In the age when diet defines health, knowing about the nutrition of food ingredients becomes very essential. In this work, a mechanism using machine learning is presented for classifying food ingredients in two categories, and the healthiness nature depends on composition and flavor types. Classification methods were trained using an Indian food dataset with ingredient percentage and categorical flavor data. The features studied are average ingredient content and main taste types in different correlations with health indicators.

To check the predictability of performance, six machine-learning algorithms, namely Decision Tree, Random Forest, SVM, Logistic Regression, KNN, and XGBoost, were implemented and compared. Among all these methods, XGBoost had consistently proved its superiority with higher accuracy and a well-balanced performance in all evaluation metrics. Since it was able to take care of complex patterns and intricate interactions between the compositions of ingredients, it becomes particularly well suited for this classification problem. The results give a way to demonstrate that XGBoost can be used ideally as the model for nutrition-driven decisions and recommendations to public health.

The process starts with collecting raw data and applying data cleaning to remove noise and irrelevant data. This is further followed by feature encoding and handling imbalanced data. An important step in preprocessing involves calculating a domain specific feature “unhealthy\_ratio” which signifies the risk posed by each ingredient. Post train – test split we applied six machine learning models. Evaluation metrics like accuracy, precision, recall, f1 – score and confusion matrix are used to assess the model’s performance. The best parameters are retained to improve generalization of the model. Through all these steps the main goal of our work evaluating and classifying food ingredients as healthy or not for human being is achieved.

The performance of various machine learning models such as DT, KNN, SVM, Logistic Regression, RF, and XGB was evaluated on the food ingredient classification task. The findings show that XGB with an accuracy of 94% outperformed others in terms of generalization ability and predictive power. Its ensemble-based gradient boosting approach helped it perform better than other models in this binary classification task.

In this study, we first attempt to separate food ingredients based on their nutritional characteristics into healthy and unhealthy. The classification algorithms with which the performance was tested are RF, XGB, KNN, DT, SVM, and LR, based on the measures of classification efficiency; accuracy, precision, recall, and f1-score. In all the experiments carried out, the ensemble method, XGB, has proven to be the most powerful, showing its robustness and competence in handling this classification problem.

Traditionally, machine learning models have shown good predictive power, especially for binary classification. However, these have room for potential further development. Adding deep learning models might enhance generalization and accuracy even more when we consider bigger and more complex datasets. To understand deeper patterns hidden in the data, future practitioners can explore networks and more sophisticated architectures, such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs).