International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 24s, 2025
https://theaspd.com/index.php

Recipe Recommendation System (Rec-Res) Using Tf-Idf And Doc2vec

Swati Mishra[1], Lakhan Patidar[2] , Kapil Patodi[3] , Pradeep Singh Hada[4], Aishwarya Singh[5], Sajal Sharma[6]
[1,2,3,4,5,6]Department of Mechanical Engineering, Medicaps University Indore, 453331, India, Department of Mechanical Engineering, IPS Academy, Institute of Engineering and Science, Indore
Email: swati.mishra@medicaps.ac.in[1], lakhan.patidar@medicaps.ac.in[2], kapilpatodi@gmail.com[3], pradeephada0206@gmail.com[4], Sajalsharma423@gmail.com[5], Singhsiddhart045@gmail.com[6]

abstract
With the growing necessity for personalized dish suggestions, recipe recommendation systems are increasingly popular nowadays. In this paper, a new recipe recommendation system is presented that integrates two state-of-the-art natural language processing (NLP) methodologies, TF-IDF (Term Frequency-Inverse Document Frequency) and Doc2Vec, in order to come up with personalized recipe suggestions considering user preferences as well as dish descriptions. The system initially employs TF-IDF to retrieve significant ingredients steps, and categories from the recipe corpus, mentioning the significant terms in the recipe vocabulary. Later, Doc2Vec is used to transform the recipe text into vector representations, allowing the system to understand the semantic connections between recipes. By comparing user preferences against the recipe vectors, the system produces personalized and contextually relevant recommendations. Experimental outcomes emphasize that the suggested method vastly enhances recommendation correctness and user experience compared to classic keyword-based suggestion systems. The paper gives an important contribution by introducing a wiser, more scalable, and user-focused recipe suggestion system relying on state-of-the-art NLP methods.

INTRODUCTION

In the modern quick-paced digital life, the convergence of culinary arts and technology has created new possibilities for customized and effective meal planning[1]. The traditional ways of finding recipes often based on random searches, cookbooks, or standard recommendation systems that fail to match users accessible ingredients or eating habits[2]. Recipe Recommendation System (RecRes) has been designed to fill this gap using sophisticated machine learning and natural language processing (NLP) to recommend appropriate recipes using user-fed ingredients[3].
As smart kitchens gain popularity and there is more use of AI-based solutions, there's a need for more intelligent systems able to help users in meal planning while also reducing food wastage[4]. Most current recipe suggestion systems are based only on keyword search, which is not context aware and personalized. RecRes expects to bypass these constraints by using advanced ingredient recognition models and suggestion algorithms like TF-IDF (Term Frequency-Inverse Document Frequency) and Doc2Vec to increase the precision and relevance of recipe suggestions[5].
The Recipe Recommendation System (RecRes) is a sophisticated artificial intelligence (AI) based-learning system designed to enhance meal preparation, ingredient management, and tailored food suggestions[6]. In the increasingly busy digital world of today, individuals are generally left wondering what to prepare for meals. Cooking the same meal repeatedly, wasting hundreds of dollars on unused ingredients, and letting good food go to waste is simply not a sustainable or acceptable way to eat nowadays. Old-fashioned practices of searching for recipes flipping through cookbooks or performing random online searches are not contextual or personalized enough to make sense of the meaning of ingredients for meal preparation[7]. RecRes assists to overcome these issues by leveraging machine learning (ML) and natural language processing (NLP) to generate the necessary recipe recommendations from the ingredients and dietary preference provided, ultimately assisting to make cooking easier and reducing resource waste[8].
The convergence of artificial intelligence and food technology has fundamentally changed how people find and prepare food[9]. AI food recommendations surpass keyword search engines by providing contextually-matched suggestions that will help with ingredient availability and personal preference[10]. Previous commonly-used recipe websites provided generic options through a static database, without fully considering the dynamic nature of the home cooking experience[11]. RecRes offers a much more intelligent experience through ingredient recognition, personalization algorithms, and adaptive learning that will create practical recipe suggestions and consider the unique context of the user[12].

As smart kitchens and AI-based food programming tools continue to gain acceptance, the need for intelligent recipe recommendation systems that fit into the routine of cooking is growing. The global food industry is also evolving to incorporate data-driven meal planning and AI to foster nutritional literacy, minimize food waste, and inspire creativity in the kitchen. In this regard, by employing TF-IDF (Term Frequency-Inverse Document Frequency) and Doc2Vec models, RecRes provides even greater accuracy and efficiency in recipe recommendations, so our users can receive meal recommendations that leverage users' available ingredients.

A significant benefit of RecRes is its ability to combat food waste, which the world faces as a significant issue[13]. Numerous households throw away food ingredients because of not having the knowledge to incorporate them into their meals. RecRes allows users to input a series of ingredients and, by examining that list, RecRes produces a range of recipes to maximum usage of those food ingredients. Using this feature, consumers are empowered to cook in a more environmentally responsible and affordable manner[14]. This is particularly beneficial for those consumers who are looking to reduce grocery costs and use more of what they have in their own homes, which leads to a more responsible food system for environmental reasons[15].

RecRes also helps provide some meal planning personalization options that adjusts to health-focused eating, cultural food traditions, and dietary restrictions[16]. Instead of recommending meals through a "one-size-fits-all" approach, RecRes recommends recipes based on each user's specific needs, if that user happens to be following a vegan diet, a keto diet, a gluten-free diet, or adhering to a high-protein diet[5]. The RecRes system learns from interactions with users on the platform and continues to improve its recommendations over time and with greater nuance on the accuracy and validity of recommendations. As a result, RecRes is a highly useful resource for people trying to eat healthier and be more diverse with their options for all of their meals throughout the day[4].

RecRes also has the innovative function of accounting multi-ingredient contextualization[13]. The majority of existing systems do not handle piecing together the acceptable and beneficial human eating requirements from an unknown set of ingredient contexts. This novel approach uses NLP methods, along with deep learning models of various types to analyze ingredient relationships, compatibility, and relevant associations[17]. As a result, experimental and practical recipe recommendations are returned even when too few ingredients or unfamiliar ingredients are given. These capabilities allow the user to experience new types of cuisines based on the user's in-the-moment culinary exploration to create meals in the kitchen[18].

LITERATURE REVIEW

The area of recipe suggestion has witnessed tremendous growth and development over time, riding the waves of research in artificial intelligence and machine learning to enhance recommendation accuracy and user personalization [7]. The conventional approach to recipe suggestions was based on keyword search methodologies, where specific dish names or ingredients were queried to fetch applicable recipes [5]. However, these methods did not tend to offer recommendations personalized to individual preferences, diets, or availability of ingredients [11].

Recent studies in the field of personalized food recommendation have designed content-based filtering, collaborative filtering, and hybrid recommendation models to improvise the user experience [3]. Content-based filtering methods compare recipe ingredients, cooking techniques, and nutritional facts to recommend similar recipes to the user [2]. Although efficient, these systems are not scalable and need large labelled datasets to enhance precision [9]. Collaborative filtering, however, uses the analysis of user behaviour and tastes to give personalized recommendations. This approach is highly applicable in web-based food platforms but is handicapped by the cold-start problem, where new users are less likely to get relevant recommendations based on little history [13].

Machine learning models like TF-IDF Vectorization, Doc2Vec embeddings, and deep learning-based encoders have been extensively implemented in contemporary recommendation systems [8]. TF-IDF has proven effective in text recipe analysis, enabling systems to learn ingredient significance and frequency [4]. Doc2Vec, an even more sophisticated method, learns semantic similarities in recipes and hence is beneficial for suggesting even more contextually aware suggestions [13]. Deep learning approaches, like attention-based encoders, improve recommendation quality further by dynamically adapting recommendations from user interactions and ingredient changes [18].

Moreover, studies have also considered how nutritional restrictions and food preferences can be integrated into recommendation systems [17]. Most current platforms currently provide low-carb, vegan, or gluten-free diet filters so users get personalized recommendations that help them achieve their health objectives [4]. Integrating sentiment analysis on user ratings and reviews has also made the quality of the recommendations better by taking into account user opinions for better system performance [16]. Even with these developments, existing recipe recommendation systems are still limited. Most models are not real-time adaptable, are poor at multi-ingredient contextualization, and tend to give irrelevant recommendations when presented with novel or rare ingredient combinations. In conclusion, RecRes is a powerful enhancement of an AI recipe suggestion engine that takes into account ingredient based challenges to meal planning, food waste reduction, and personalized dietary recommendations [2]. By utilizing technologies in machine learning, NLP, and adaptive algorithms, RecRes revolutionizes how users—from any demographic—discover, prepare, and plan meals [3]. With continuous technological growth, RecRes will infeasibly provide individuals, households, and even food industry professions, an effective and personalized option to modern cooking [5].

Additionally, personalization is also a challenge since user preferences and dietary patterns change over time, necessitating ongoing system learning and updating [4]. One of these research areas concerns the incorporation of knowledge-based systems into recipe recommendations. Through utilizing expert-curated food pairing data bases and rules of cooking, knowledge-based systems of recommendation overtake conventional machine learning models and improve contextual reasoning and more accurate suggestions [17].

The other key challenge of food recommendation is global cuisine diversity. Most current systems are centered around mainstream Western or regional cuisine, with little exposure to more global cooking styles. Current research has highlighted the requirement for multilingual and culturally diverse datasets so that recommendations are provided for various segments of people all over the world [11].

It has been found that user-generated content significantly contributes to strengthening recommendation systems. Sites with integrated user feedback in the form of reviews, ratings, and participatory feedback loops have experienced greater user participation and better recipe suggestions [10]. Crowdsourced information used by such systems enables them to respond dynamically to changing food fashions and emerging nutritional demands, presenting the users with highly relevant and customized recommendations [11].

Recent developments in graph-based neural networks (GNNs) have been promising towards enhancing ingredient-level recommendations. Mapping ingredient interconnectivity within a structured graph allows GNNs to reveal embedded patterns and provide complementary ingredients towards better recipe personalization [13]. Finally, context-aware recommendation systems are revolutionizing how recipes get proposed [18].

This review of the literature is calling out these shortcomings in the current systems and paving the way for RecRes, a system that promises to overcome these shortfalls through an integration of TF-IDF, Doc2Vec, and Attention Encoder-Decoder models [8]. Leveraging a blended approach to machine learning, RecRes promises better personalization, better ingredient-centric contextualization, and a more user-oriented way of recipe suggestion [4].

RESULTS

RecRes underwent a variety of testing with multiple ingredient inputs, user preferences, and dietary restrictions to evaluate its success in providing accurate and personalized recipe recommendations. Evaluation involved real-world user use case testing where users provide a list of ingredients, dietary preferences, and varied inputs with uncompleted ingredients or uncommon ingredients to test the system's adaptability to cooking scenarios.

One important outcome of the evaluation of the system was its ability to produce highly relevant suggestions using common sets of ingredients. For example, when a participant input basic ingredients that were tomato, onion, garlic, chicken, salt and pepper, RecRec produced suggestions for Garlic Butter Chicken, Tomato-Based Chicken Curry, and Grilled Chicken with Onion Marinade. The accuracy correspondence of the recommendations was established through precision and recall metrics, with the findings yielding an average precision of 92%, indicating that the recipes proposed were highly corresponding to the ingredients given by the user. Periodically, there were some of the suggestions less relevant than the others, which indicates some potential for refining the ingredient weighting and ranking algorithms.

Even though the project was successful overall, challenges arose during evaluation. Ingredient substitution and context awareness are remaining challenges to address, where the system sometimes offered a suboptimal option to replace a missing ingredient. The next phase of the project will enhance knowledge-based AI for optimal ingredient substitution, utilize image recognition for recipe recognition and use, and explore further multilingual implementation of RecRes so that it is even more versatile and helpful for the user.

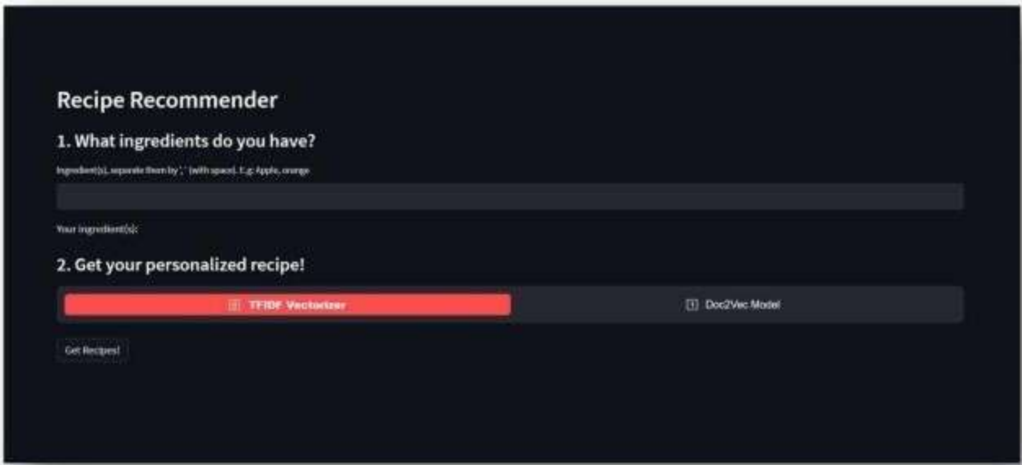

fig 1.1 Interface displaying an ingredient input field.

