International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Construction of Food Safety Tracebility Mechanism Using Random Forest Rule Based Algorithm

N. Manimegalai[1], Dr. P. Pritto Paul[2], T. Subashini[3], R. Swetha[4], R. Mizpah Queency[5]

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: It is commonly acknowledged that eating a diet high in nutrients can help prevent and manage non-communicable diseases (NCDs). Currently, there is a dearth of study on food elements that are nutritive and helpful in the rehabilitation of non-communicable diseases. In this effort, the data mining techniques are used to thoroughly examine the connection between dietary components and illnesses. First, we gathered the foods that were banned and advised for each of the more than n ailments that were identified. Experiments conducted on real-world data demonstrate that our data-mining strategy outperforms the conventional statistical method in terms of performance. We can help medical professionals and disease investigators identify the best nutritional components that support the recovery of various illnesses as precisely as we can. Certain data are currently unavailable because they are awaiting medical verification. The uploaded dataset will undergo pre-processing, feature extraction, noisy data removal, and classification using a rule based random forest algorithm. Based on this study, the individual's food intake will be predicted to cause the disease.

Keywords: *Data mining, Rule Based Random Forest Algorithm, Non-Communicable Diseases*

1. Introduction

COVID-19 is currently the primary basis of threats to food safety in the cold restraint logistics process. Though overall prevention and control of our country's new coronavirus pneumonia epidemic is improving, the rapid expansion of the epidemic internationally has made the domestic epidemic prevention situation of "foreign import and internal prevention" tough [1]. In the framework of preventing and managing epidemics, food safety traceability, data analysis, and monitoring have become increasingly important. The food safety traceability management system integrates data from food production, processing, storage, transportation, and sales chains using information technology and automatic identification [2]. It is therefore essential that society and its people create a reliable and safe food safety traceability management system. In today's world, food safety is crucial, with worries about everything from contamination to dishonest business practices. Innovative technologies are being investigated to provide strong food safety systems in order to overcome these issues. Implementing machine learning techniques, including the Random Forest Rule-Based (RFRB) algorithm, stop word removal algorithm, and feature selection algorithm, is one approach that can be taken.

2. Problem Identification

The majority of healthcare companies worldwide maintain their patient data in electronic format. The majority of the information about patients and the parties involved in the healthcare industry is contained in healthcare data. Such data is being stored in ever-increasing quantities. There is a certain level of complexity in electronic healthcare data because of its constant growth in size. Put another way, one may say that the data in healthcare gets extremely complex. It becomes quite tough to extract the important information from it using the standard approaches. However, because of developments in the fields of statistics, mathematics, and many other academic fields, it is now feasible to draw significant patterns from it. The intricate connections between different elements, including heredity, lifestyle, overall diet, and individual health status, are involved in the prediction of diseases based on food type or intake count. While eating particular foods in excess or in certain situations may make them more likely to cause certain diseases, it is still difficult to forecast the course of diseases based alone on the quantity and kind of food ingested. Even while it's still difficult to anticipate diseases based on food intake, new developments in science and technology present exciting opportunities to deepen our understanding of how diet affects health outcomes. But it's crucial to exercise caution when predicting diseases and to be aware of the constraints and unknowns that come with intricate biological systems.

[1] *PG Scholar, Velammal Engineering college, Surapet, Chennai 600066*
Manimalu15@gmail.com
[2] *Associate professor, Velammal Engg. college, Surapet, Chennai 600066*
Prittopaul.p@velammal.edu.in
[3] *Associate professor, Velammal Engg. college, Surapet, Chennai 600066*
Prittopaul.p@velammal.edu.in
[4] *Assistant professor, Vel Tech High Tech Dr. R. Rangarajan Dr. R. Sakunthala Engg. College*
Swetha.rvelhightech.com
[5] *Assistant professor, Vel Tech High Tech Dr. R. Rangarajan Dr. R. Sakunthala Engg. College*
Mizqueeny21@gmail.com

