Check for updates

# Food Label Analyzer using Nutri-Score Algorithm and Content Based Filtering

Samundiswary S*, Arnab Samanta, Iliyas Shah, and Suhail Khan

SIES Graduate School of Technology, Nerul, Navi Mumbai, India

samundiswarys@sies.edu.in

**Abstract.** Consumers struggle to interpret food labels, leading to uninformed dietary choices that may impact long-term health. Our study addresses this issue by providing a user-friendly solution that analyzes the nutritional quality of packaged foods and delivers clear, actionable insights. Using the Nutri-Score system, we assign a color-coded A-E rating based on both beneficial (fiber, protein, fruits) and harmful (sugar, saturated fat, salt) ingredients. As part of this study, we have compared Nutri-Score with the Health Star Rating to determine the most effective nutritional assessment method.

To make nutritional information more accessible, the platform enables users to retrieve key details effortlessly. Consumers can scan product barcodes to instantly access nutritional breakdowns, including ingredient compositions, potential allergens, and additives. Additionally, food label images can be analyzed to identify important details related to nutritional content. The platform also suggests healthier alternatives by comparing ingredient compositions and nutritional quality. By simplifying food label interpretation, our research aims to empower consumers to make healthier dietary choices with ease. ...

**Keywords:** Food labels, Nutri-Score, ML (Machine Learning), Content based filtering, Health score.

## 1 Introduction

In the modern age, when packaged foods are consumed on the mass scale, consumers face serious difficulties in taking nutrition-based decisions because labeling of food is inherently complicated and opaque. Unhealthy food items are unintentionally consumed because nutritional breakdown, the lengthy ingredient lists, and regulatory restrictions are frequently disregarded or misinterpreted. Obesity, diabetes, and cardiovascular disease are among the non-communicable diseases that are more common as a result of this inadequate information. Regional variations in food labeling procedures exacerbate this problem by adding to customer confusion and misunderstanding. Because of this, people frequently base their decisions on front-label marketing promises like "low-fat" or "sugar-free" without fully comprehending the product's nutritional makeup thus, deceptive claims.



R. Magar et al. (eds.), *Proceedings of the MULTINOVA: First International Conference on Artificial Intelligence in Engineering, Healthcare and Sciences (ICAIEHS-2025)*, Advances in Intelligent Systems Research 198,
https://doi.org/10.2991/978-94-6463-852-3_14

In order to address these issues, the Food Label Analyzer is proposed as an AI-based solution that could simplify the reading of food labels and enable customers to easily choose healthier options. Users can get comprehensive nutritional information by scanning product barcodes, which are only needed to identify the product and retrieve previously stored data from the system's internal database. Product images can be used and analyzed using a Large Language Model (LLM) to extract important information such as ingredients, additives, allergies, and the nutritional value of the Big 7 nutrients in cases when product information is unavailable.

A smart recommendation system that uses ingredients, composition similarity, and Nutri-score to provide users with healthier alternative options is based on Term Frequency-Inverse Document Frequency (TF-IDF) and cosine similarity.

The goal of this study is to use a thorough, scientifically based strategy to close the gap between advanced nutritional information and consumer understanding. By offering precise and comprehensible information, the Food Label Analyzer helps people make better food choices and contributes to broader improvements in public health by raising understanding of nutrition.

## 2 Related Work

Growing awareness of nutrition and health issues in recent years has motivated scientists and engineers to develop methods that tell customers about the quality and composition of packaged foods. Health Star Rating (HSR) and Nutri-Score are two front-of-pack labeling systems that have grown in popularity because they help consumers make healthier choices. However, the majority of these systems do not have the real-time data processing capabilities, comprehensive ingredient-level data, or user-focused personalization features necessary for complete dietary choices.

In order to suggest healthier food substitutes, Loesch et al. [2] used graph neural networks with Nutri-Scores, demonstrating advancements in algorithmic substitution. However, the study does not include real-time comparison based on semantic similarity metrics such as cosine similarity and TF-IDF, nor does it include direct label extraction. This leads to a lack of personalization in dynamic consumer applications. The proposed platform addresses this by utilizing the Gemini Vision API for real-time label interpretation, which reads images and interprets text data to support substitution recommendations and ingredient-level content analysis.

A comprehensive review of HSR labels and future research goals were given by Hasni et al.[3], although the paper is not an implementation that can be easily turned into an end-user mobile or online application. The created solution addresses this limitation by allowing end users to add new items and offers a personalized interface tailored to Indian products that has been pre-curated and cleaned from Open Food Facts.

PaddleOCR and ChatGPT were used by Rosyadi et al. [25] to extract ingredient information from labels, which is a useful step for label comprehension. However, neither an incorporated recommendation module nor a conclusive scoring mechanism is present in the architecture that has been described. By including Nutri-Score calculations, the detection of hazardous components, and product recommendations based on vectorized ingredients, the developed approach improves on existing work.

Seitaj and Elangovan [4][24] looked into text extraction using machine vision, while Shah et al.[21] looked into OCR-based nutrition table and ingredient understanding. However, neither study offers a downstream application such as contextual product suggestion or health rating; they simply address content extraction. Based on ingredient similarities, the integrated method, on the other hand, makes it easier to calculate Nutri-Score and suggests healthier substitutes.

A multi-criteria grocery recommender was developed by Hafez et al. [8], and real-time personalization for healthcare use cases was found to have limitations by Tran et al. [22]. Despite their insightfulness, such attempts do not handle ingredients at the real-time level and mostly rely on information. By examining the contents of real labels and providing context-aware recommendations, the system described in this research goes above and beyond.

The growing interest in employing digital technologies to examine food labels and the health of customers is demonstrated by these recent studies.

