

# Personalized Food Recommendation System by using Machine Learning Models

M.S.N.V. Jitendra<sup>1</sup>, Maddula Lakshmi Jyosna<sup>2</sup>, Sai Sri Varsha Veeraghanta<sup>3</sup>, Shanmuk Srinivas A<sup>4</sup>, K Bhargav<sup>5</sup>  
<sup>1,2,3,4,5</sup> Department of Computer Science and Engineering, GITAM School of Technology,  
GITAM University, Visakhapatnam, AP, India

**Abstract:-** A recommender system's primary function is to make recommendations to users. So, in this study, we developed a "Food Recommendation System" that suggests foods based on the age provided and a rating greater than 4. This work connects users to recipe ideas, ingredient lists, and cooking times, serving as a hub for kitchen-related information. This study has the dataset with 12 features, namely name of the dish, style of cooking, category, calorie count, flavor profile, cooking skills, course, healthy, preparation time, ingredient count, rating and age, where data cleansing is done and is divided into 80/20 training and test data, respectively. Decision trees, random forest, K-Nearest Neighbour, and logistic regression models were applied to the training dataset, and the expected value for the test dataset was then produced. For each of these models, a combined confusion matrix and categorization report were created. Based on the confusion matrix and classification report numbers, validate the results with Precision, Recall, F1-score, and Accuracy for each model. Calculated the misclassification rate for each algorithm and showed the most accurate model. This dataset consists of 488 records.

**Keywords:-** Decision Tree, K-Nearest Neighbour, Logistic Regression, and Random Forest.

## I. INTRODUCTION

They all consider what to eat, where to eat, how much dietary worth a food has, and whether it will help them lose weight or build muscle, among other things.<sup>[1]</sup> Thus, this method enables the user to make decisions in these complex information spaces quickly. A meal recommendation is a personalized service that works wonders for helping with diet planning. This recommender may abuse the meal's nutritional value to justify its suggestions. The application aims to provide users with a platform to learn about their favorite foods' nutritional values and find new foods to try. Anyone who wants to lose weight is worried about their health, or is dealing with a particular health issue can gain something from this. Systems for calorie counting are getting much attention as they have begun to replace the old with and stylus techniques. These programs offer services and educational resources that influence users to change their behavior [1]. Due to the widespread use of these calorie-counting offices, these systems have access to a vast amount of client preference data that can be used to tailor intuitive features and boost user engagement with the system and the slim-down plan [2]. The user of this food recommendation app can find their favored foods and learn more about their nutritional value in light of any health conditions or

particular diseases they may be experiencing. This is done by looking at the foods that contain one of the main components and how healthy they are rated [3]. The recommender system is the foundation of this endeavor. Tools that use supervised learning for recommendations. When using supervised learning, we already know the user's needs and use our ML model to make recommendations [4]. Nutrition is getting the food and nutrients needed for development and good health. We become weaker, get sicker, and so on. In the worst-case scenario, we might die if we do not consume [5]. We cannot regularly subject our bodies to taxing mental and physical activities and fall short of developmental benchmarks. An ideal diet must be balanced for nutrition and well-being. It protects against chronic, noncommunicable diseases like diabetes, heart disease, and cancer [6]. All living things, including humans, plants, animals, and other creatures, depend on them to survive. Nutrition recommendation systems are one of the most widely used technological innovations in nutrition knowledge. (NRS). They are being researched as a potential tool to assist users in forming better eating routines and achieving their goals [7]. A recommender system can anticipate users' preferences for unrated goods and recommend new products.

### A. Activity Diagram

The progression of one action through a system or process is shown in an "activity diagram," a flowchart in the Unified Modeling Language (UML). It is used to describe the different dynamic components of a system and is known as a "behavior diagram" because it specifies what should happen in the modeled system [8].

In this study, the initial action for a user is to log in and input an age range, as shown in Figure 1. The food items are sorted, after which suggestions are generated.

