Analysis of Indian Food Based on Machine learning Classification Models

Article *in* Journal of Scientific Research and Reports · July 2021

DOI: 10.9734/JSRR/2021/v27i730407

4 authors:

Sasmita Nayak
Centurion University of Technology and Management
31 PUBLICATIONS **168** CITATIONS
SEE PROFILE

Mamata Beura
KIIT University
3 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Mohammed Siddique
Centurion University of Technology and Management
37 PUBLICATIONS **167** CITATIONS
SEE PROFILE

Siba Prasad Mishra
Centurion University of Technology and Management
215 PUBLICATIONS **888** CITATIONS
SEE PROFILE

Journal of Scientific Research & Reports

27(7): 1-7, 2021; Article no.JSRR.71268
ISSN: 2320-0227

Analysis of Indian Food Based on Machine learning Classification Models

Sasmita Kumari Nayak[1], Mamata Beura[2], Mohammed Siddique[3] and Siba Prasad Mishra[4*]

[1]Department of CSE, Centurion University of Technology and Management, Odisha, India.
[2]Department of CSE, College of Engineering, Bhubaneswar, Odisha, India.
[3]Department of Mathematics, Centurion University of Technology and Management, Odisha, India.
[4]Department of Civil Engg., Centurion University of Technology and Management, Odisha, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JSRR/2021/v27i730407
Editor(s):
(1) Shunpu Zhang, University of Nebraska- Lincoln, USA.
Reviewers:
(1) Eliane Rodrigues, Brazil.
(2) Michela Sperti, Italy.
Complete Peer review History: https://www.sdiarticle4.com/review-history/71268

Original Research Article

Received 20 May 2021
Accepted 23 July 2021
Published 30 July 2021

ABSTRACT

For human life, Food is highly necessary and essential for human to live the life. The objective of the current study is to characterise, classify and compare the food consumption patterns of many Indian food diets such as non-vegetarian and vegetarian. Given data about different Indian dishes, we try to predict here the dish is vegetarian or not. To get the best predictive model, this study is conducted with the comparison of Decision Tree, K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Random Forest algorithms. In this study, the concept and implementation of all these four models be made for prediction of Indian food. For training and testing the models, Indian food dataset is used that contains, in total 255 records to fit with all these four models. In short, the classification and prediction of Decision tree and KNN model provides less performance than the other models used here. However, the Random Forest model was generally more accurate than SVM, KNN and Decision Tree model, which have got from the simulation.

Keywords: Indian food; k-nearest neighbor; support vector machine; decision tree; random fores.

**Corresponding author: E-mail: 2sibamishra@gmail.com;*

1. INTRODUCTION

A Healthy diet is very essential to human health [1]. Natural products are commonly utilised as food. Even, itmightprocess to fulfilthe need of people. The attributes of Food (processed as well asnatural food) likenutrients, compositions, type, and process styles are related to issues of healthy food. There is a kind to say: “food is the god of people” [2]. People from various locales have distinctive dietary patterns. In all over the world, safety and the quality of food can be examined by knowing the attributes of food, which is very essential for consumers. Hence, food classification could be anexampleof healthy life style [3,4,5].

Health is the main driver forveggie lover food utilization among non-vegans.Mainly, Food computing employs the strategies from technology for food-related study.This includes the acquisition as well as analysis of Indian food dataset containing various modalities (e.g., smell, taste, recipe, food logs and food images) [6].This analysis employs to data mining, machine learning, computer vision and other modern technologies to connect between the human and food for supporting human-centric services. The human-centric services likeunderstanding the human culture, guiding the human behaviour, improving the human health etc. [6]. In this part, we tend to introduce commonly used data sources for classification and prediction of Indian food diet. Furthermore, it provides the overview and comparison on datasets of existing food with various kinds.

The position of India is first among countries with the biggest populace of veggie lovers, and 40% of Asian Indians are vegetarian.It feels like a "nutrition transition" among vegan in India with a lower consumption of entire plant food substance and replace with refined carbohydrates, fried foods, and processed foods. This explores the relationship between the consumption of a veggie lover diet.

The remainder of this article is as per the following: the proposed methodology is introduced in section III. The outcomes of this study have been discussed in section IV and finally, the conclusion and future work has been projected in section V.