Earlier studies have also explored various mechanisms. In their comparison of NLP and machine learning methods for food classification and prediction, Hu et al. [26] showed remarkable categorization abilities. Real-time processing of product labels or health-centered recommendations, however, are not part of their job. By using the Gemini Vision API for label extraction and semantic similarity for comparison, the system in this case offers ingredient-based personalization and recommendations.

Consumer choices about food intake have also been impacted by the introduction of front-of-pack labeling technologies such as Nutri-Score [1] and Health Star Rating (HSR) [3][9]. Hafner and Pravst [9] used a large database of branded foods to compare Nutri-Score and HSR, while Julia and Hercberg [1] presented the effectiveness of Nutri-Score towards healthier food choices. However, real-time product analysis mechanisms, dynamic fine-grained label content extraction, and ingredient-driven product suggestion are not included in these studies, which are primarily focused on the scoring schemes themselves.

The MobiHealth app [6] and FoodSwitch [7] both use barcode scanning to help users make healthier decisions. However, these platforms do not offer dynamic new product addition or ingredient-level semantic analysis; instead, they rely on structured databases derived from barcodes. This suggested approach provides real-time text-by-image analysis through the Gemini Vision API, which may internally integrate OCR techniques for label data extraction. It uses cosine similarity and TF-IDF to offer substitute options.

Movie recommendation is only one of the many fields that have made extensive use of textual similarity-based models like TF-IDF and cosine similarity

[5][10][11][12]. Wang et al. [14] showed hybrid upgrades for collaborative filtering, whereas Guo and Yang [13] enhanced TF-IDF for keyword relevance. Although there hasn't been much direct use of these models for food product labels, the current method makes use of them by turning food ingredients into vectors that may be used to identify semantically comparable products to suggest.

The psychological influence of labels and clear food information is highlighted by consumer sentiment research, such as those conducted by Sundar et al. [15], Pagliai et al. [16], and Martini et al. [19]. These tend to be survey and observation-based, but the combined platform uses these observations to create a useful digital tool that provides current, transparent, and localized label information.

The health risks associated with ultra-processed foods and the shortcomings of current labeling systems were highlighted by Juul and Hemmingsson [17] and Mertens et al. [18]. Improved transparency is emphasized in these grassroots studies that attempt to raise public health awareness. The implementation created in response to these research permits centralized access to nutritional content and promotes better alternatives, even though they don't specify a technical system.

Consumer behavior in response to labeling tactics has been studied by Roberto et al. [20] and Song et al. [23]. Although these studies aid in the creation of effective labeling, they fall short of addressing technical implementation. By using label content to inform suggestions, the suggested platform offers a tangible response to these findings.

It is evident from the body of literature that previous studies have only looked at recommender systems, label text recognition, or health assessment separately. The combination of these elements, real-time image-based label parsing using Gemini Vision, Nutri-Score calculation, allergen and additive detection, and content-based food recommendation does not exist. The recommended method closes these gaps with a hand-crafted interface based on Open Food Facts' Indian product data, employing TF-IDF and cosine similarity for astute matching and offering a user-friendly interface for making educated food choices.

## 3 Methodology

The proposed system, Food Label Analyzer, is made to give consumers a simple and efficient way to understand complex food label information, enabling them to make better dietary decisions. The system workflow diagram in the below figure 1, illustrates how the system integrates barcode scanning, machine learning techniques, natural language processing, and similarity algorithms into an interconnected framework.

The process begins when a user queries the system, either by scanning a barcode or entering the name of a product. The backend processes this to see if the product is already in the internal database. Information about the product, including its ingredients, allergies, and nutritional statistics, is gathered and transmitted through the Health Scoring Engine if it is identified. The main component of the system is the Nutri-Score algorithm, which rates food items on both positive (such as fiber and protein) and negative (like sugars, saturated fat, and sodium) nutrients. A standardized color-coded scale from A (healthy) to E (unhealthy) is then created using the score. Users can quickly assess a product's healthiness thanks to this abstraction, which eliminates the need to decipher the nutritional data.

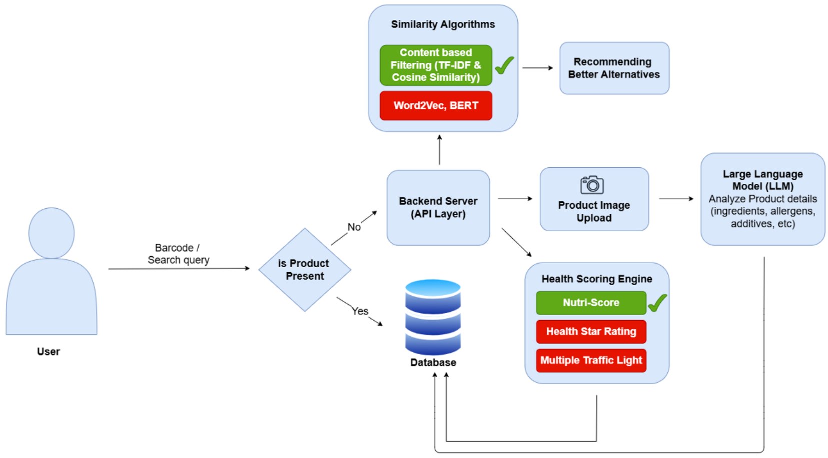



**Fig. 1.** Food Label Analyzer - System Workflow

Users are asked to upload a picture of the product's package label if the item is not in the database. A Large Language Model (LLM) is used to process this image, applying OCR and semantic analysis to obtain comprehensive nutritional information, such as ingredient lists, allergy percentages, and additive usage. After that, the collected data is entered into the same Nutri-Score evaluation pipeline.

