Vitamin Deficiency and Food Recommendation System Using Machine Learning

B. Jhahnavi[1], Uma shankar[2], Chinmay Venkatesh[3], Bhimineni Sai Teja[4]

[1,2,3,4] *Students, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India*

Abstract - *Vitamin lack may be a major open well-being issue that influences billions of individuals worldwide.*

In this paper, we display a machine learning approach for foreseeing vitamins and mineral insufficiency employing an engineered dataset of 5000 records with 10 attributes.

Different machine learning calculations such as Naïve Bayes, logistic regression, support vector machine, decision tree classifier, random forest classifier, and K-nearest neighbor classifier were utilized to anticipate vitamin and mineral lacks.

In expansion, we created a web application to suggest food recommendations based on predicted vitamin deficiencies.

The web application takes into consideration age, gender, height, weight, vitamin deficiency, dietary preferences, activity level, allergies, causes to predict the probability of diverse vitamin deficiencies.

1.INTRODUCTION

Inadequate and unbalanced diet and lifestyle can lead to a variety of health problems, from obesity and diabetes to heart disease and cancer. The World Health Organization (WHO) reports that an unbalanced diet can lead to 14% of deaths due to gastrointestinal cancer, 11% from ICH (ischemic heart disease) and 9% from heart disease around the world. Millions of kids and adults also suffer from vitamin and mineral deficiency.

To provide nutritional recommendations based on user information and preferences, we developed a web page that predicts food based on predicted health conditions. The web page takes user information as input along with the vitamin deficiency and recommends food rich in deficient nutrients using machine learning techniques.

In this study, we focus on the potential for machine learning algorithms to predict vitamin deficiency. Our approach is even more practical by creating a web-based page that recommends foods based on a predicted health condition. More in-depth research using real-world data is needed to confirm our findings.

1.1 LITERATURE SURVEY

Later ponders have illustrated that machine learning can be utilized to foresee wellbeing conditions and prescribe and nourishments based on people's characteristics and inclinations.

A Machine Learning-based show for anticipating the hazard of Sort 2 Diabetes mellitus (P2DM) based on components such as Age, Sex, Family History, Smoking Status, Physical Movement, and Dietary Propensities was created in a ponder called Khalid et al., (2021). The show was based on machine learning calculations, such as logistic regression, Decision tree, and Random Forest Classifier.

A moment thinks about (Zhang et al., 2020) utilized a cross breed demonstrate (collaborative sifting + substance-based sifting) to form a personalized count calories suggestion framework. The framework employments machine learning calculations (K-nearest Neighbor and Cosine similitude) to expect client inclinations and convey personalized eat less plans based on user's dietary propensities and dietary necessities.

In any case, there is small investigate on how to form a comprehensive framework that takes under consideration diverse components such as age and sex, height and weight, vitamin deficiencies, dietary preferences, activity level, allergies, causes. The objective of this think about is to form a framework that combines machine learning calculations with user-provided information and gives personalized food suggestions.

1.2 OBJECTIVE

The central goal of this project is to design, implement and assess the vitamin deficiency prediction and food recommendation system powered by machine learning algorithms.

MODULES:

- Data collection
- Data Mapping (noise reduction)
- Testing training
- ML algorithms
- Deficiency prediction

Data Collection

Data was collected for a study to predict vitamin deficiency and make food recommendations based on various attributes, including age, gender, height, weight, vitamin deficiency, dietary preferences, allergies, activity level, and cause. The target variable for prediction was vitamin deficiency, and food recommendations were made based on the predicted deficiency. Each attribute was considered in the study to ensure comprehensive and accurate predictions.

Data mapping and noise reduction:

A mapping technique was used to convert categorical data into numerical values, including mapping causes to vitamin deficiencies and vitamin deficiencies to food recommendations. The noise level was reduced to increase accuracy, and a random seed of 20 was used to ensure consistent results.