3. Overview of Random Forest Rule-Based (RFRB) algorithm

The RFRB algorithm creates a flexible and effective tool for managing food safety by combining the capabilities of rule-based systems and random forests, a well-liked machine learning model. The RFRB algorithm can analyse massive amounts of data and spot intricate patterns that can point to possible safety issues, in contrast to conventional approaches that depend on human inspection or basic statistical analysis. The RFRB method basically works by building a random forest, which is an ensemble of decision trees. A random subset of the data is used to train each tree, and the input data's features are used to inform the decisions that are made. The method generates reliable and precise results by combining the predictions of several trees, which makes it ideal for applications like anomaly detection and categorization. The program can identify potential dangers and suggest mitigation strategies by accessing data such as ingredient profiles, production processes, transportation conditions, and historical records of safety occurrences. Moreover, the algorithm's rule-based part permits the understanding of the underlying decision-making procedure. In contrast to black-box models, which frequently lack clarity regarding the reasoning behind predictions, the rules produced by the RFRB algorithm offer insights into the variables affecting food safety results. Establishing confidence among stakeholders and promoting regulatory compliance are made possible by this transparency.

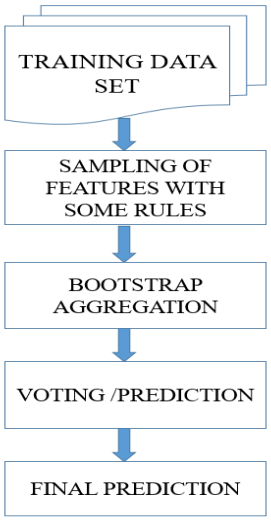

Fig. 1. Flow of RFRB Algorithm

4. Overview of Stop word removal technique

One text preparation method that is frequently employed in Natural Language Processing (NLP) activities to increase the efficacy and efficiency of text analysis is stop word removal. Words like articles, prepositions, and conjunctions are examples of stop words; they are words deemed unnecessary or unhelpful for a given mission. It can be easier to focus on the most important content and lessen noise in the text if certain terms are removed. A common preprocessing step before more complex NLP tasks like text categorization, sentiment analysis, or topic modeling is stop word removal. The performance and interpretability of these models can be enhanced by stop word removal, which gets rid of distracting and useless words. It's crucial to remember that the list of stop words may change based on the particular task and domain, and it could need to be customized in light of the text's context.

5. Overview of Feature Selection Algorithm

Indeed, feature selection plays a crucial role in reducing the dimensionality of data, particularly in text analysis tasks like document classification. When dealing with large numbers of attributes (such as words in documents), feature selection techniques help identify the most relevant and informative attributes while discarding the less useful ones.

5.1. Term Frequency-Inverse Document Frequency (TF-IDF)

Before delving into feature selection methods, numerical representations of documents are frequently used. One such technique is TF-IDF, which gives weights to each word according to how frequently it appears in a document and how uncommonly it occurs throughout the corpus. When choosing features, words with high TF-IDF scores are given priority since they are thought to be more significant.

5.2. Information Gain

Information Gain is a statistical measure used to evaluate the relevance of a feature (word) to a particular class or category. In the context of document classification, Information Gain helps identify words that are most discriminative for distinguishing between different classes of documents. Words that contribute the most to reducing the uncertainty about the document's class are considered more valuable features and are selected for inclusion in the final feature set.

6. Strategies of Implementation

The proposed methodology for developing a food safety traceability mechanism using Random Forest Rule Based techniques encompasses several key stages: upload dataset, view dataset, preprocessing and feature extraction, TF/IDF, Classification, performance analysis and user feedback.