In addition to health grading, the platform has a recommendation engine that suggests healthier solutions to improve user experience. This is accomplished by comparing the nutritional and component makeup of items using content-based filtering techniques, primarily TF-IDF and Cosine Similarity.

The system's overall goal is to promote nutritional transparency and customer self-awareness in addition to simplifying label comprehension. The Food Label Analyzer aims to enable consumers to make more informed, healthful decisions by deciphering food label information and exposing hidden contaminants. Our overall objective is to leverage the power of easily accessible and intelligent food analysis to encourage better feeding practices and build a community that is health - focused.

### 3.1 LLM-Based Food Label Evaluation

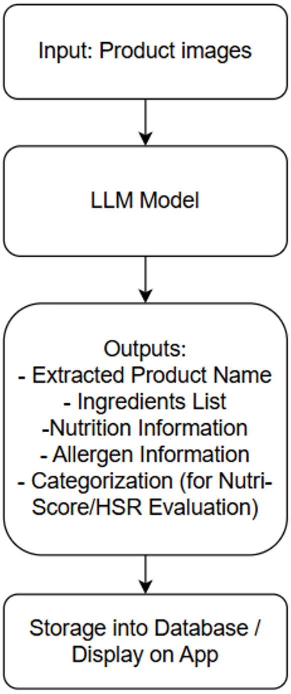



**Fig. 2.** LLM - based Product Detail Extraction Pipeline

The interface integrates an end-to-end information extraction module driven by Large Language Model (LLM) that processes user-uploaded product label photos to handle newly encountered packaged food items. To extract and identify key product aspects such the product name, components list, nutritional value, and allergies, we specifically use Gemini LLM, which can perform multi-modal analysis.

As illustrated in Figure 2, the LLM first processes the image, semantically splitting the label's visual and textual components. For jobs involving product classification and subsequent health grading algorithms (like Nutri-Score), the output includes structured data categorized by category. The extracted data is then concurrently made available for user viewing within the program and stored in the internal database.

This module enables the system to adapt to a changing food product environment, significantly increasing its scalability and reliability. Support for user-initiated, real-time product consumption is ensured, removing the need for static databases and enabling dynamic, personalized nutritional analysis.

### 3.2 Nutri Score Algorithm

The Nutri-Score algorithm is a nutritional rating system that was developed scientifically and evaluates food products' healthiness according to their composition. To evaluate the nutritional value of food products, the algorithm generates a numerical score that takes into account both positive and negative factors. In order to assist customers in making well-informed food selections, this score is then assigned a letter grade ranging from A to E.

$$\text{Nutri-Score Points} = (\text{Total Negative Points}) - (\text{Total Positive Points}) \tag{1}$$

In the Nutri-Score system, components are divided into positive and negative to balance the healthfulness of a food product.

a. **Negative Points:** These are assigned to components that negatively impact health when consumed in excess. A higher value increases the total Nutri-Score, making the product less healthy.
   - **Energy (kJ):** Contributes to obesity and metabolic issues.
   - **Total Sugars (g):** Excess sugar increases diabetes and weight gain risk.
   - **Saturated Fats (g):** High intake raises cholesterol level and heart disease risk.
   - **Salt (g):** Elevated salt intake is linked to high blood pressure.

b. **Positive Points:** These are assigned to components that enhance nutritional quality. A higher value decreases the total Nutri-Score, making the product healthier.
   - **Fiber (g):** Aids digestion, supports gut health, and increases satiety.
   - **Proteins (g):** Essential for muscle growth, tissue repair, and overall metabolism.
   - **Fruits, Vegetables, and Legumes (%):** Provide essential vitamins, minerals, and antioxidants, promoting better health.

The points for each component are assigned based on their amounts per 100g of the product. For details on how these points are allocated, refer to the table 1 and 2 below.

**Table 1.** Negative Points Allocation of Nutri-Score Calculation

| Points | Energy (kJ/100g) | Sugars (g/100g) | Saturated Fat (g/100g) | Salt (g/100g) |
|---|---|---|---|---|
| 0 | $\leq 335$ | $\leq 3.4$ | $\leq 1$ | $\leq 0.2$ |
| 1 | $> 335$ | $> 3.4$ | $> 1$ | $> 0.2$ |
| 2 | $> 670$ | $> 6.8$ | $> 2$ | $> 0.4$ |
| 3 | $> 1005$ | $> 10$ | $> 3$ | $> 0.6$ |
| 4 | $> 1340$ | $> 14$ | $> 4$ | $> 0.8$ |
| 5 | $> 1675$ | $> 17$ | $> 5$ | $> 1.0$ |
| 6 | $> 2010$ | $> 20$ | $> 6$ | $> 1.2$ |
| 7 | $> 2345$ | $> 24$ | $> 7$ | $> 1.4$ |
| 8 | $> 2680$ | $> 27$ | $> 8$ | $> 1.6$ |
| 9 | $> 3015$ | $> 31$ | $> 9$ | $> 1.8$ |
| 10 | $> 3350$ | $> 34$ | $> 10$ | $> 2.0$ |
| 11 | | $> 37$ | | $> 2.2$ |
| 12 | | $> 41$ | | $> 2.4$ |
| 13 | | $> 44$ | | $> 2.6$ |
| 14 | | $> 48$ | | $> 2.8$ |
| 15 | | $> 51$ | | $> 3.0$ |
| 16 | | | | $> 3.2$ |
| 17 | | | | $> 3.4$ |
| 18 | | | | $> 3.6$ |
| 19 | | | | $> 3.8$ |
| 20 | | | | $> 4.0$ |