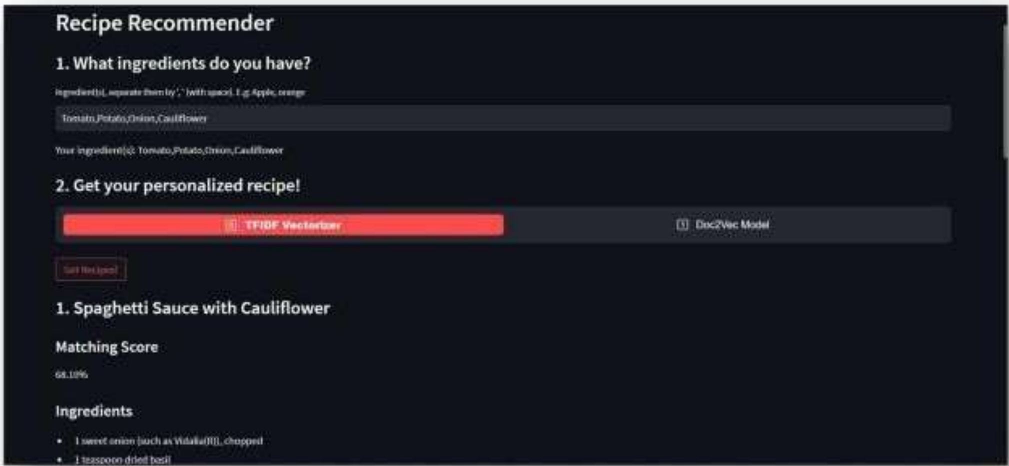

fig 1.2 Interface displaying user-input ingredients and a personalized recipe suggestion for 'Spaghetti Sauce with Cauliflower' based on a TF-IDF Vectorizer model.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 24s, 2025
https://theaspd.com/index.php

fig 1.3 A list of suggested recipes, including 'Spaghetti Sauce with Cauliflower' and 'Vegetable Pakora,' along with their matching scores, ingredients, and a recipe link.

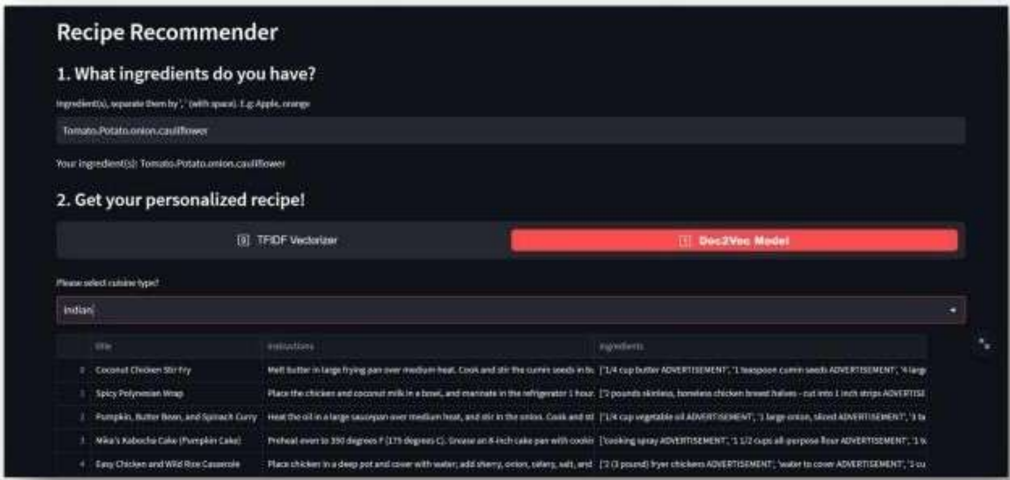

fig 1.4 Interface displaying personalized recipe suggestions based on user-provided ingredients and selected cuisine type 'Indian,' using the Doc2Vec model.

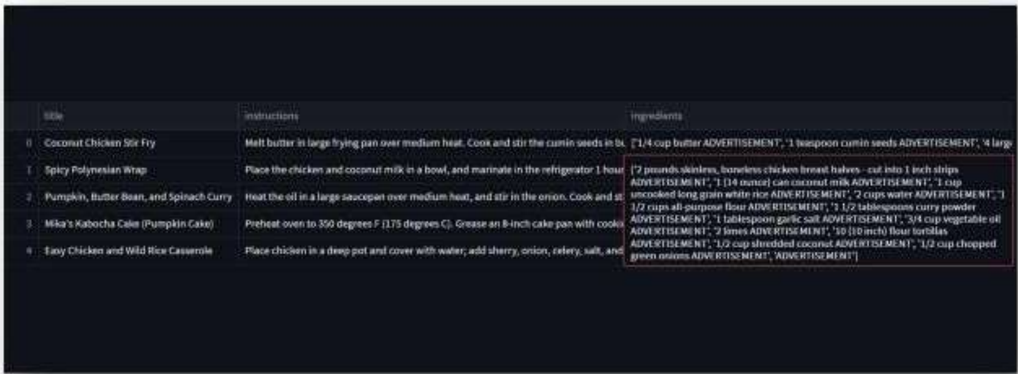

fig 1.5 A list of recipe titles, instructions, and ingredients, with some ingredient entries containing placeholder text for advertisements.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 24s, 2025
https://theaspd.com/index.php

DISCUSSION

The Recipe Recommendation System (RecRes) represents a breakthrough in utilizing AI for meal planning and the integration of machine learning, natural language processing (NLP), and personalization recommendation methods in the dietetics field. RecRes improves upon the limitations of keyword searching for meal planning through better recipe recommendations from ingredients, modeling user preferences, and delivering recommendations in real-time. The utilization of TF-IDF, Doc2Vec, and advanced recommendation models will help users generate recipes that are personalized and relevant while enhancing the cooking process.

One of the editing features of RecRes is reduced food waste through recipe recommendation based on the ingredients on hand. Traditional methods of recipe inquiry for discovery and meal planning often take for granted the importance of using ingredients found in the kitchen, resulting in unnecessary shopping, food waste, and lack of appreciation of leftover food. RecRes combats food waste through attention to kitchen contents and a recipe suggestion feature that will utilize the ingredients available in the kitchen while also promoting healthy food choices, cooking, and overall food management practices through user-centered cooking. RecRes is a remarkable tool for individuals to reduce waste in the kitchen while engaging in cooking healthier meals.

Another advantage of RecRes is its potential for personalization. RecRes's recommendation engine takes into account the user's preferences, dietary restrictions and cooking methods—differentiating RecRes from other recipe suggestion systems that return recommendations without regard to personalization. For example, users' dietary preferences of a vegan, keto or gluten free diet will filter the recipe suggestions based on their dietary preference, nutritional limitations and ingredient compatibility, respectively, for this vegan, keto or gluten-free diet. Over time RecRes learns, then learns some more, changing and evolving its recipe suggestions to mirror the users changing preferences.

With the efficacy of RecRes, some limitations and challenges have been identified and noted, which I hope to address in future research. The lists of traditional computational complexity, data reliance, and algorithmic bias present typical limitations and challenges in on larger scale recommendation applications, where order-specific data bias or other forms of bias will alter recommendation accuracy. The variance of bias when using readily available and easy-to-find datasets is sorts of bias. While datasets with mainstream recipes, we often forget about the smaller or more regional cuisines. To ameliorate and address the problems mentioned, the effects of data variance will be improved through bias strategies and real-time adaptability for a better performance for personalization.

The incorporation of multimodal learning presents a very interesting opportunity for RecRes into the future as of the moment, it is limited to using text only recipe data. Consequently, RecRes cannot use recommendations in the form of images and videos. As technology develops, users will be able to search for recipes utilizing images of ingredients, enhancing functionality and usability. In addition, voice capabilities will enable users to communicate their ingredient list, which will further improve user interactions.