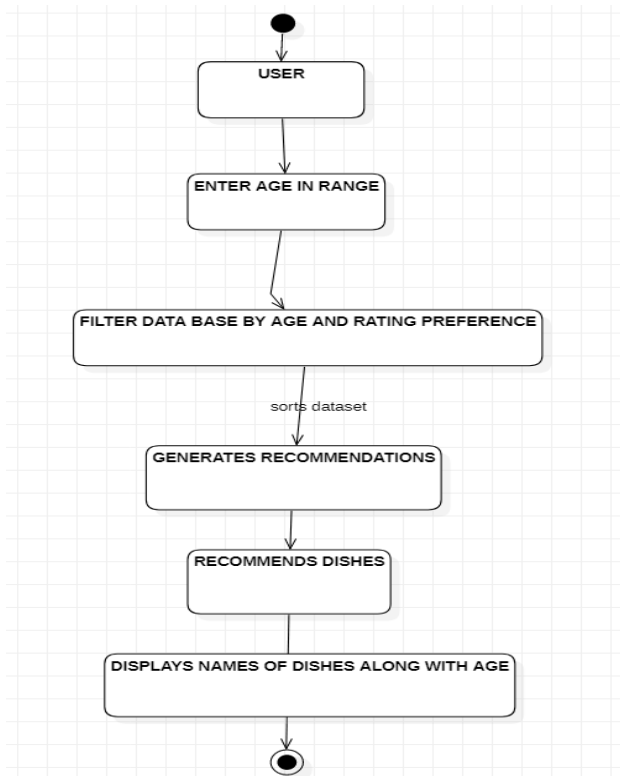


Fig. 1: Activity Diagram

## II. LITERATURE REVIEW

The study articles taught how recommender systems filter out the necessary information and offer a suggestion to a user based solely on their current needs. In this personalized and non-personalized recommender system, they are essentially types. In personalized recommender systems, the information is first gathered from the user, for instance, by conducting surveys or distributing a questionnaire to understand the user's likes and dislikes and recommendations [9]. In contrast, in non-personalized recommender systems, the information is given based on observation of likes. Also, the top trending hits at the time, for example, Netflix displays the top 10 shows in each region, and one more type is based on the one thing that is watched. Others are decided using machine learning algorithms, categorized, and displayed as similarities [10].

According to Anna Gatzoura and Miquel Sanchez, the recommendation system aims to provide users with valuable and pertinent material while actively using the site. (items).[11,12] Recommendation systems have become increasingly popular in recent years. In the middle of the 1990s, the first publication on collaborative filtering appeared after recommendation systems emerged as a popular research subject. To filter and retrieve info, the recommendation system technology is used [13]. These technologies also help e-commerce websites and other networks increase their sales. These systems are a piece of work that helps customers find the products they prefer and connects them to the services and products they are interested in [14]. Recommendation systems are a type of machine learning technology that falls under unsupervised machine learning models, in which the data is not categorized [15].

We are currently in an information age. Data are abundant in the form of comments on numerous websites, reviews, blogs, publications, etc. Since 1995, there has been a 40% growth in the number of people using the internet globally and reaching 3.2 billion in number. More opportunities have emerged due to the increased information flow, yet it has also confused the user. In the midst of all of this massive data, the process of making certain decisions can be challenging. The ad age that one should make an educated decision is accurate. However, having too much knowledge can sometimes make it difficult to make decisions. Recommender systems were thus developed to relieve users of this uncertainty and improve their online browsing experience. Lior Rokach, Bracha Shapira, and Francesco Ricci developed fine recommender systems as work that provides users with suitable suggestions.

## III. MATERIALS AND ALGORITHMS

The technologies used in our study are NumPy, pandas, sci-kit-learn, confusion matrix, and K-Nearest Neighbor.

### A. Numpy

A popular Python package for numerical processing is called NumPy. It works exceptionally well with arrays and matrices. It offers multidimensional arrays and different mathematical functions that can be used on them in a high-performance implementation. Due to their contiguous memory allocation and ability to vectorize operations, NumPy arrays are more effective than Python lists, making them ideal for scientific computing, data analysis, and machine learning apps.

Numerous linear algebra operations are available in NumPy, including matrix multiplication, decomposition, and eigenvalue computations. A fast Fourier transform (FFT) tool for signal processing and spectral analysis is also included in NumPy. The work is extensively used in various industries, including data science, engineering, finance, and scientific computing.

Several built-in functions in NumPy, such as reshaping, slicing, and indexing, make manipulating and changing array data simple. For additional analysis and visualization, they can also be coupled with other data structures, such as pandas DataFrames. NumPy is a vital instrument for Python's effective and efficient numerical computing.

### B. Pandas

A popular Python tool for data analysis and manipulation is called Pandas. It offers flexible, high-performance data structures like Series (1-dimensional) and DataFrame (2-dimensional) for managing tabular data. NumPy, which Pandas is based upon, enables the practical computation of sizable datasets.

Filtering, combining, grouping, and pivoting are just a few of the functions that Pandas offers for cleaning, transforming, and analyzing data. The library also provides solutions for managing duplicate data and handling missing

data. Pandas is a well-liked option for data analysis across a variety of sectors because it is made to be flexible and intuitive.

Pandas support loading data from various sources, including CSV and Excel files, SQL databases, and online APIs. [5]Pandas offers solid data analysis and visualization tools such as summary statistics, scatter plots, and histograms, once the data has been loaded.

Additionally, Pandas is very extendable and configurable. Pandas' apply technique allows users to build and apply custom functions to their data. Pandas also works with other Python libraries like Matplotlib and Seaborn for advanced visualization and statistical analysis.

Pandas is a well-liked option among data scientists, analysts, and developers because it is a practical Python data analysis and manipulation instrument.

### C. Confusion Matrix

A confusion matrix table is used to assess how well a machine learning algorithm works. [3] A square matrix compares a model's actual and predicted values for a data set.

The most common divisions are true positives, false positives, true negatives, and false negatives quadrants of the matrix. False positives are instances where the model predicted a positive outcome, but the actual result was negative. True positives are instances where the model accurately predicted a positive outcome. [4] False negatives are instances where the model predicted a negative outcome, but the actual result was positive. True negatives are instances where the model accurately predicted a negative outcome.

Accuracy, precision, recall, and F1 score are just a few performance measures that can be calculated using the confusion matrix. For instance, precision is computed by dividing true positives by the sum of true positives and false positives, while accuracy is determined by dividing the sum of true positives and true negatives by the total number of cases.

A confusion matrix is a helpful tool for assessing how well a machine-learning model performs because it offers a clear and succinct summary of the model's advantages and disadvantages. We can learn more about how the model is doing and spot areas for development by examining the various quadrants of the matrix.

### D. Scikit Learn

With various supervised and unsupervised learning methods, and tools for data preprocessing, model selection, and evaluation, Scikit-learn is a well-known open-source machine-learning library for Python. [5] Based on NumPy, SciPy, and Matplotlib, Scikit-learn is intended to be user-friendly and effective for big data processing.

A range of Classification, regression, clustering, and dimensionality reduction algorithms is available in the library, along with feature extraction and selection tools.

Additionally, cross-validation, hyperparameter tuning, and model analysis techniques are offered by Scikit-learn. The library has thorough documentation and provides a variety of examples and tutorials to assist users in understanding machine learning.

Applications in natural language processing, image identification, and predictive analytics benefit the most from Scikit-learn. The library can manage numerical and categorical data and covers various input data types, including sparse matrices.

Linear regression, logistic regression, k-nearest neighbors, decision trees, random forests, support vector machines, and neural networks are some of the most well-liked methods in sci-kit-learn. Scikit-learn also offers ensemble learning tools, which combine various models to boost prediction efficacy.

### E. Decision Tree

The decision tree, a non-parametric supervised learning method, can solve classification and regression problems. It is arranged in a hierarchical structure with internal roots, branches, internal nodes, and child nodes. [6] In data mining and machine learning, the decision tree is a particular type of predicting model. Its layout is similar to a tree, with each node denoting a judgment call or test on a specific element of the incoming data and each branch denoting one or more potential values. The tree's leaves represent the ultimate result or forecast. Both categorization and regression issues can be solved using decision trees. The tree is constructed in a classification problem to forecast the class label for a given occurrence. The tree is constructed to forecast a serial number in a regression issue. [7] The Decision Node and Leaf Node are the two parts of a decision tree. While Leaf nodes are the results of choices and do not have any additional branches, Decision nodes are used to make decisions and have numerous branches. Because it begins at the root node and adds additional branches to mimic a tree, it is known as a decision tree. Selecting the best algorithm for the available data and the issue at hand is essential to consider when developing a machine learning model. The two reasons for employing the decision tree are as follows: Decision trees are frequently designed to imitate the elements that decision-makers consider, which makes them simple to comprehend. The decision tree's reasoning is easy to comprehend because it uses a tree-like structure.

