemicals") and foods that you should avoid (e.g., "Do not use home canned foods").Data analytics and IoT-gathered health indicators are probably the foundation of this advice. Additionally, it shows the amount of food consumed, which could reveal how much of the suggested food the user has eaten. By allowing users to rate and comment on the recommendations, the system uses big data approaches to improve subsequent ideas. All things considered, this module improves food traceability and public health safety by providing users with well-informed, data-driven food recommendations that are tracked and overseen by the administrator.

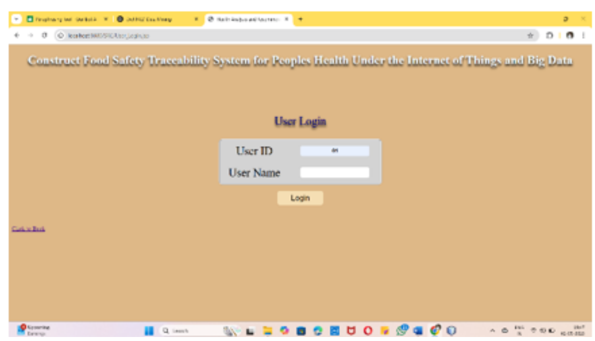

Fig 3: User Login

The Construct Food Safety Traceability System for People's Health's Admin Login Module The safe gateway for system administrators is part of the Internet of Things and Big Data initiative. This module makes sure that only individuals with permission can use the system's backend features.

The interface indicates that a thereusername and password are needed for authentication. The administrator has access to a number of system functions, including dataset administration, food safety analysis, user monitoring, and performance evaluation, once the proper login credentials have been entered. The integrity and confidentiality of data processed through the Internet of Things and big data platforms are ensured by this login method, which is essential for safeguarding sensitive health and food-related data. This module's existence aids responsible administration control and strengthens system security.

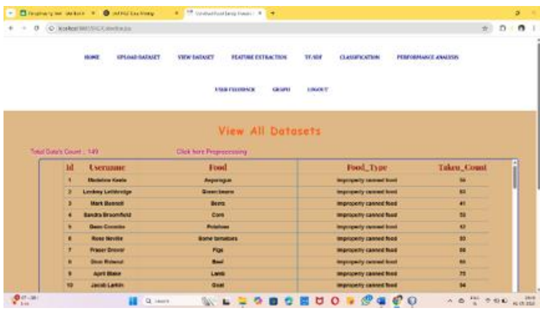

Fig 4: Dataset Management and Food Consumption Monitoring Module

The Construct Food Safety Traceability System for People's Health's View All Datasets Module The administrator can access and manage all gathered data entries pertaining to food consumption through an organized interface made possible by the Internet of Things and Big Data project. Username, Food Item, Food Type, and Taken Count (i.e., the amount or frequency of consumption) are among the detailed details displayed in this module. It facilitates real-time tracking of eating habits and food safety issues, like consuming "improperly canned food," which might be harmful to one's health. While the preprocessing option allows the administrator to get data ready for additional analysis, including categorization or risk assessment, the total data count at the top provides a brief summary of the dataset volume. Finding the cause of health problems, identifying foods that pose a risk to consumers, and facilitating data-driven decision-making are all made possible by this module. It is a prime example of how the system combines big data and the Internet of Things to improve accountability and transparency in the food supply chain.

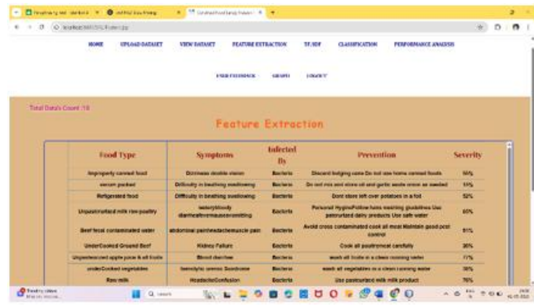

Fig 5: Food Safety Feature Extraction and Risk Analysis Module

A crucial part of the Construct Food Safety Traceability System for People's Health Using the Internet of Things and Big Data project is the Feature Extraction Module, which examines food-related data to find important characteristics pertinent to public health and food safety. Essential features including food type, related symptoms, "infected by" (bacteria or toxins, for example), prevention techniques, and severity levels (represented as a percentage) are extracted in this module. The module facilitates informed decision-making by emphasizing the connections between particular food items and the health hazards they bring, such as raw milk being connected to bacterial diseases or inadequately canned food producing botulism. For subsequent steps like classification, risk analysis, and producing preventive recommendations, the collected features are essential. Through the use of big data analytics and Internet of Things inputs, it enables the system to better comprehend the trends and risks present in the food data, ultimately facilitating proactive interventions and improving food safety.