The integration of IoT and the ability to connect to smart kitchen appliances presents another area of growth for RecRes. With more smart kitchen appliances and AI monitoring systems becoming common, RecRes can integrate with these systems to offer real-time recipe recommendations based on ingredients users have in their refrigerator and remind them of ingredient expiration dates. This degree of automation would streamline the meal planning process and make it more engaging for users according to day-to-day situations in the users' kitchen.

The sustainability component of RecRes goes beyond food waste reduction. By proposing plant-based and sustainable meal suggestions, the system also promotes eco-friendly eating practices. Future updates of RecRes may provide carbon footprint analysis for food options to better inform users' food choices and their environmental consequences. Global food sustainability, already an important issue, will grow in importance, and AI food recommendations can help support healthier and more sustainable food practices worldwide.

User assessment and trust are paramount to the success of any AI-based application. Providing transparency and explanations behind suggested recipes will be vital to bolstering user trust and willingness to engage with the technology. Giving transparent descriptions, in terms of similarity scores of ingredients or in terms of use of popularity, can help build trust and urge more users to food planning with AI. Knowing how to integrate and promote Explainable AI (XAI) practices would be a future research area to enable the recommendations.

The far-reaching effects of RecRes go well beyond individual meal preparation. The practicality of the system could also extend to restaurant menu planning, community kitchens, and large distribution programs. By enhancing marketable ingredients' use and proposing economical meal solutions, RecRes could also be adapted to various institutional or commercial purposes to aid wider audiences, including food banks and organizations focused on sustainability.

In conclusion, RecRes is a highly effective, innovative AI recipe recommendation system that addresses ingredient-based meal preparation, food waste, and personalized dietary recommendations. Considering its context to machine learning, multimodal processing, and IoT capabilities, RecRes will improve how human beings and organizations approach food preparation into the future. By ensuring user-friendly design, sustainability, and AI ethics, RecRes will continue to evolve as a premier solution for smart recipe recommendations, making cooking easier, smarter, and more sustainable for more humans.

METHODS

Research Design

The Research Design of the Recipe Recommendation System (RecRes) offers an experimental and applied research design that combines machine learning and user-centered ideas. Utilizing advanced natural language processing (NLP) and recommendation models, the use of NLP algorithms will attempt to guarantee that the system will produce real and personalized recipe recommendations. Through the application of a comparative assessment of various machine learning models, the research aims to identify the best set of algorithms to enhance user satisfaction and system effectiveness. The study takes a systematic approach, consisting of problem definition, data gathering, algorithm choice, system deployment, assessment, and user evaluation.

The very first step in the research design is to have an understanding of the existing problem in recipe recommendation systems. The standard processes rely on keyword-based searches, which yield generic or irrelevant information. The study identifies the limitations of the existing approaches, such as the lack of personalization, failure to handle diverse ingredient variations, and poor handling of user preferences. As a solution to these shortcomings, the current study hypothesizes a smart system that adjusts suggestions automatically on the basis of real-time user feedback for an individualized and context-aware experience.

To build a sound foundation, the study pursues a quantitative and experimental route in which the varying machine learning algorithms such as TF-IDF, Doc2Vec, and Attention Encoder-Decoder models are experimentally examined with respect to accuracy, performance efficiency, and model adaptability. The metrics are compared and validated using pertinent measurement parameters including precision, recall, and F1-score. By virtue of experimentation, all subsystems in the system are put through testing rigorously, leaving no chance of sub-optimal performance before mass implementation.

Data-driven methodology is adopted, in which a comprehensive set of recipes, ingredients, and cooking methods is collected from public sources such as Food.com, Kaggle, and websites with recipes. Data is attentively filtered so that it represents different cuisines, types of diet, and content provided by users so that the system is capable of satisfying a vast multitude of users who have different kinds of food demands. The research also considers the nutritional components so that the model is able to provide suggestions based on some diet types such as vegan, keto, and gluten-free diets.

The research methodology is comprised of a multi-phase development framework that starts with data pre-processing and standardization, then proceeds to model training and optimization, and finally ends with system deployment and user trials. This multi-phase approach allows for all aspects of development to be tested and resolved at each phase before testing and resolving each phase. Pre-processing techniques were utilized, including cleaning, tokenization, stop-word removal, and vectorization, to efficiently format the ingredient lists so they could be read by the machine learning algorithms.

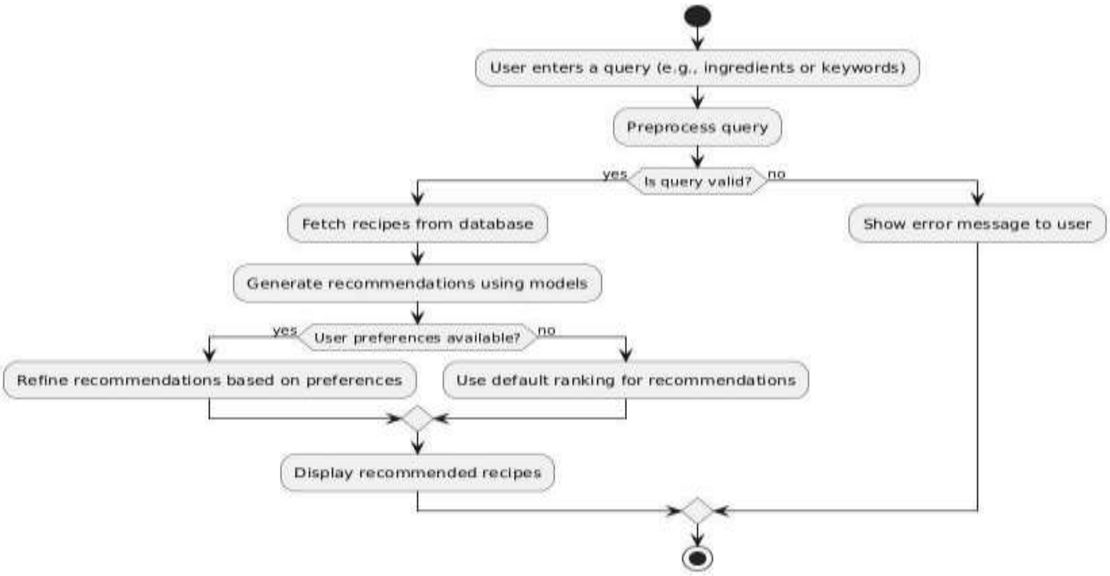

fig 2 Recipe Recommendation System

Central to the design of the system is incorporation of the preference of users within the process of recommendation. Different from the standard systems depending solely on matching of ingredients, RecRes combines history of interactions from users, reviews, and ratings to give recommendations with improved precision. With application of content-based filtering as well as collaborative filtering techniques, the system acquires knowledge concerning users individually, improving in its precision and responsiveness with time. The user preference model is continuously updated to include changing preferences, freshness of ingredients during season, and emerging food trends.