### F. K-Nearest Neighbor

The assisted learning method K-Nearest Neighbor is additionally utilized. [8] The K-NN algorithm puts the new case in the group that most closely resembles the categories for the existing cases, indicating a relationship between the new case and the existing cases. The K-NN algorithm saves every accessible information and categorizes new material based on similarities. This indicates that the K-NN algorithm can accurately and speedily categorize new data. The K-NN algorithm gives the new data point in the classification assignment the name of the prevalent class among its K nearest neighbors. K-NN calculates the average or median of the values of the K closest neighbors to predict the value of the new data point in the regression

assignment. <sup>[9]</sup> K is a crucial parameter in the K-NN algorithm because it determines the precise Classification or forecast. Underfitting may result from a high value of K, whereas overfitting may result from a low value of K. The K-NN algorithm has several benefits, such as managing non-linear decision boundaries, simplicity, and interpretability. However, meticulous preprocessing and data normalization are necessary to guarantee accurate results, which can be computationally expensive for big datasets. The k-NN classification technique produces a class membership. The majority of its neighbors determines the Classification of a particular object, and it is assigned to the category in which its k closest neighbors are most likely to agree. (k is a large positive number). If k = 1, the item is added to the class of its nearest neighbor.

*G. Logistic Regression*

Logistic regression is a machine learning technique that fits into the supervised learning group. A predetermined collection of independent variables is used to predict the categorical dependent variable. Forecasting the outcome for a binary dependent variable uses logistic regression. Therefore, the result must be a number that is discrete or categorical. Instead of the precise values between 0 and 1, it provides the probability values that fall between those numbers. Either True or False, 0 or 1, or Yes or No, are possible outcomes. Logistic regression is widely used in many sectors, such as healthcare, finance, and marketing, to forecast the likelihood of an occurrence based on a set of predictor factors. <sup>[10]</sup> For instance, it can predict whether customers will buy a product based on their demographic information and prior transactions. The simplicity, interpretability, and ability to manage non-linear decision boundaries are just a few benefits of logistic regression. It functions well with small and medium-sized files as well. The predictor variables and the response variable may have missed complex relationships and may not perform well when the predictor variables are highly correlated.

*H. Random Forest*

The guided learning approach includes the machine learning program, Random Forest. It can be applied to ML issues that combine categorization and regression. <sup>[11]</sup> Its foundation is the idea of ensemble learning, a method for merging various classifiers to handle challenging issues and enhance model performance. As its name implies, Random Forest is a classifier that increases the forecast accuracy of the dataset by averaging several decision trees applied to different parts of the provided dataset. Instead of depending solely on one decision tree, the random forest uses predictions from each decision tree and determines the result based on the votes of the majority of predictions. <sup>[12]</sup> The woodland's larger tree population prevents better accuracy and overfitting. Random forests have several benefits compared to individual decision trees, such as managing high-dimensional data, noisy data, and absent values. It also improves the model's extension skills and reduces the likelihood of overfitting. Random forest is used in many fields, including those where exact forecasts are crucial, like finance, healthcare, and marketing. For instance, predicting a patient's likelihood of contracting a particular illness is possible using their medical history and

demographic information. One of the disadvantages of random forest is its computational complexity, which can be challenging for large databases. Comparing the predictions from different decision trees to those from random forest models can be difficult.

**IV. RESULTS & DISCUSSIONS**

The data set was created for us in this study, and we manually added the records and features. Since our study is about food recommender systems, we have decided to divide and create a dataset and add various cuisines and courses for all kinds of people to make, i.e., from easy to complex, from regional to international and from minimal time to lots of preparation time. This data will then be classified into testing and training models and the best accuracy will be shown. The program will classify the data into 20% testing and 80% training data.

The average of five readings from each model is used to calculate the readings in Table 1. The accuracy, precision, and recall F1-Scores for Decision Tree are 62.64%, 71.16%, 73.12%, and 75.1%, respectively. Accuracy, Precision, Recall, and F1-Score for K-Nearest Neighbor are 56.92%, 65.14%, 73.12%, and 71.6%, respectively. Accuracy, Precision, and Recall, F1-Scores for Logistic Regression, are 64.06%, 65.70%, 88.40%, and 77.7%, respectively. The accuracy, precision, and recall F1-Scores for Random Forest are 66.12%, 72.14%, 79.38%, and 76.4%, respectively.

Table 1: Parameters of different Machine Learning Models

S.NO	Model Name	Accuracy	Precision	Recall	F1-Score
1.	Decision Tree	62.64	71.16	73.12	75.1
2.	K-NN	56.92	65.14	73.12	71.6
3.	Logistic Regression	64.06	65.70	88.40	77.7
4.	Random Forest	66.12	72.14	79.38	76.4

*A. Pie Chart Representation*

Figure 2 demonstrates that 65.4% of the food is nutritious and 34.6% is unhealthy. By considering the calorie count feature, this result is reached. A portion of food falls into the unhealthy group if it has more than 500 calories and the healthy category if it has fewer than 500 calories.