**Table 2.** Positive Points Allocation of Nutri-Score Calculation

| Points | Fruits/Veg/Nuts (%) | Fiber (g/100g) | Proteins (g/100g) |
|---|---|---|---|
| 0 | $< 40$ | $\leq 3.0$ | $\leq 2.4$ |
| 1 | $\geq 40$ | $> 3.0$ | $> 2.4$ |
| 2 | $\geq 60$ | $> 4.1$ | $> 4.8$ |
| 3 | $\geq 80$ | $> 5.2$ | $> 7.2$ |
| 4 | – | $> 6.3$ | $> 9.6$ |
| 5 | – | $> 7.4$ | $> 12.0$ |
| 6 | | | $> 14.0$ |
| 7 | | | $> 17.0$ |

Based on the Nutri-Score points, the Nutri-Score method divides food products into five groups, from A (healthiest) to E (least healthy). To enhance user understanding, each class is associated with a distinct color:

– **Dark Green (A)** represents the healthiest products.

– **Light Green (B)** indicates a slightly lower but still good nutritional quality.
– **Yellow (C)** represents moderately healthy products.
– **Light Orange (D)** denotes products with lower nutritional value.
– **Dark Orange (E)** is assigned to the least nutritious products.

The following table 3, provides the classification based on the range of Nutri-Score points, assigned class, and corresponding color code:

**Table 3.** Final Nutri-Score Thresholds

| Nutri-Score Points | Class | Color |
|---|---|---|
| Min to 0 | A | Dark green |
| 1 to 2 | B | Light green |
| 3 to 10 | C | Yellow |
| 11 to 18 | D | Light orange |
| 19 to Max | E | Dark orange |

This classification enables consumers to make informed dietary choices at a glance by comparing products.

An illustration of Nutri-Score calculation in the Food Label Analyzer program is shown through Lays American Style Cream & Onion. This well-known packaged snack was selected for analysis due to its widespread use and well-known nutritional value. Table 4 displays the nutritional content breakdown used to calculate the Nutri-Score.

**Table 4.** Key nutritional values per 100g of Lays American Style Cream & Onion

| Nutrition | Values |
|---|---|
| Energy | 537 kcal $\approx$ 2246.808 kJ |
| Sugars | 3.4 g |
| Saturated Fat | 14.8 g |
| Salt | 1.61 g |
| Protein | 6.7 g |

Nutri-Score points are assigned based on predefined threshold values for both negative and positive factors referring to Table 1 and Table 2:

**a. Negative Points:** These are based on energy, sugars, saturated fat, and salt using Table 1.
   – **Energy (2246.808 kJ):** Falls between 2010 and 2345 kJ, contributing **6 points**.

  - **Sugars (3.4 g):** Equal to 3.4 g, earning **0 points**.
  - **Saturated Fat (14.8 g):** Contributing **10 points**.
  - **Salt (1.61 g):** Contributing **8 points**.

b. **Positive Points:** These are awarded for protein, fiber, and fruit/vegetable /legumes using Table 2.
  - **Protein (6.7 g):** Falls between 4.8 and 7.2 g, earning **2 points**.
  - **Fiber and Fruits/Vegetables/Legumes:** Not provided, contributing **0 points**.

The final Nutri-Score is calculated as 22 (24–2), placing the product in Grade E referring Table 3. This classification indicates a poor nutritional profile primarily due to high sugar content, with minimal beneficial nutrients like fiber or fruits/vegetables.

### 3.3 Healthier Product Recommendation

The Food Label Analyzer encourages users to choose foods that are higher in nutrients and healthier by using a hybrid recommendation system. By looking at text-based features like ingredients and categories, it first uses content-based filtering to identify related products. After that, it refines the recommendations by examining the Nutri-Score of each comparable item and suggesting products that are not only comparable but also healthier.

The following dataset columns are taken into account by the system while calculating product similarity: product name, text categories for ingredients (e.g., General Food, Beverages).

Each product has a single textual representation created by joining these columns. Prior to vectorization, the text is normalized by removing punctuation and stop words, changing it to lowercase, and maybe lemmatizing it. This reduces vocabulary noise and ensures consistency in the product descriptions.

By using the Term Frequency–Inverse Document Frequency (TF-IDF) approach, the preprocessed text is converted into numerical vectors. Each phrase is given a weight by TF-IDF according to its significance in a particular product in relation to its usage in all products.

TF-IDF Formula:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \tag{2}$$

where

- $t$ = term (e.g., *"sugar"*, *"wheat*),
- $d$ = document (i.e., the combined text of a product's category and ingredients),
- $\text{TF}(t, d)$ = Term Frequency = number of times term $t$ appears in document $d$,
- $\text{IDF}(t)$ = Inverse Document Frequency = measure of how unique term $t$ is across all documents.

For example, consider these Two very basic product descriptions:

- "Baked potato chips" is Product A,
- "Fried potato chips" is Product B.

Since they appear in both items, common terms like "potato" and "chips" will have a lower IDF, whereas unique words like "baked" and "fried" will have higher IDF values. Therefore, TF-IDF prioritizes what sets products apart from one another rather than what they have in common. This gives each product a TF-IDF vector that is projected onto a high-dimensional feature space.

We compute the cosine similarity between the TF-IDF vectors of two products to determine how similar they are. With 1 denoting complete similarity and 0 denoting no similarity, cosine similarity computes the cosine of the angle between two vectors.

Cosine Similarity Formula:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \, \|B\|} \tag{3}$$

where:

- $A \cdot B$ is the dot product of two TF-IDF vectors
- $\|A\|$ is the magnitude (Euclidean norm) of vector $A$
- $\|B\|$ is the magnitude (Euclidean norm) of vector $B$
- $\theta$ is the angle between the two vectors

Returning to the Example:

Since "fried potato chips" and "baked potato chips" share two of the three terms, their cosine similarity will be strong. However, there will be a small angle deviation because "baked" and "fried" are used interchangeably, indicating that they are similar but not precisely the same. This method ensures that products with comparable categories and ingredients are positioned closer together in the vector space and are therefore identified as similar.