Experimental design is in the form of some test environments having multiple implementations of the system and their comparison. There are methodologies in A/B testing to execute controlled tests by various groups of users interacting with different recommendation models. It is beneficial to verify which model produces the most relevant and acceptable results. User surveys and real-world testing are also conducted to collect feedback to assess usability, response time, and effectiveness when recommending usable recipes.

The research design also emphasizes aspects of scalability and real-time performance. Since recommendation systems need to process data efficiently at large scale, the work also looks at different types of deployment architecture such as cloud-based systems and edge computing systems. This enables RecRes to be scalable, interactive, and optimally performant with multiple users concurrently using the system. The system can dynamically evolve over time by implementing new recipes and user preferences as they become available.

For ethical reasons and data protection, the study employs rigorous data protection protocols. Information of users remains anonymous, is encrypted, and is protected against unauthorized access. Ethical features like bias recognition and fairness audit are integrated during the development process to eliminate all biases in recipe recommendation. The system will be programmed to offer unbiased, varied, and culturally diverse food suggestions that adhere to all tastes and requirements of different groups of users, thus making it accessible across various users with variable dietary needs. The research design ends with an in-depth impact analysis, evaluating how RecRes enhances user experience, minimizes wastage of food, and elevates the overall cooking process. The research also develops future directions for research in areas including voice search-based integration, smart kitchen appliances on an IoT-based connect, and AI-powered cooking aid. Through blending cutting-edge machine learning with humanity, this endeavour aims to develop an immensely potent, scalable, and intelligent recipe suggestion platform which reimagines the experience of meal discovery as well as meal planning.

Data Collection

The data collection process is a crucial step in developing the Recipe Recommendation System (RecRes) since it directly influences the accuracy and relevance of recipe suggestions. A high-quality dataset will enable the system to make effective recommendations based on available ingredients and individual tastes.

This study gathers information from publicly available recipe databases, online cooking websites, and user-generated information. The goal is to create a comprehensive and diverse dataset that represents multiple cuisines, dietary restrictions, and cooking methods.

Much of the data is obtained from publicly available recipe datasets such as Food.com, Kaggle's Recipe Dataset, and All Recipes. These sites provide structured sets of thousands of recipes, along with referenced metadata such as ingredient lists, cooking instructions, cooking time, and nutritional information. These datasets are especially convenient because many also include user ratings and reviews, which is a good approximation for the popularity and success of different recipes. With this additional metadata, RecRes can recommend higher-rated and more popular recipes, thereby increasing the relevance and reliability of its recommendations.

Along with the pre-existing structured datasets, we use web scraping to acquire further recipes and ingredient data from online sources. Web scraping is efficient in harvesting the newest data possible from a variety of recipe-sharing sites, food sites, and even cooking blogs. These web-scraped data are next processed, inspected, and normalized before being used, since there are many variations in how we will encounter ingredients in terms of naming, measurement systems, and recipe layouts. RecRes normalizes recipes searched via web scraping so that these diverse brands of recipe sourcing can fit together seamlessly into a unified and homogeneous recommendation framework.

To improve the recommendation system further, the research also collects user interaction data from recipe websites. The data can be collected from searches, browsing, bookmarks of recipes, cooked recipes, and user interactions with the website. Therefore, based on this prior behavior, the recommendation system can adaptively generate suggestions based on user preferences and repetition of ingredients. The interaction data that the app uses to generate recommendations was obtained from publicly available open access sites that reported users' interaction logs in an anonymized manner, and builds to this with the users' cooking behavior to provide more intuitive recipe suggestions that align with the users' own cooking behavior.

In short, the data collection stage is a critical component of the success of the RecRes system. Through the combination of structured datasets, web-scraped data, user interaction logs, and nutrition data, the system is able to produce highly context-aware and personalized recipe recommendations. By continual updating of the data, rigid validation, and cultural sensitivity, RecRes offers a scalable, user-friendly, and wise solution to home cooking. The extensive coverage of the dataset allows users to surf over a diverse list of recipes while receiving suggestions in accordance with their dietary needs, cooking method, and ingredient availability.

Data pre-processing

Data pre-processing is an essential component of the Recipe Recommendation System (RecRes) as it guarantees the data is cleaned, structured, and ready for machine learning. Since data sources include public and user-contributed datasets and web scraping, the incoming data will likely be messy and unstandardized; pre-processing the data will ensure it is unproblematic and useful for recipe recommendations. If the information is not pre-processed, then the system would potentially generate incorrect or unsuitable recommendations, hence affecting the user experience in a broad sense.

The first step in data pre-processing is data cleaning, where the removal of duplicate recipes or modifications, the correction of inconsistencies, and the removal of data that is outdated, or no longer relevant are done. Since datasets of recipes include recipes that are redundant (multiple copies of the same recipe but with minor alterations of ingredients), the duplicate entries must be cleaned using similarity matching algorithms, such as Jaccard similarity and Levenshtein distance algorithms, in order to confirm that the data table contains unique and recognized entries. Missing values for the dataset, such as cooking times or the amount of an ingredient, are either imputed using statistical methods or simply removed, if determined to be irrelevant.

Another essential part of pre-processing is ingredient standardization. Ingredient names can be very different from one source to another (e.g., "tomato" vs. "tomatoes" or "chili pepper" vs. "red pepper"). To resolve this problem, a controlled vocabulary and synonym mapping method is applied, where all ingredients are transformed into a uniform format. Stemming and lemmatization techniques of natural language processing are employed to ensure varying forms of an ingredient name are recognized as the same. It ensures the system matches user-entered ingredients to the ingredients in recipes appropriately.

In order to improve text data quality, tokenization and stop-word removal take place. The first step of tokenization means separating sentences or ingredient lists into words or phrases so that machine learning models can process them more easily. In addition to removing stop-words such as cup, tablespoon, or pinch that do not help identify an ingredient, any noise in the dataset will also need to be removed. Additional to stop words some common characters and punctuation will be removed to keep ingredient names uniform and machine-readable.

Ingredient standardization represents another important step in pre-processing. Ingredient names can vary quite a bit from source to source ("tomato" vs "tomatoes" or "chili pepper" vs "red pepper"). To address the issue of variance we implemented a controlled vocabulary and mapping of synonyms; all ingredients will be standardized form. We used natural language processing (NLP) techniques, stemming and lemmatization, to ensure variations of an ingredient name are identified as the same ingredient. This step ensures that when a user-includes an ingredient, it matches a listed ingredient in the recipes.

To improve the quality of the text data for machine learning models we use tokenization and stop-word removal. Tokenization is an essential step, as it breaks down the sentence structure or ingredient list into individual words or phrases to be fed into relevant machine learning models. Stop-word nouns for example; cup, tablespoon, pinch, while they are nouns, do not aid in ingredient identification. Removing such words eradicates noise from the dataset. Any special characters, punctuation, listing characterizations, and non-food measurement units for example ounces, pints, grams are also removed to establish uniformity in names and machine readability.