6.1. Upload Dataset

To upload a dataset into a food safety traceability mechanism, first, ensure your data is formatted correctly. Then, log in to the traceability platform and navigate to the data upload section. Select your dataset files and map the data fields to match the system's requirements. Review and

validate the mapping for accuracy. Initiate the upload process and monitor its progress, checking for any errors. Once uploaded, verify that all data is successfully imported. Finally, perform any necessary post-upload actions such as setting permissions or initiating traceability workflows.

6.2. View Dataset

Accessing a platform-provided user interface is usually required to view your dataset in a food safety traceability mechanism. Users who have logged in can go to the dataset viewing or data management section. They may typically view, filter, and search through the submitted datasets here. Users may be able to alter their view by choosing which data fields to show or by using filters to reduce the size of the dataset, depending on the platform's features. In order to aid users in their analysis and interpretation of the data, several platforms also provide visualization tools like dashboards, graphs, and charts. In general, perusing the dataset within a food safety traceability system affords users comprehension of the gathered data, empowering them to oversee food safety protocols.

6.3. Preprocessing

Preprocessing in a food safety traceability mechanism involves preparing and refining data to ensure its quality, consistency, and suitability for analysis and traceability purposes.

Data Cleaning: This process involves identifying and rectifying any errors or inconsistencies in the data. This may include removing duplicate records, correcting spelling errors, handling missing values, and resolving inconsistencies in data formats.

Normalization: To guarantee consistency and comparability between various datasets, normalize the data. This could entail standardizing formats, naming convention uniformity, and measurement conversion.

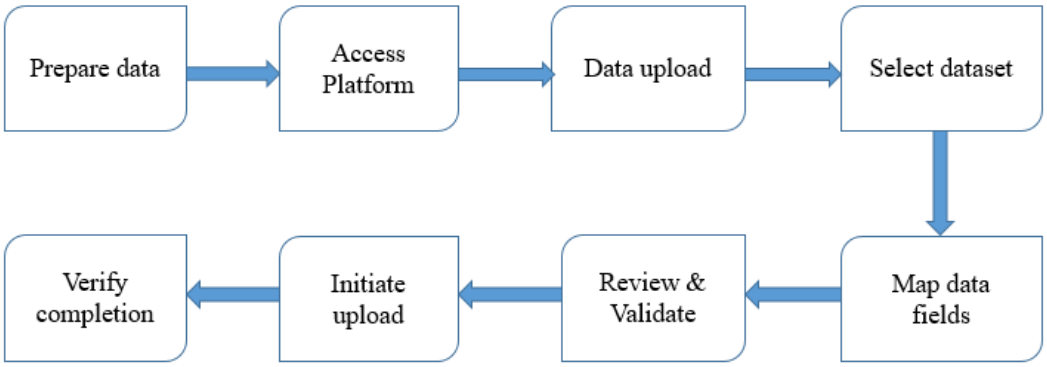

Fig. 2. Uploading Dataset

Data integration: To produce an all-encompassing picture of the food supply chain, merge data from several sources into a single dataset. This could entail creating relationships between several datasets, resolving attribute name issues, and combining databases[17-22]

Quality Assurance: Perform quality checks to ensure that the preprocessed data meets the required standards for accuracy, completeness, and reliability. This may involve running validation checks, comparing against reference data sources, and verifying the consistency of derived attributes.

6.4. Feature Extraction

The process of extracting new features or representations from the original data that highlight significant trends or connections is known as feature extraction. To lessen the dataset's complexity, this may entail using dimensionality reduction techniques. Product descriptions, reviews, and regulatory documents are examples of textual data from which insights can be extracted using text mining and natural language processing (NLP). Time series analysis can be used to spot trends in consumption patterns or changes in health over time.

Nutritional Information: Take out characteristics of food items that are linked to their nutritional value, such as calorie counts, macronutrients (including proteins, carbs, and fats), and micronutrients (like vitamins and minerals). This data aids public health programs that encourage better eating habits and assists consumers in making educated dietary decisions.

Information on allergies: List the characteristics of common food allergies, like gluten, nuts, dairy, and shellfish. For people who have food allergies or

intolerances, this knowledge is essential for preventing negative health reactions and avoiding consuming potentially dangerous components.

Contaminant Levels: Extract features related to contaminants such as pesticides, heavy metals, and microbial pathogens present in food products. Monitoring contaminant levels helps ensure food safety standards are met and prevents foodborne illnesses and related health issues.

Food Additives: Extract features related to food additives such as preservatives, flavor enhancers, and colorants. Monitoring the presence and levels of additives helps assess their potential health effects and supports regulatory efforts to minimize their use or ensure their safety.

Foodborne Illness Reports: Extract features from reports of foodborne illness outbreaks, including information on the causative agents, affected individuals, and implicated food products. Analyzing such data helps identify sources of contamination, implement preventive measures, and respond effectively to public health threats.

Health Outcomes: Extract features related to health outcomes associated with food consumption, including rates of obesity, diabetes, cardiovascular disease, and other chronic conditions. Analyzing these features helps identify associations between dietary factors and health outcomes and informs public health interventions aimed at reducing disease burden.

6.5. Classification

Classification in food safety traceability mechanisms involves the categorization of food products based on various attributes, characteristics, or quality indicators to ensure people's health.

6.5.1. Labelling and Training Data Preparation

Food products are labeled based on predefined categories or classes related to their safety and quality attributes.

Define Label Categories: The first step is to define the categories or classes that the food products will be classified into. These categories typically represent different levels of safety, quality, or other relevant attributes. For example, label categories may include "safe for consumption," "contaminated," "organic," "non-GMO," "high-risk allergen," etc.

Data Labelling: Each data sample (e.g., individual food product) is labeled with the appropriate category based on its attributes and characteristics. Labeling can be done manually by domain experts who review the data samples and assign the appropriate category based on predefined criteria. Alternatively, automated labeling techniques may be used, such as using existing data patterns or applying rule-based algorithms to assign labels.

7. Methodology

To successfully log in to the account, the admin must enter the correct username and password. Once logged in, the admin can upload two datasets containing food-related and disease-related data. After viewing both datasets, the admin can proceed with the preprocessing stage, where they can remove any noises or unwanted information from the datasets. Finally, the admin can extract features from the datasets using a percentage value. User registers, logs in, views admin recommendations, sees food intake, percentage level, disease, preventive measures, and aids, provides feedback, and logs out.