The Construct Food Safety Traceability System's Fitness & Exercise Module for Human Health The project uses the Internet of Things and Big Data as a visual analytics tool to track and assess users' weekly patterns of physical activity. This module shows the amount of time spent exercising on each day of the week, broken out by intensity level (High, Medium, and Low) using a stacked bar chart. It provides useful information for upholding a balanced lifestyle by establishing a correlation between physical activity, dietary consumption, and health risk levels. This module's integration with the food safety system enables the administrator or user to determine whether appropriate exercise is enhancing dietary practices, particularly in cases when food-related health hazards are identified. By demonstrating how regular or diverse their physical activity is over time, it promotes individualized health advice and motivates people to adopt healthier routines.

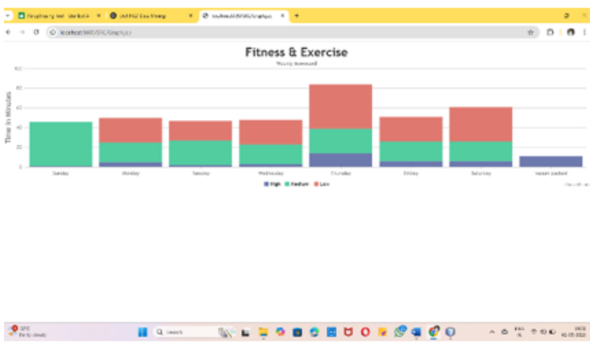

Fig 6: Weekly Fitness and Exercise Monitoring Module

IV. CONCLUSION

In conclusion, adopting a balanced diet and cultivating healthy eating habits are critical elements of a healthy lifestyle. Although it takes time to develop these habits, they have substantial long-term advantages that improve mental clarity, physical performance, and general quality of life. In addition to highlighting the value of eating healthily, our initiative incorporates contemporary technology like big data analytics and the Internet of Things (IoT) to provide individualized dietary advice. We build a dynamic, user-centered system that continuously improves based on real-world facts and user preferences by allowing users to evaluate and comment on food recommendations. When a food safety traceability system is successfully implemented, it increases transparency, fosters customer confidence, and guarantees that people have access to wholesome, safe food. In the end, our research shows how technology may be used to advance public health, support dietary literacy, and help theprevent diseases linked to poor nutrition. The groundwork for more intelligent and responsive health-support systems in the digital era is laid by this integrated approach.

REFERENCES

1. http://www.mayoclinic.org/diseases-conditions/heart disease/basics/definition/con-20034056.http://www.idf.org/aboutdiabetes.http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-what-is-breast-cancer.Kumara, M., Vohra, R., Arora, A. (2014).[1] Prediction of diabetes using Bayesian network. International Journal of Computer Science and Information Technologies, 5(4), 5174-5178.Thirumal, P. C., & Nagarajan, N. (2015). [2]
2. 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).[3]
3. 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). [4]
4. Data Mining: Practical Machine Learning Tools and Techniques, (2nd Ed), Morgan Kaufmann Publishers. Witten, I. H. & Frank, E. (2005). [5]
5. Data Mining: Practical machine learning tools and techniques, (2ndEd.) San Fransisco: Morgan Kaufmann. WEKA: Data Mining Software in Java. Retrieved fromhttp://www.cs.waikato.ac.nz/ml/weka/.elen, D., Walker, G., & Kadam, A. (2005). [6]
6. Predicting breast cancer survivability: A comparison of three data mining methods.

Artificial Intelligence in Medicine, June, 34(2), 113-127[7]

7. S. Liao, P. Chu and P. Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011", Expert Systems with Applications, vol. 39, no. 12, pp. 11303-11311, 2012. [8]
8. N. Ghiya, S. Godbole, P. Hol, G. Deortare and M. Chavan, "Forecasting
9. Diseases by Classification and Clustering Techniques", International Journal of Innovative Research in Science, Engineering and Technology, vol. 4, no. 1, pp. 18958-18961, 2015.[9]
10. Ashok Kumar, "Clustering Dichotomous Data for Health Care", IJIST, vol. 2, no. 2, pp. 23-33, 2012.[10]
11. K. Rajalakshmi, Dr.S.S. Dhenakaran and N. Roobini, "Comparative Analysis of K-Means Algorithm in Disease Prediction", International Journal of Science, Engineering and Technology Research (IJSETR), vol. 4, no. 7, pp. 2697-2699, 2015.[11]