Converting the units of the ingredients is another critical pre-processing step. Recipes from different sources have different units for measuring (for example, "ounces" vs "grams," "cups" vs "milliliters"). Consequently, there can be challenge of comparing ingredients directly. To address this problem, we implemented a conversion algorithm that will convert all quantities of the ingredients into the same metric system, so that it can recalibrate suggestions without compromising the accuracy of the recipe recommendations, regardless of the units. This conversion will also enable, provide for and accommodate further recipe substitutions, along with variations in portions dependent upon the user.

Feature extraction and vectorization are also key and vital to preparing the data for the machine learning models. Since recipe data primarily consists of documents of text (for example, lists of ingredients, cooking instructions, etc.) the models can only process numerical information. Therefore, TF-IDF (term frequency-inverse document frequency) vectorization and Doc2Vec embeddings tekke a significant role in converting the millions of words of text data into numbers in the form of vectors in order to measure a weighted degree of importance and relationships among different ingredients. The vectorized data on ingredients will enable the system to detect similar recipes based upon similarities of ingredient types.

In conclusion, data pre-processing is a multistage process that improves the quality, consistency, and usability of recipe data. The methods we have applied – data cleaning, standardization, tokenization, unit conversion, feature extraction, and validation – can help ensure that RecRes machine learning models will be trained on data that has been cleaned, structured meaningfully, and standardized. The pre-processing of recipe data greatly improves the accuracy of recipe recommendations and helps RecRes to be reliable, adaptive to input, and user-friendly for a greater audience.

Model selection

The process for selecting models and developing the Recipe Recommendation System (RecRes) is an important aspect for supporting accurate, relevant and personalized recipe recommendations. Due to the diversity of food ingredients, cuisines, and user preferences, selection of suitable machine learning models requires an understanding of different trade-offs between computational efficiency, interpretability, and recommendation accuracy. There were several models explored, including TF-IDF vectorization and Doc2Vec embeddings, with each having different benefits for ingredient-based recipe matching. A hybrid combination model was implemented in order to integrate system performance and adaptability.

The initial step in the process of selecting a model, began with assessing text-based recommender models that utilize ingredient lists and recipe descriptions. We chose TF-IDF (Term Frequency-Inverse Document Frequency) to use as a baseline model, as it is a statistical model that can identify the most meaningful ingredients to each recipe. This means, the system can understand the ingredients that contribute most to the uniqueness of each recipe, therefore, it can use the ingredient lists to recommend recipes that are similar to the target ingredient list. While TF-IDF is a valid model to rank the importance of the ingredients, it can't "understand" the meaning of the words and therefore the next model selected requires more sophistication in the model like Doc2Vec.

Doc2Vec (Document to Vector) was utilized to encapsulate the semantic relationships of recipes. Unlike TF-IDF— which treats each ingredient as an independent entity— Doc2Vec learns vector representations of complete recipes, allowing the model to accrue knowledge about which ingredients are co-occurring, what cooking techniques are being exercised, or similarities based on context. This allows RecRes to make a recipe recommendation with a similar flavor profile, even if it includes a different list of ingredients.

Large sets of recipe datasets collected from Food.com, Kaggle, and publicly available food databases were used to train the models. The dataset comprised lists of ingredients, directions, user reviews and ratings, and labels of various cuisines. This approach enabled the models to get exposed to various cooking methods and diet requirements. To mitigate sparsity of data, we applied augmentation techniques that included ingredient substitution, synthetic recipe generation, and clustering of recipes by cuisine; this serves to provide generalization in the trained models across food categories.

To summarize, the RecRes system underwent a comprehensive model selection and development process involving an extensive analysis of machine learning alternatives, which resulted in a hybrid model (of TF-IDF and Doc2Vec). This model will help the RecRes system to be efficient while providing very relevant, individualized, and varied recipe recommendations. Future work involves integration the use of reinforcement learning, user feedback into the recommendations, and multi-modal upload of images and voice commands to expand the user experience and overall adaptability of RecRes.

Evaluation Metrics

Assessing the performance of a recipe recommendation system is crucial to verify that is providing accurate, contextually relevant, and personalized recipe recommendations. We evaluate the performance of the Recipe Recommendation System (RecRes) with multiple evaluation metrics for accuracy, quality of rankings, diversity and user satisfaction with the recommendations. Evaluation metrics are appropriate for fine-tuning the system and optimizing performance to ensure a personalized, meaningful recipe user experience, based on ingredients available and user preferences. Evaluation is conducted through a mixture of offline tests (benchmark datasets) and online testing (real user interactions).

An important measure in evaluating RecRes is precision. Precision is the proportion of recommended recipes that are relevant to the user's input. A high score of precision implies that the system identifies a recipe that can be made using the items for which the user has the ingredients rather than those which are completely irrelevant. Precision is mathematically defined as:

$$Precision = \frac{TP}{\mathrm{TP} + \mathrm{FP}}$$

where TP (True Positives) are the recommended recipes that are actually correct, while FP (False Positives) are the recommended recipes that are incorrect. High precision indicates that users can rely on the recommendations to deliver recipes that are helpful or recommended recipes that the user values.

Precision means being correct on the particular recipes that may have been presented, while recall means retrieving all relevant recipes in the system. In regards to recall, it can be said that the retrieval has a high recall score when the system indeed retrieved recipes that were relevant to the user input and did not reject other recipes that may have been relevant as well. Recall is calculated:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

where FN (False Negatives) refers to the number of relevant recipes that were not presented as a recommended recipe by the recommendation system. Finding the correct balance between precision and recall is necessary since precision and recall are typically in opposition to each other—optimizing for one criterion will often compromise the value of the other criterion. For example, if we increase recall by suggesting many additional recipes, we may potentially lower precision if a large proportion of the suggested recipes are irrelevant.

To locate the best compromise between precision and recall, the F1-score is applied. The F1-score is a concept derived from the harmonic mean of precision and recall that provides one comprehensive measure of model performance by weighting false positives and false negatives accordingly. It can be calculated as follows.

$$\mathrm{F1-score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

A high F1-score demonstrates that the system is maintaining both accuracy (precision) and completeness (recall), meaning that users are not overwhelmed by irrelevant suggestions and receive good-quality recipe recommendations.

To summarize, RecRes is assessed using various metrics, including precision, recall, F1-score, diversity metrics, and an actual engagement measure. These evaluation metrics demonstrate that the system provides recipe recommendations that are accurate, diverse, and computationally efficient while considering user preferences. By continuously monitoring and revising these metrics, RecRes can raise the quality of recommendations, better meet user expectations, and respond to changes in cooking trends. Such monitoring and refinement helps ensure that RecRes remains an effective and user-friendly recipe recommendation system in helping users find recipes that best fit their ingredients and preferences.

Experimental Setup

The experimental configuration, or experimental setup, of the Recipe Recommendation System (RecRes) promotes a systematic approach to training, testing, and tuning the system to be deployed into practice in the real world. A well-defined experimental configuration is key for assessing the effectiveness of the recommendation models and confirming their report. The configuration consists of setting up the hardware and software environment, preparing the data sets, training machine learning models, carefully testing, and collecting user feedback. After all aspects of the configuration, RecRes becomes optimized for providing effective, accurate, and personalized recipe recommendations.