A filter is applied to the second layer using the Nutri-Score, a normalized health grading system (A to E, with A being the healthiest), following the identification of a group of related items using cosine similarity. Every food has a predetermined Nutri-Score based on its nutritional makeup (energy, saturated fat, sugar, sodium, etc.). Comparable items with Nutri-Scores that are equal to or higher than the reference product are given preference by the algorithm.

For instance, the system promotes Lays American Style Cream & Onion as a healthier choice if it has a Nutri-Score of E and, among comparable goods, a baked version or a reduced-sodium brand has a Nutri-Score of D. This is seen in figure 3.

The final outcome is a carefully chosen selection of goods that are:

- textually similar to the selected product (with regard to categories, labeling, and ingredients),
- improved in health, as indicated by a higher Nutri-Score.

This two-step suggestion process ensures that customers are guided toward better dietary choices.

### 3.4 Dataset

The data used in this study is sourced from the Open Food Facts database, an open, crowdsourced database that provides comprehensive data for packaged food products worldwide, provided the data used in this study. More than 1.7 million recordings from all across the world were included in the first data. By choosing just products marked as India, the data were filtered according to the study's interest in the Indian market, reducing the total number of entries to roughly 8,000.

Further filtering was done to preserve the relevance and quality of the data. Only entries that had a valid `ingredients_text` field that wasn't empty were retained. The dataset was then cleaned to get 1,680 high-quality entries.

The dataset includes features such as:

– Product Barcode
– Image URLs (for front pack, ingredients, and nutrition)
– Product name
– Ingredients list
– Nutritional values (energy, fat, saturated fat, carbohydrates, sugars, protein, salt)
– Additives & Allergens
– Category

Preprocessing procedures comprised:

– Removing records of ingredients and nutritional values that are incorrect or missing.
– Eliminating unnecessary columns such as origins, countries, states, labels, retailers, and packaging.
– Grouping products into categories: Water, Cheese, Dairy, Red Meat, Fats/Nuts/Seeds/Oil, and General Food.

These procedures all contributed to the dataset's consistency, cleanliness, and preparedness for nutritional analysis and intelligent recommendation systems.

## 4 Results

### 4.1 Product Details Page: Comprehensive Nutritional Insights and Healthier Alternatives

The most significant feature for users of the Food Label Analyzer is the Product Details page, which provides comprehensive and understandable nutritional data. After selecting a product, consumers can view its Nutri-Score, which is a color-coded rating ranging from A to E that indicates the product's nutritional content. The page provides important product details in an orderly and user-friendly manner, including a list of the product's components, additives, and allergens. Values for the Big 7 nutrients—calories, fat, saturated fat, carbohydrates, sugar, fiber, and protein—are shown in the nutritional information.

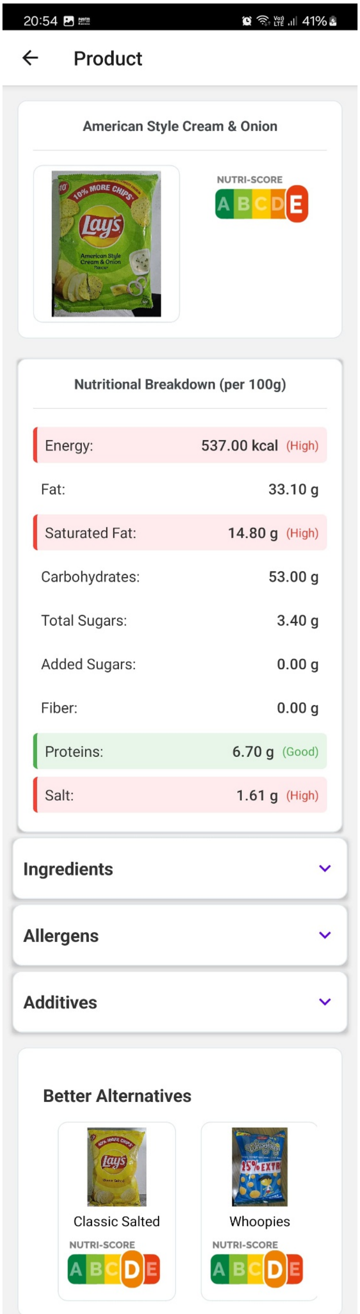



**Fig. 3.** Product Detail page

Each nutritional value is color-coded (green, yellow, and red) to make it clear whether it exceeds safe intake levels or is within suggested healthy limits. With the use of this visual assistance, consumers may quickly assess a product's healthiness. Additionally, the system presents customers with healthier options by recommending better selections based on the product's Nutri-Score and constituent composition. This page supports better eating habits and increased general well-being by giving users concise and important information to help them choose healthier foods. Users are able to understand the nutritional value of items and choose healthier alternatives thanks to the combination of real-time information, customized insights, and recommendations for substitute products. Above figure 3, shows the product detail page.

### 4.2 Product Recommendation System

Word2Vec with Euclidean distance and TF-IDF with cosine similarity were the two algorithms used to evaluate the recommendation model and assess how well the suggested approach performed. Table 5 discusses the performance metrics that are implemented.