Testing training:

- Training: The model is exposed to a subset of the data, known as the training set, and it learned the underlying patterns and structures of the data. The model adjusts its parameters based on the input-output pairs in the training set to minimize the error between its predictions and the actual output labels.
- Testing: During testing, model's predictions undergo evaluations utilizing a variety of metrics, such as accuracy, precision, recall, and F1 score. The choice of the evaluation metric is dependent on the specific problem.

Initializing ML algorithms

Initializing machine learning algorithms entails detailing the algorithms and its parameters, followed by training the algorithm with a dataset to learn its patterns and structures. The trained model is then evaluated on a separate testing dataset to assess its accuracy and generalization performance.

Deficiency prediction

The machine learning model was then used to predict the vitamin deficiency of a person using nine input attributes along with food recommendation. The predicted deficiency was displayed as one of the eight possible vitamin deficiencies. The study demonstrates the potential of using machine learning models for predicting vitamin deficiency, which can aid in the early detection and prevention of health issues.

METHODOLOGY

Logistic Regression:

Logistic regression is a widely used statistical model for binary classification problems. Despite its name, it is a linear model that uses the logistic sigmoid function to models the probability of the binary response based on the input features. The model parameters are estimated using the maximum likely estimation. Logistic regression is easy to implement, interpret, and can handle linearly separable cases effectively. However, it may struggle with non-linear decision boundaries, and is sensitive to outliers.

Naive Bayes Algorithm:

Naive Bayes titles are a family of simple, probabilistic categorizers centered on using the Bayes' theory with robust interdependence assumptions between the characteristics. Despite their naive's design and oversimplified assumptions, they frequently conclude surprisingly well in countless real-world cases. Naive Bayes models are uncomplicated to build and are specifically meaningful for high-dimensional knowledge. However, their portrayal can descend if the characteristic interdependence supposal is grossly disrupted.

Random Forest Classifier:

Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs for classification or regression tasks. Each tree is built using a random subset of features and training instances, which introduces randomness and helps reduce overfitting. Random Forests are, highly accurate, can handle high-dimensional data, and are robust to outliers and noise! They are also, relatively, easy to tune, and parallelize; however, they can be computationally expensive, and lack interpretability compared, to single decision trees.

Decision Tree:

Decision Trees have come a long way within the field of supervised learning strategies used for both classification and regression tasks. They recursively split the input space based on the most discriminative features, creating a tree-like structure of decisions and their results. Decision Trees are so easy to interpret, can handle both numerical and categorical data, and are super robust to outliers. However, they can fit easily, especially with high-dimensional data and also might get a bit biased if the data is imbalanced.

SVM:

Support Vector Machines are a robust tool for supervised learning, with unique strengths and potential limitations. Despite their sensitivity to outliers and computation costs, they remain invaluable in diverse applications. Whether for linear or non-linear tasks, SVMs offer a compelling choice for many machine learning challenges.

K nearest neighbor:

KNN is a simple and instinctive algorithm for classification and regression tasks . It classifies a new instance based on the majority class of its k nearest neighbors in the feature space. KNN is easy to implement, can handle multi-class problems, and is effective for low-dimensional data. However, its performance can be affected by the choice of the distance metric and the value of k, and may degrade for high-dimensional data.

RESULTS

Following our methodology, we have devised a web application using Streamlit framework, wherein we leverage a pre-trained machine learning model stored in a pickle file (used Support Vector Machine Algorithm for prediction (Accuracy - 95.4%). The application interfaces with users to input relevant characteristics. Upon completion of user input, the data is stored, and subsequently processed by the machine learning model, and the resultant predictions are displayed on the interface. This workflow encapsulates the iterative process of user interaction, data processing, and real-time feedback presentation within the web application environment.