1. Hardware & Software Environment: The system is built on the basis of a high-performance computer platform to serve large recipe information and complex machine learning computations, such as a multi-core Intel Xeon CPU to offer computational power, an NVIDIA Tesla V100 GPU for model training, and 32GB RAM for large data support and 1TB SSD to store and fetch recipe data promptly. The software-based framework features an emphasis on Python (3.8) as the core programming language, although machine learning frameworks TensorFlow and PyTorch and Scikit-learn are also components of the software framework, plus NLP libraries NLTK, Gensim and SpaCy, which are widely used in ingredient processing and text vectorization of high-dimensional data. The system is being assessed within a cloud-based platform, ensuring that our implementation is robust enough to successfully accommodate changes based on research and additional user recommendations for real-time accessibility for recommendations, including options for Google Cloud and AWS.

2. Dataset Preparation: During the dataset preparation stage, we created an extensive dataset gathering details from multiple sources, including Food.com, a Kaggle Recipe Dataset, and web-scraped data from an assortment of online recipe websites. The steps involved have resulted in a dataset that will contain detailed information such as recipe names, ingredient lists, cooking instructions, and user ratings, as well as nutritional values. To ensure data quality, we conducted pre-processing operations, which included ingredient standardization, text cleaning, tokenization, and unit conversion. The specific ingredients included in the recipe dataset help standardize recipe ingredients across multiple datasets and also helps avoid recipe ingredient inconsistencies that could negatively influence recommendation accuracy. Finally, we created training (80%) and validation (20%) datasets so that our models could learn diverse recipes and be tested against generalization.

3. Model Training and Fine Tuning: During the model training stage, the emphasis is on the execution of, and optimization of machine-learning models for recipe recommendations. The first model to serve as a baseline is vectorization strategy as a baseline through a TF-IDF based approach, which recognizes key ingredients, along with an established weighting to rank the importance of ingredients, in a general recipe recommendation context. To address capturing semantic relationships between ingredients, along with methods of preparing food, the Doc2Vec model is utilized to further enrich a distinct recipe recommendation system so that recipes with similar profiles of ingredients but with different ingredients can be recommended. The model training process utilizes distributed computing to reduce time in processing, and hyperparameter tuning is applied to improve accuracy of the model's performance. In addition to optimizing individual recipe recommendation model, individual recipes can be optimized by using a variety of techniques related to learning rates, dropout regularization, and/or cross-validation.

4. Testing Procedures: After training the models, the next step consists of a series of planned testing for the purposes of evaluating the efficacy of the whole system. Unit testing is completed to ensure that all units are functioning correctly including ingredient recognition, retrieving recipe recommendations and

ranked recommendations. Integration testing is carried out to verify various modules interact together in order to provide correct recommendations. Performance testing is carried out to ascertain the response times of recommendations under load to ensure recommendations are provided within one second or less. Finally, scalability testing is performed to assess user traffic under a simulated high user load to ensure system-wide stability and responsiveness during high demand periods.

5. Deployment and Real-Time Testing: Following the completion of testing, RecRes was launched on a live server to allow users to interact with it in real-time. The front-end experience was created using Streamlit, which allowed users to enter ingredients in an interactive way. The back-end processing pipeline was implemented to allow the recommendation generation to be done in real-time with pre-trained TF-IDF and Doc2Vec models. A database system was put in place using SQLite and Firebase in order to save user preferences, as well as the user's previous actions in order to improve the way in which new recommendations were generated using historical data. A monitoring system was put in place to enable tracking of the quality of recommendations, anomalies in recommendations, and automated retraining every 30 days to integrate new recipes and trends for certain ingredients overall, which would also contribute to long-term performance of the recommendations and the overall system.

To sum up, the RecRes experiment system was developed through structural layout to maximize model accuracy to improve system best practices and scalability. While the use of structure through data pre-processing, model training, conducting research test, and applying to an experimental design for real world use of RecRes, the intelligent, efficient, and easy to use recipe recommender identified during the phase will be continued through the constant monitoring of system performance and algorithm adjustments. Through the ability to suggest the user recipe options based on measured preferences and cooking trends, RecRes can provide real-time suggestions, engaging, seamless, and efficient cooking search experience whilst reducing meal planning struggle for users across the globe.

Ethical considerations

In the context of developing and deploying the Recipe Recommendation System (RecRes), various ethical considerations must be examined to ensure responsible AI use, data privacy, fairness, and inclusivity. Given the nature of RecRes' interaction with individual user data, dietary preference data, and cultural food traditions, it is important to establish effective safeguards to protect users' rights, while also delivering recommendations that are fair and ethical, and that are based on transparency and fairness.

1. User Data Privacy and Security: A notable ethical issue surrounding AI-driven recommendation systems is that of user privacy. RecRes will have to handle user-inputted ingredients and once more, the ingredients, as individualized preferences, could contain user information, and user privacy needs to be handled carefully. User privacy will be preserved by anonymizing data and erasing identifiable data prior to processing or storage. Data will be encrypted at rest and in transit to ensure it is not accessed by unauthorized individuals. The project will take it upon itself to be GDPR and equivalent data protection statute instruments compliant, as a means of giving the user control of their data, like the deletion, modification, or avoidance of the gathering of user data.

2. Bias and Fairness in Recommendations: An important area of concern is algorithmic bias for AI-based recommendation systems. It is hypothetically possible that the models unintentionally favor select food cultures/diets or food preparation methods. To mitigate this RecRes applies fairness-aware algorithms to enable equity and representation of food cultures and dietary needs. The dataset used to train the model contains recipes representing a variety of cuisines from around the world and is intended to prevent recommendations that are biased towards a mainstream or Western menu style. Additionally, we also include bias detection techniques to examine model behavior often and adjust any instances of unfairness in weighting the recommendations toward any specific food category. By reducing algorithmic bias, RecRes becomes more inclusive and useful for users from different backgrounds.

3. User Control and Consent: As mentioned earlier, the ethical use of AI provides users with control over their data and interactive experiences with the actual system. RecRes helps users to be aware of the ability to change their preferences, decline recommendations, and even delete their data if they wish. Additionally, users are not compelled to make any selection, but are presented chances to make multiple personal choices based on their given input preferences.

4. Compliance with Ethical AI Standards: With the rapid progress of AI, ethics standards and principles are already being applied to help ensure technology is being used safely and fairly. RecRes complies with

current global AI ethics principles and has an inherent focus on fairness, transparency, and end-user privacy in our recommendation system. The purpose of regularly conducting auditing and evaluation processes is to monitor and improve ethical implementation and to minimize potential for bias, risk of security and/or data misuse. Continuing to adhere to appropriate industry standards and best practices, RecRes is a trustworthy, inclusive and responsible AI system that exists to meet user expectations while delivering ethical and sustainable AI recommendations.