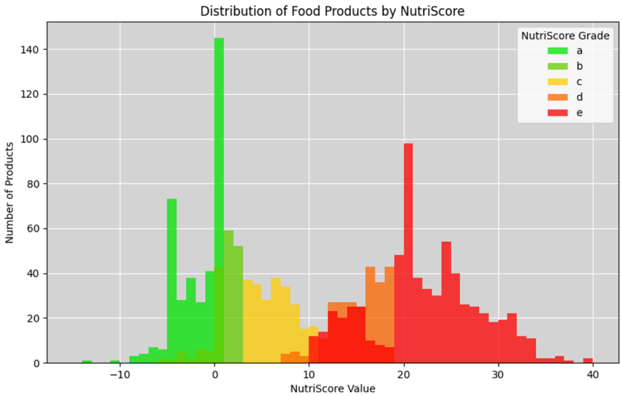
**Table 5.** Algorithm Comparison

| Algorithm | Precision | Recall | Average |
|---|---|---|---|
| TF-IDF and Cosine Similarity | 0.86 | 0.0159 | 0.0152 |
| Word2Vec and Euclidean Distance | 0.732 | 0.0085 | 0.0078 |

A high precision of 0.86 was attained by TF-IDF with cosine similarity, as indicated in Table 5, suggesting that the interface primarily suggests important items.

### 4.3 Health-Score Algorithm Analysis

Although both the Nutri-Score and Health Star Rating (HSR) approaches are used to evaluate the healthiness of foods, the Nutri-Score graph clearly displays a bimodal distribution, segmenting products into healthy (A, B) and unhealthy (D, E) categories to make it simpler for customers to quickly identify products. On the other hand, products are distributed throughout a continuous range of 0.5 to 5.0 stars on the Health Star Rating (HSR) graph. For decisions that need to be taken right away, Nutri-Score offers a faster visual contrast, while HSR offers more thorough comparisons. Figure 4 below, depicts a graphical representation of Nutri-Score and Health Star Rating on the same dataset.




**vs**

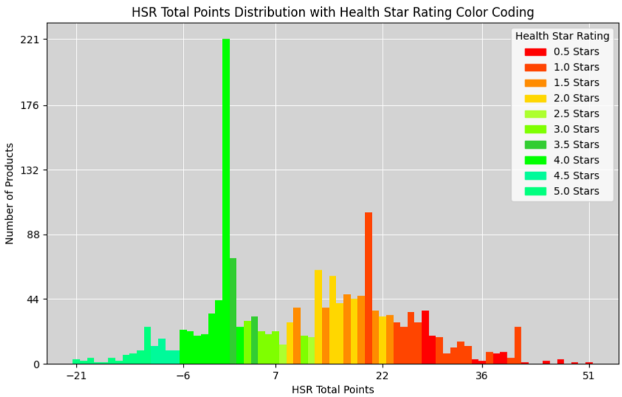



**Fig. 4.** Graphical Comparison of Nutri Score vs HSR

There are significant differences between Nutri-Score and the Health Star Rating (HSR) when it comes to how they evaluate cheese products. Cheese products are evaluated differently using the Health Star Rating (HSR) and Nutri-Score. Foods with high fat and salt content typically receive lower rankings from Nutri-Score, which takes into account both positive and negative factors. However, HSR places greater emphasis on vital nutrients like protein and calcium, which results in higher scores even when foods contain unhealthy components.

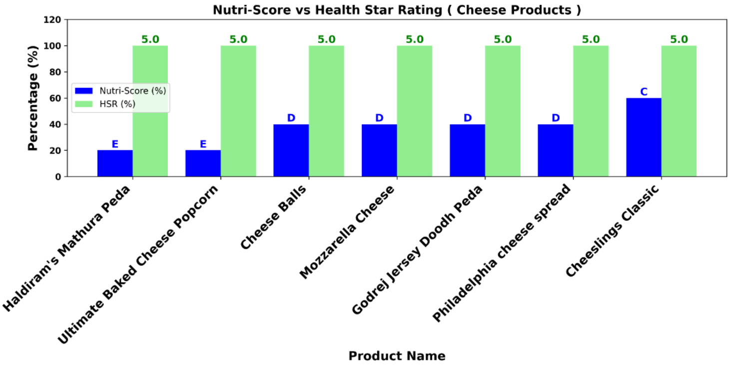



**Fig. 5.** Nutri Score vs HSR (Cheese Product)

Cheese products such as cheese popcorn, mozzarella cheese, and cheese spread are categorized under D or E in the Nutri-Score system due to their high fat and salt content, as seen in Figure 5. These products, however, frequently get a score of 5.0 under the Health Star Rating (HSR) system, indicating that its grading scheme is not very stringent. This highlights how Nutri-Score takes into account both nutritious components that are good and those that are bad, offering a more thorough and stringent health score evaluation.

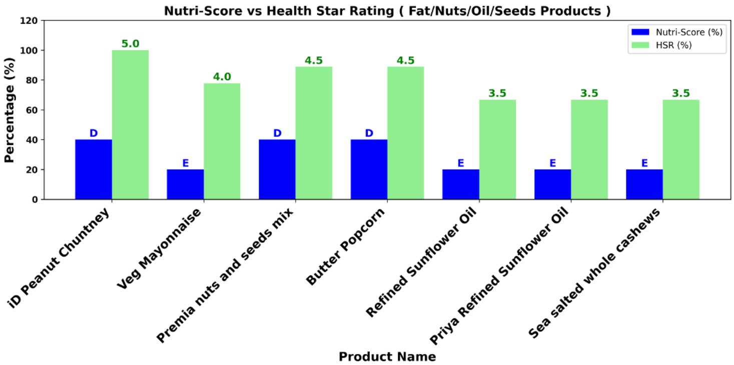



**Fig. 6.** Nutri Score vs HSR (Fat/Nuts/Oil/Seeds Product)

When it comes to assessing fat, nut, and oil goods, the Nutri-Score and Health Star assessing (HSR) systems differ significantly, as shown in Figure 6. Due to issues with energy density and processing, Nutri-Score assigns lower scores (D or E) to items including butter popcorn, vegetable mayonnaise, and refined sunflower oil. On the other hand, HSR frequently assigns the same items higher ratings (3.0 to 5.0 stars), possibly as a result of its nutrient-based algorithm, which may not account for processing level.