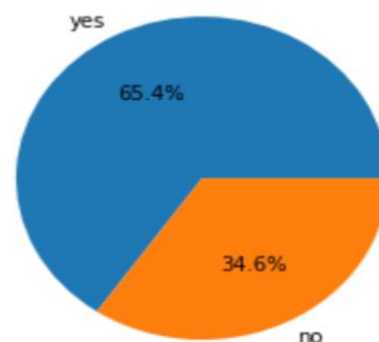


Fig. 2: Pie Chart

## V. CONCLUSION

Our age-based food suggestion system is helpful for people who wish to make wise and healthy food decisions. The system may offer tailored advice from a pre-existing database of foods by inputting the user's age.

The system employs algorithms to analyze the user's age and nutritional requirements to provide suitable food recommendations. The system uses the database, which may contain nutritional data for various items, to recommend meals appropriate for the user's age and dietary needs.

Overall, the meal suggestion system is a valuable tool that promotes healthy eating practices and offers users individualized nutritional guidance. Ongoing research and development could significantly enhance the accuracy and efficiency of the system's dietary recommendations.

## REFERENCES

- [1.] Mishra, R., Mulik, D., Narharshettiwar, A., & Patil, D. FOOD RECIPE RECOMMENDATION ANDROID APPLICATION.
- [2.] Jitendra, M., & Radhika, Y. (2021). An Automated Music Recommendation System Based on Listener Preferences. In *Recent Trends in Intensive Computing* (pp. 80-87). IOS Press.
- [3.] Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In *data democracy* (pp. 83-106). Academic Press.
- [4.] Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, 6(10), 74-78.
- [5.] Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2), 130.
- [6.] B. J. Jaidhan, B. D. Madhuri, and K. Pushpa, "(2019.)Application of Big Data Analytics and Pattern Recognition Aggregated With Random Forest for Detecting Fraudulent Credit Card Transactions (CCFD-BPRRF)," vol. 7, no. 6, p. 6,
- [7.] Wang, L. (2019, December). Research and implementation of machine learning classifier based on KNN. In *IOP Conference Series: Materials Science and Engineering* (Vol. 677, No. 5, p. 052038). IOP publishing.
- [8.] Shanmuk Srinivas Amiripalli, VeeramalluBobba, 'Research On Network Design and Analysis of TGO Topology', *International Journal of Networking and Virtual Organizations*. Vol. 19, No. 1, pp 72-86, 2018 DOI: 10.1504/IJNVO.2018.10015031
- [9.] Shanmuk Srinivas Amiripalli, VeeramalluBobba(2019), Impact of trimet graph optimization topology on scalable networks, *Journal of Intelligent & Fuzzy Systems* 36 (3), 2431-2442. DOI: 10.3233/JIFS-169954.
- [10.] Amiripalli, S.S., Bobba, V. A fibonacci based TGO methodology for survivability in Zigbee topologies (2020) *International Journal of Scientific and Technology Research*, 9 (2), pp. 878-881.
- [11.] Ahmed, J., & Ahmed, M. (2023). Classification, detection and sentiment analysis using machine learning over next generation communication platforms. *Microprocessors and Microsystems*, 98, 104795.
- [12.] Amiripalli, S. S., & Bobba, V. (2020). An Optimal Graph based ZigBee Mesh for Smart Homes. *Journal of Scientific & Industrial Research* Vol. 79, April 2020, pp. 318-322
- [13.] Renjith, S., Sreekumar, A., & Jathavedan, M. (2020). An extensive study on the evolution of context-aware personalized travel recommender systems. *Information Processing & Management*, 57(1), 102078.
- [14.] Shanmuk Srinivas Amiripalli, Mukkamala S N V Jitendra, Surendra Talari, Sannith Akkireddi, D Sateesh Kumar, (2020) "Design and Implement an Artificial Intelligence based Zombie's Application using Unity3D," vol. 83, no. 6, p.16541-16547.
- [15.] Maalouf, M. (2011). Logistic regression in data analysis: an overview. *International Journal of Data Analysis Techniques and Strategies*, 3(3), 281-299.