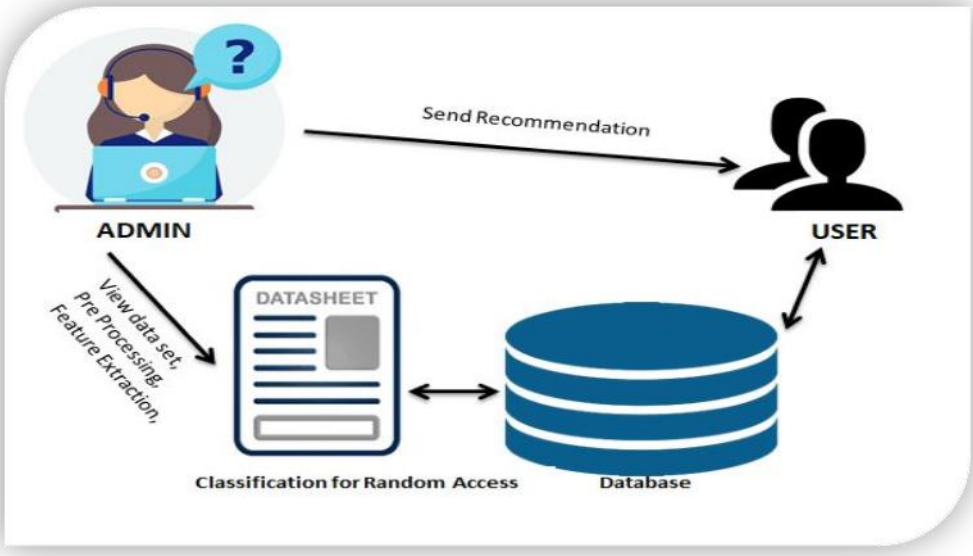

Fig. 3. Architectural Flow

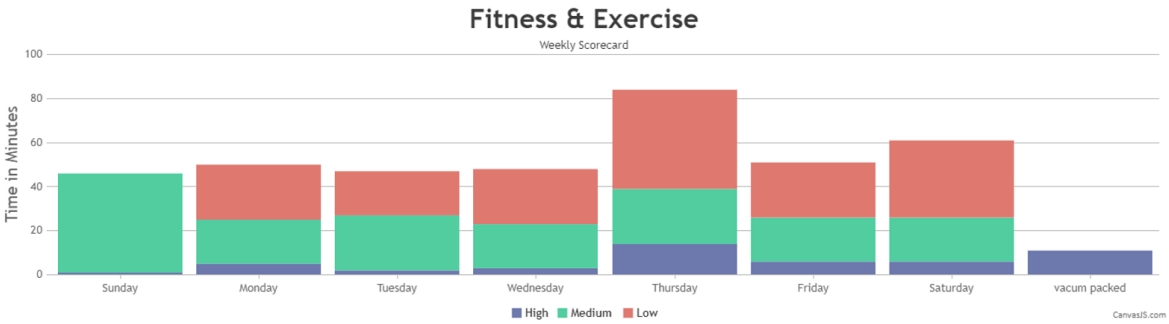

Fig. 4. The above figure represents the maintenance chart of the patients with the tracking of their information.

8. Conclusion

An unbalanced meal does not suddenly ruin good nutritional habits or a balanced diet, they take time to acquire. A higher performance in the classroom, the gym and on the dance floor will result from eating a balanced, healthful diet and keeping your weight within appropriate ranges. Giving feedback by user to admin and rating for admin recommendation foods. The Random Forest Rule Based algorithm is a machine learning technique that can enhance food safety by identifying potential risks associated with food items. It provides accurate predictions based on factors like ingredients, storage conditions, and handling practices. The model also offers real-time decision support, enabling stakeholders like food inspectors, retailers, and consumers to make informed decisions. Random Forest models can be trained on large datasets and adapt to different food products and safety criteria, making them applicable across various industries. The mechanism can continuously improve its performance and accuracy through ongoing monitoring, feedback collection, and model updates. Additionally, automating food safety evaluation using machine learning techniques can potentially reduce the time, cost, and human resources required for manual inspection and assessment.

9. Future Work

In our next effort, we will combine the several strategies discussed above to increase the diagnosis accuracy in cases with imperceptibly recognized data sets. Finding the main cause of young people's drug use and indulgence was the main objective of the study. The goal of ongoing initiatives is to expand the data set's size. The research is very helpful in analyzing the several aspects that contribute to the current drug boom. The system is extremely helpful to the user in identifying the many aspects of drug addiction, which will enable the provider to prescribe the appropriate prescription for the individual and assist preserve their valuable human life.

References

[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] Utilization of data mining techniques for diagnosis of diabetes mellitus - A case study. ARPN Journal of Engineering and Applied Sciences, January, 10(1), 8-13. Gomathi, K. (2012).

[3] An empirical study on breast cancer using data mining techniques. International Journal of Research in Computer Application & Management, July, 2(7), 97-102.Witten, H. I., & Frank, E. (2005).

[4] Data Mining: Practical Machine Learning Tools and Techniques, (2nd Ed), Morgan Kaufmann Publishers. Witten, I. H. & Frank, E. (2005).

[5] Data Mining: Practical machine learning tools and techniques, (2ndEd.) San Fransisco: Morgan Kaufmann. WEKA: Data Mining Software in Java. Retrieved from http://www.cs.waikato.ac.nz/ml/weka/.Delen, D., Walker, G., & Kadam, A. (2005).