For example, because to its fatty acid makeup, refined sunflower oil receives a 'E' from Nutri-Score but 3.0 to 3.5 stars in HSR. This disparity demonstrates Nutri-Score's stricter stance on highly processed and calorically dense foods. According to a comparison by Hafner and Pravst [9], Nutri-Score would give olive oils a higher rating than HSR since it places greater weight on the quality of fats rather than just their calorie or macronutrient amount. Overall, Nutri-Score appears to offer a more health-conscious evaluation than HSR.

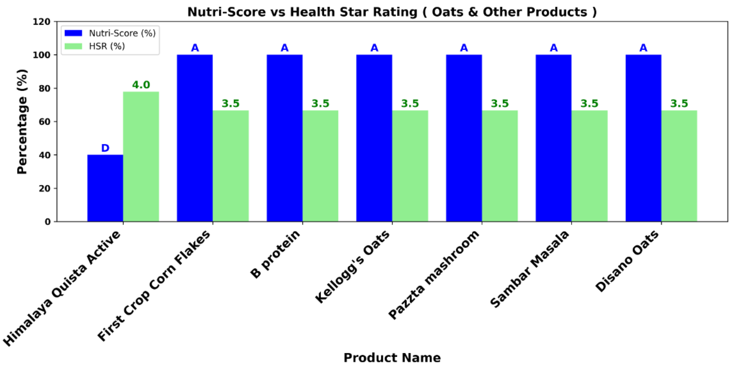



**Fig. 7.** Nutri Score vs HSR (Oats and Other Food Product)

As seen in Figure 7, Nutri-Score often assigns high (A) ratings to oat-based goods such as corn flakes, oats, and ready-to-eat combinations based on its priority for sugar levels, fiber intake, and whole grain content. In contrast, HSR displays a more conservative assessment by giving these same products a moderate rating of 3.5 to 4.0 stars. A notable example of a discrepancy is when a cereal like Himalaya Quista Active receives a 4.0 from HSR but a D from Nutri-Score. Consequently, it appears that Nutri-Score is more effective in identifying healthier cereal options.

## 5 Conclusion

In conclusion, this study presents the Food Label Analyzer, a comprehensive tool designed to give consumers easily understandable and useful information on the nutritional value of packaged foods. Main contribution is the use of Gemini, a Large Language Model (LLM), for product picture and text semantic understanding. This model enables robust label extraction even when dealing with non-standard formats. The technology provides the most recent health assessment for a range of food items and allows for the real-time calculation of Nutri-Score for new or unknown goods. Using nutritional value and component similarities, a content-based recommendation engine suggests healthier substitutes. Additionally, the platform supports goods that were not initially included in the Open Food Facts dataset and enables user-driven product importation, dynamically replenishing the database. Additionally, allergy detection is integrated, enhancing user safety by taking particular dietary requirements into account.

The recommendation system performed well in offering healthy substitutes, as evidenced by its precision of 0.86 when TF-IDF and Cosine Similarity were combined. However, despite the good ideas, the system only retrieves a limited number of choices, as indicated by the low recall value of 0.0159, suggesting room for improvement in recommendation coverage. This proposed system is a more dynamic, multimodal, and customized solution than earlier approaches that mostly relied on static data or single evaluation criteria. In summary, the Food Label Analyzer satisfies a critical need for transparent, health-conscious food decision-making. By combining real-time analysis, personalization, and a dynamic recommendation framework, it presents a powerful step forward in consumer-focused nutritional guidance.

## 6 Future Work

Further updates may incorporate a user feedback loop to enhance recommendations through collaborative filtering, allowing the system to gradually improve by learning from users' preferences. By connecting food intake and health monitoring, interoperability with fitness tracking apps like MyFitnessPal or Google Fit can offer a comprehensive view of users' eating habits. Multilingual label interpretation will enable the system to support users from different languages and geographical locations, hence improving accessibility. Users will be better equipped to make morally and nutritionally responsible decisions with additional elements like deceptive packaging alertness and environmental footprint indicators (such as carbon footprint and packaging recyclability). Real-time product evaluations and hands-free inquiries are two other ways that voice assistant capabilities can facilitate user involvement.

## References

1. Julia, Chantal, and Serge Hercberg. "Nutri-Score: Evidence of the effectiveness of the French front-of-pack nutrition label." Ernahrungs Umschau 64, no. 12 (2017): 181-187.
2. Loesch, Julie, Ilse van Lier, Alie de Boer, Jan Scholtes, Michel Dumontier, and Remzi Celebi. "Automated identification of healthier food substitutions through a combination of graph neural networks and nutri-scores." Journal of Food Composition and Analysis 125 (2024): 105829.
3. Hasni, Muhammad Junaid Shahid, Mohsin Abdur Rehman, Nicolas Pontes, and Muhammad Zafar Yaqub. "Health Star Rating Labels: A systematic review and future research agenda." Food Quality and Preference (2024): 105310.

4. Seitaj, Hansi, and Vinayak Elangovan. "Information Extraction from Product Labels: A Machine Vision Approach." International Journal of Artificial Intelligence & Applications 15 (2024): 57-76.
5. Chiny, Mohamed, Marouane Chihab, Omar Bencharef, and Younes Chihab. "Netflix recommendation system based on TF-IDF and cosine similarity algorithms." no. Bml (2022): 15-20.
6. Wong, Man Wai, Qing Ye, Yuk Kai Chan Kylar, Wai-Man Pang, and Kin Chung Kwan. "A Mobile adviser of healthy eating by reading ingredient labels." In Wireless Mobile Communication and Healthcare: 6th International Conference, MobiHealth 2016, Milan, Italy, November 14-16, 2016, Proceedings 6, pp. 29-37. Springer International Publishing, 2017.
7. Dunford, Elizabeth, Helen Trevena, Chester Goodsell, Ka Hung Ng, Jacqui Webster, Audra Millis, Stan Goldstein, Orla Hugueniot, and Bruce Neal. "FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of national food composition data." JMIR mHealth and uHealth 2, no. 3 (2014): e3230.
8. Hafez, Manar Mohamed, Rebeca P. Díaz Redondo, Ana Fernández Vilas, and Héctor Olivera Pazó. "Multi-criteria recommendation systems to foster online grocery." Sensors 21, no. 11 (2021): 3747.
9. Hafner, Edvina, and Igor Pravst. "Comparison of nutri-score and health star rating nutrient profiling models using large branded foods composition database and sales data." International Journal of Environmental Research and Public Health 20, no. 5 (2023): 3980.
10. Mehta, Raghav, and Shikha Gupta. "Movie recommendation systems using sentiment analysis and cosine similarity." International Journal for Modern Trends in Science and Technology 7, no. 01 (2021): 16-22.
11. Muthurasu, N., Nandhini Rengaraj, and Kavitha Conjeevaram Mohan. "Movie Recommendation System Using Term Frequency-Inverse Document Frequency and Cosine Similarity Method." International Journal of Recent Technology and Engineering (IJRTE) (2019).
12. Singh, Ramni Harbir, Sargam Maurya, Tanisha Tripathi, Tushar Narula, and Gaurav Srivastav. "Movie recommendation system using cosine similarity and KNN." International Journal of Engineering and Advanced Technology 9, no. 5 (2020): 556-559.
13. Guo, Aizhang, and Tao Yang. "Research and improvement of feature words weight based on TFIDF algorithm." In 2016 IEEE information technology, networking, electronic and automation control conference, pp. 415-419. IEEE, 2016.
14. Wang, Dawei, Yuehwern Yih, and Mario Ventresca. "Improving neighbor-based collaborative filtering by using a hybrid similarity measurement." Expert Systems with Applications 160 (2020): 113651.
15. Sundar, Aparna, Edita Cao, Ruomeng Wu, and Frank R. Kardes. "Is unnatural unhealthy? Think about it: Overcoming negative halo effects from food labels." Psychology & Marketing 38, no. 8 (2021): 1280-1292.
16. Pagliai, G., M. Dinu, M. P. Madarena, M. Bonaccio, L. Iacoviello, and F. Sofi. "Consumption of ultra-processed foods and health status: a systematic review and meta-analysis." British Journal of Nutrition 125, no. 3 (2021): 308-318.
17. Juul, Filippa, and Erik Hemmingsson. "Trends in consumption of ultra-processed foods and obesity in Sweden between 1960 and 2010." Public health nutrition 18, no. 17 (2015): 3096-3107.

18. Mertens, Elly, Chiara Colizzi, and José L. Peñalvo. "Ultra-processed food consumption in adults across Europe." European journal of nutrition 61, no. 3 (2022): 1521-1539.
19. Martini, Daniela, Justyna Godos, Marialaura Bonaccio, Paola Vitaglione, and Giuseppe Grosso. "Ultra-processed foods and nutritional dietary profile: a meta-analysis of nationally representative samples." Nutrients 13, no. 10 (2021): 3390.
20. Roberto, Christina A., Shu Wen Ng, Montserrat Ganderats-Fuentes, David Hammond, Simon Barquera, Alejandra Jauregui, and Lindsey Smith Taillie. "The influence of front-of-package nutrition labeling on consumer behavior and product reformulation." Annual review of nutrition 41, no. 1 (2021): 529-550.
21. Shah, Yaksh & Jariwala, Nandan & Kachhia, Bhakti & Shah, Prachi. (2023). "Delving Deep into NutriScan: Automated Nutrition Table Extraction and Ingredient Recognition." International Journal for Research in Applied Science and Engineering Technology. Volume 11. 1596. 10.22214/ijraset.2023.56852.
22. Tran, Thi Ngoc Trang, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. "Recommender systems in the healthcare domain: state-of-the-art and research issues." Journal of Intelligent Information Systems 57, no. 1 (2021): 171-201.
23. Song, Jing, Mhairi K. Brown, Monique Tan, Graham A. MacGregor, Jacqui Webster, Norm RC Campbell, Kathy Trieu, Cliona Ni Mhurchu, Laura K. Cobb, and Feng J. He. "Impact of color-coded and warning nutrition labelling schemes: A systematic review and network meta-analysis." PLoS medicine 18, no. 10 (2021): e1003765.
24. Seitaj, Hansi, and Vinayak Elangovan. "Information Extraction from Product Labels: A Machine Vision Approach." International Journal of Artificial Intelligence & Applications 15 (2024): 57-76.
25. Rosyadi, Ahmad Wahyu, Siti Ma'shumah, Muhammad Qomaruz Zaman, and Moh Rizki Fajar. "Ingredients Identification Through Label Scanning Using PaddleOCR and ChatGPT for Information Retrieval." Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 8, no. 6 (2024): 758-767.
26. Hu, Guanlan, Mavra Ahmed, and Mary R. L'Abbé. "Natural language processing and machine learning approaches for food categorization and nutrition quality prediction compared with traditional methods." The American Journal of Clinical Nutrition 117, no. 3 (2023): 553-563.