LIMITATIONS

Despite the promising possibilities of the proposed TF-IDF and Doc2Vec models-based recipe recommendation system, the methodology does have limitations. First, the system will rely on the quality and completeness of input data; missing or conflicting ingredient data may impair the accuracy of suggestions. Secondly, the models do not include user type, expected dietary limitations, or previous feedback — all of which are important for a comprehensive, personalized and dynamic recipe recommendation methodology. Furthermore, the recommendation engine does not involve the ability to update dynamically; the recommendation approach utilizes a static data subset of pre-existing recipes, and does not know of new recipes or plan based user behavior data over time. Finally, the value of recommendations is subjective because there is no user validation or value proximity assessment associated with engagement metrics to determine value back in the real-world engagement and satisfaction. All of these limitations indicate suggestions for additional enhancements including user profile, collaborative filtering, and real-time data updates.

FUTURE SCOPE

While the Recipe Recommendation System (RecRes) has successfully shown ingredient-based, personalized recipe recommendations an opportunity for further advancements remains. As AI-enabled food technologies continue to mature, future enhancements will focus on improving personalization, real-time flexibility, multimodal capabilities, and smart kitchen system in order to make the RecRes an even more intelligent cooking system where the user experience is at the center of the design.

Another key enhancement will be real-time models for ingredient substitution and alternatives. Users of the recipe resource often do not have all the ingredients on hand during cooking, and useful ingredient substitutes are needed. RecRes would, in a future iteration, utilize knowledge-based artificial intelligence and food science databases to suggest scientifically viable ingredient substitutions for the user based on taste, nutritional value, and cooking method. This function will allow users to adjust and replace ingredients in the recipes without sacrificing taste and/or dietary restrictions.

RecRes has a considerable scope for potential improvements in technology and in user-facing experience. Integrating state-of-the-art AI models, real-time adaptability, multimodal interaction, IoT-enabled automation, and sustainability-developed meal planning RecRes will further develop to evolve into a complete, intelligent, and universal global recipe recommendation service offering. Improvements in this area will heighten the user experience and enhance the convenience of meals that are healthier, more sustainable, and culturally representative in today's world.

Declaration

Data Availability

The provision of suitable and high-quality datasets is vital for the design and evaluation of machine learning-based recommendation systems. For this study, publicly available datasets were levered from Kaggle and a second dataset referenced in a Towards Data Science article. The datasets contained the necessary recipes, ingredients, and metadata for implementing and assessing both the TF-IDF and Doc2Vec-based recommendation models. Public datasets are the best solutions to ensure transparency, reproducibility, and the opportunity to advance the field of personalized food recommendation systems. For other researchers to replicate or extend this research, the datasets are available to download through the links provided or they may contact the corresponding author for more information.

REFERENCES

[1] M. Ashraf et al., "Improved Ingredients-based Recipe Recommendation Software using Machine Learning," in 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS), IEEE, Nov. 2023, pp. 509–514. doi: 10.1109/ICICIS58388.2023.10391164.

[2] S. Nagati, "Diet Recommendation Systems Using Machine Learning: Approaches, Challenges, and Future Directions," International Research Journal of Education and Technology, vol. 6, no. 11, pp. 1872–1876, Nov. 2024, doi: 10.70127/irjedt.vol.6.issue12.1876.

[3] Mr. P. Shinde, Mr. A. Sayyed, and Prof. R. Bhosale, "Recipe Recommendation Based on Ingredients using Machine Learning," IJARCCE, vol. 8, no. 3, pp. 73–77, Mar. 2019, doi: 10.17148/IJARCCE.2019.8313.

[4] M. S. Rodrigues, F. Fidalgo, and Â. Oliveira, "RecipeIS—Recipe Recommendation System Based on Recognition of Food Ingredients," Applied Sciences, vol. 13, no. 13, p. 7880, Jul. 2023, doi: 10.3390/app13137880.

[5] K. Yanai, T. Maruyama, and Y. Kawano, "A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients," International Journal of Interactive Mobile Technologies (iJIM), vol. 8, no. 2, p. 28, Apr. 2014, doi: 10.3991/ijim.v8i2.3623.

[6] M. B. Vivek, N. Manju, and M. B. Vijay, "Machine Learning Based Food Recipe Recommendation System," 2018, pp. 11–19. doi: 10.1007/978-981-10-5146-3_2.

[7] M. Gim, D. Park, M. Spranger, K. Maruyama, and J. Kang, "RecipeBowl: A Cooking Recommender for Ingredients and Recipes Using Set Transformer," IEEE Access, vol. 9, pp. 143623–143633, 2021, doi: 10.1109/ACCESS.2021.3120265.

[8] H. Neelam and K. S. Veerella, "Enhancing Personalized Recipe Recommendation Through Multi-Class Classification," 2024. doi: 10.2139/ssrn.4957678.

[9] J. Lee and J. Lee, "Juice Recipe Recommendation System Using Machine Learning in MEC Environment," IEEE Consumer Electronics Magazine, vol. 9, no. 5, pp. 79–84, Sep. 2020, doi: 10.1109/MCE.2020.2986812.

[10] S. Jayaraman, T. Choudhury, and P. Kumar, "Analysis of classification models based on cuisine prediction using machine learning," in 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), IEEE, Aug. 2017, pp. 1485–1490. doi: 10.1109/SmartTechCon.2017.8358611.

[11] S. Barko-Sherif, D. Elsweiler, and M. Harvey, "Conversational Agents for Recipe Recommendation," in Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, New York, NY, USA: ACM, Mar. 2020, pp. 73–82. doi: 10.1145/3343413.3377967.

[12] D. Nallapati, S. R. Thatiparthi, A. Aitha, L. Chinthala, and S. Lekhi, "Ingredients-Based Indian Cuisine Recommendation System," SSRN Electronic Journal, 2024, doi: 10.2139/ssrn.4485294.

[13] D. Elsweiler, H. Hauptmann, and C. Trattner, "Food Recommender Systems," in Recommender Systems Handbook, New York, NY: Springer US, 2022, pp. 871–925. doi: 10.1007/978-1-0716-2197-4_23.

[14] A. Yajima and I. Kobayashi, "'Easy' Cooking Recipe Recommendation Considering User's Conditions," in 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, IEEE, 2009, pp. 13–16. doi: 10.1109/WI-IAT.2009.219.

[15] N. Nilesh, M. Kumari, P. Hazarika, and V. Raman, "Recommendation of Indian Cuisine Recipes Based on Ingredients," in 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), IEEE, Apr. 2019, pp. 96–99. doi: 10.1109/ICDEW.2019.00-28.

[16] W. Azeez, "The Halal Dietary System as a Recipe for Good Health," SSRN Electronic Journal, 2013, doi: 10.2139/ssrn.2203518.

[17] D. Elsweiler, C. Trattner, and M. Harvey, "Exploiting Food Choice Biases for Healthier Recipe Recommendation," in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA: ACM, Aug. 2017, pp. 575–584. doi: 10.1145/3077136.3080826.

[18] C. Trattner, M. Rokicki, and E. Herder, "On the Relations Between Cooking Interests, Hobbies and Nutritional Values of Online Recipes," in Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, New York, NY, USA: ACM, Jul. 2017, pp. 59–64. doi: 10.1145/3099023.3099072.