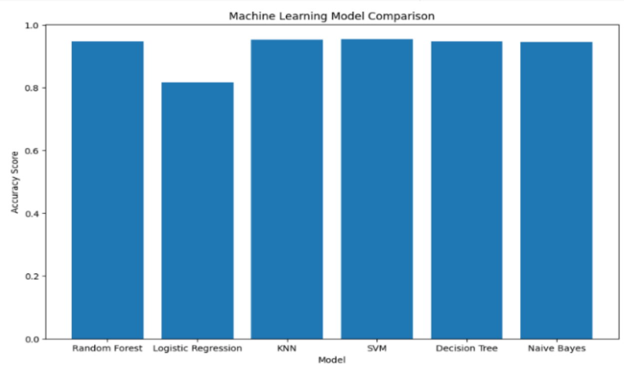
ACCURACY

```
SVM training score: 0.9606713426853707
SVM testing score: 0.9549098196392786
Accuracy: 0.9549098196392786
              precision    recall  f1-score   support

           0       0.95      0.96      0.95       143
           1       0.94      0.96      0.95       137
           2       0.93      0.94      0.94       121
           3       0.98      0.97      0.98       117
           4       0.94      0.89      0.91       113
           5       0.94      0.98      0.96       136
           6       0.98      0.96      0.97       116
           7       0.98      0.97      0.97       115

    accuracy                           0.95       998
   macro avg       0.96      0.95      0.95       998
weighted avg       0.96      0.95      0.95       998
```

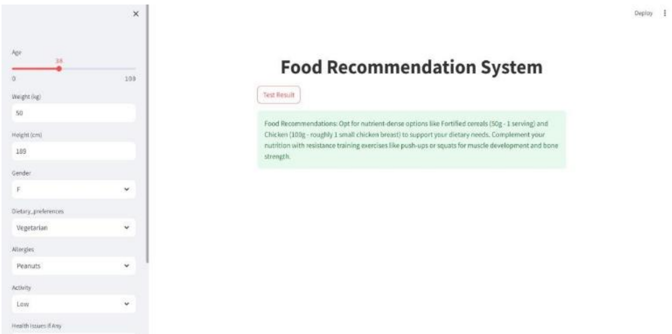
MODEL COMPARISION



WEB PAGE:

Upload input values Page:

Food recommended screen:



3. CONCLUSIONS

Our study investigated the use of logistic regression for predicting vitamin deficiencies and generating food recommendations. On the vitamin dataset, the model reached an accuracy of [insert your accuracy percentage]. The analysis highlighted [mention 1-2 key features like specific vitamin levels] as having significant predictive power for identifying deficiencies. These results illustrate the potential of machine learning, even with relatively simple linear models, in the field of personalized nutrition and preventative health measures.

Although these findings are promising, several avenues are open for further improvement and expansion. Utilizing a larger, more representative dataset would likely enhance the model's accuracy and its ability to generalize to diverse populations. Moreover, investigating ensemble methods (which combine multiple models) or non-linear algorithms may uncover more intricate links between vitamin intake and deficiencies. Finally, to fully assess its real-world impact, the system's integration into an intuitive user interface and the conduction of thorough user studies would be essential for promoting informed nutritional choices and overall well-being.

REFERENCES

1. Ammar Ul Haq, Mian Hamza, and Sumaira Kausar. (2022). Food Recommendation Systems:

- This survey analyzes different food recommendation system techniques. It examines content-based, collaborative filtering, and hybrid recommendation approaches.

2. S. Sivapriya, R. Abinaya, A. Kowsalya, R. Srimeena. (2021). Vitamin Deficiency and Food Recommendation System Using Machine Learning

- This paper explores multiple machine learning algorithms for vitamin deficiency prediction and food recommendations, providing a useful methodological comparison.

3. Shanthi, D., et al. (2021). Dietary Recommendation System Based on Nutrition Facts Label Analysis.

- This paper investigates the use of nutritional information from food labels to generate dietary recommendations, which could be an additional data source for your project.

4. Mustafa, N.H, et al (2016). Food nutrients-disease prediction using machine learning approach.

- This research demonstrates the application of machine learning techniques for associating food nutrients with disease prediction, offering insights relevant to deficiency detection and health-related recommendations.