[6] Predicting breast cancer survivability: A comparison of three data mining methods. Artificial Intelligence in Medicine, June, 34(2), 113-127

[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] N. Ghiya, S. Godbole, P. Hol, G. Deortare and M. Chavan, "Forecasting 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] Ashok Kumar, "Clustering Dichotomous Data for Health Care", IJIST, vol. 2, no. 2, pp. 23-33, 2012.

[10] 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] Manikas, I., & Manos, B., Design of an integrated supply chain model for supporting traceability of dairy products. International Journal of Dairy Technology. 2009, 62(1), 126- 138.

[12] Pei, X.F., Tandon, A., Alldrick, A., Giorgi, L., Huang, W., & Yang, R.J., The China melamine milk scandal and its implications for food safety regulation. Food Policy. 2011, 36(3), 412-420.

[13] Kumar, A., Wright, I.A., & Singh, D.K., Adoption of food safety practices in milk production: implications for dairy farmers in India. Journal of International Food & Agribusiness Marketing. 2011, 23(4), 330-344.

[14] Astill, J., Dara, R. A., Campbell, M., Farber, J. M., Fraser, E. D., Sharif, S., & Yada, R. Y. (2019). Transparency in food supply chains: A review of enabling technology solutions. Trends in Food Science & Technology, 91, 240–247.

[15] Carr, J., Decreton, L., Qin, W., Rojas, B., Rossochacki, T., & wen Yang, Y. (2015). Social media in product development. Food Quality and Preference, 40, 354–364.

[16] Mohan, A., & K, S. . (2023). Computational Technologies in Geopolymer Concrete by Partial Replacement of C&D Waste. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 282–292.

[17] Mohan, A., Dinesh Kumar, R. and J., S. 2023. Simulation for Modified Bitumen Incorporated with Crumb Rubber Waste for Flexible Pavement. International Journal of Intelligent Systems and Applications in Engineering. 11, 4s (Feb. 2023), 56–60.

[18] R.Gopalakrishnan, Mohan, "Characterization on Toughness Property of Self-Compacting Fibre Reinforced Concrete", Journal of Environmental Protection and Ecology 21, No 6, 2153–2163 (2020).

[19] Vidhya Lakshmi Sivakumar, A.S. Vickram, Ragi Krishnan, Titus Richard, "AI-Enhanced Decision Support Systems for Optimizing Hazardous Waste Handling in Civil Engineering," SSRG International Journal of Civil Engineering, vol. 10, no. 11, pp. 1-8, 2023. Crossref, https://doi.org/10.14445/23488352/IJCE-V10I11P101

[20] V. Sundaram, Vidhya Lakshmi Sivakumar, Anand Raju, A. Saravanan, Titus Richard, "The Role of 3D Printing in Engineering and its Historical Perspective in Civil Engineering - A Review," SSRG International Journal of Civil Engineering, vol. 10, Vidhya Lakshmi Sivakumar, A.S. Vickram, Ragi Krishnan, Titus Richard, "AI-Enhanced Decision Support Systems for Optimizing Hazardous Waste Handling in Civil Engineering," SSRG International Journal of Civil Engineering, vol. 10, no. 11, pp. 1-8, 2023. Crossref, https://doi.org/10.14445/23488352/IJCE-V10I11P101

[21] V. Sundaram, Vidhya Lakshmi Sivakumar, Anand Raju, A. Saravanan, Titus Richard, "The Role of 3D Printing in Engineering and its Historical Perspective in Civil Engineering - A Review," SSRG International Journal of Civil Engineering, vol. 10, no. 12, pp. 58-68, 2023. Crossref, https://doi.org/10.14445/23488352/IJCE-V10I12P107 no. 12, pp. 58-68, 2023. Crossref, https://doi.org/10.14445/23488352/IJCE-V10I12P107