2. LITERATURE REVIEW

Data mining covers research for feature selection, feature extraction and classification of problems of agriculture, Healthcare, financial data analysis, retail industry. In machine learning, supervised learning is a method to deal with all fields issues.For classification, there are numerousmodels used like KNN [7], [8], [9], SVM [8], [10], [11], [12], [9],Decision tree [8], [10], [11], [9], Random forest [7], [10], [11], [9], ANN [13], CNN [3], [5] and so on.

The popularity in classification of Indian Food is gaining slowly due to the awareness of food and health among people.As indicated by the World Health Organization (WHO) [5,14], more than 1.9 billion adults (18 years above) were overweight.It is terribly stunning to understand that 13% of the total populace includes both women and men (15% women and 11% men) are overweight.In reality, some of individualsover the globe are suffered from overweight, which has doubled since 1980. As a result, it shows that foodhas played animportant role in fitness of an individual.

According to the Statistics, 95% of the individual disobeysthedietary plan as they are very strict and restrict individualtoconsume their regular food.Some of are, youth are tracking the intake of nutrition and calories to maintain fitness, patients want to control their health through food because of various nutritional constraints, old aged individuals monitoring their intake of food etc.. Extraction of calories and image based dietary has been a difficult job. Over the recent years, many research works is going on for this issue [5].

In day-to-day life, automatic, accurate, and rapid decision of attributesfood is auseful demand [3]. For modelling, severaltechniques of data analysis areformulated to manage a large amount of data, viz. Artificial Neural Network (ANN) [10], Partial Least Squares (PLS) [15], Random Forest [7], [10], [16], [11], [9], Support Vector Machine (SVM) [10], [17],[11], [12],[9], K-Nearest Neighbor (KNN) [7], [18], [9], and so on.

Most of the research work has made on image classification for Indian food by using the deep learning models, which can be utilised for classification of image like CNN [3], [5]. For feature extraction, viz. Principal Component Analysis (PCA) [19], Wavelet Transform (WT) [20], Independent Component Correlation Algorithm (ICA) [21], Scale-Invariant Feature Transform [22], Speedup Robust Features [23], Histogram of Oriented Gradient [24]. These

techniqueshave demonstrated their great value to deal with these data.

Through this analysis, an exertion has been put to categorisethe Indian food dataset into their individualgroups with the help of classification algorithms. Regarding accuracy, a comparison has been doneamong the models.

3. PROPOSED METHODOLOGY

This type of research had been completedwith the help ofnumerous Machine Learning (ML)models, viz. KNN [7,9], SVM [10,11,12,9], Decision Tree [9,20,24] and Random Forest [7,10,11,9]. ML models have been generally employed to large and high dimensional datasets [10,25]. ML is a subset of Artificial Intelligence (AI), by which a computer algorithm learns from the past experience [9,25]. The steps of machine learning model as shown in Fig.1 [10]. The most essential part of machine learning model is the collection and pre-processing of data. This model has been applied to clean, normalize and pre-process the collected data called as Indian Food raw dataset.

In this study, Indian food dataset is considered as the dataset (available at https://www.kaggle.com/nehaprabhavalkar/indian-food-101) for all proposed machine learning methodologies. This dataset contains 255 numbers of records with 9 numbers of attributes or features. Prediction of Indian food can be made using parameters like, name, ingredients, diet (Vegetarian or Non vegetarian), preparation time, cook time, flavour of food (spicy, sweet etc.), course, state, and region.

In this article, we have considered the following machine learning models to predict the Indian food. They are K-Nearest Neighbour (KNN) [7,9], Decision Tree [10,11,9], Random Forest (RF) [7,10,11,9], and Support vector Machine (SVM) [10,11,12,9] models. These four machine learning models are supervised machine learning models. We have implemented the above four models for classification experimentally and compared with each other to find the highest accuracy.

3.3 Decision Trees (DST)

Decision tree (DST) [9], [20], [24] is based on the principle of decision rules. According to this, it represented as a tree structure form, which contains root nodes, branches, internal nodes as well as leaf nodes. Always the target values are represented as leaf nodes. It is also a supervised machine learning algorithm, which learns from the past experiences. Due to this stability as well as reliability, DST algorithm is preferable.

3.4 Random Forest (RF)

Random forest (RF) [7,10,11,9] is an ensemble learning model, which is the combination of any type of model or combine the same model with several times. Due to this RF algorithm is a powerful algorithm. In this study, RF model combines the decision trees several times that is result a forest (set of trees). After comparing all the models, RF model provides a good result to classify the Indian food.

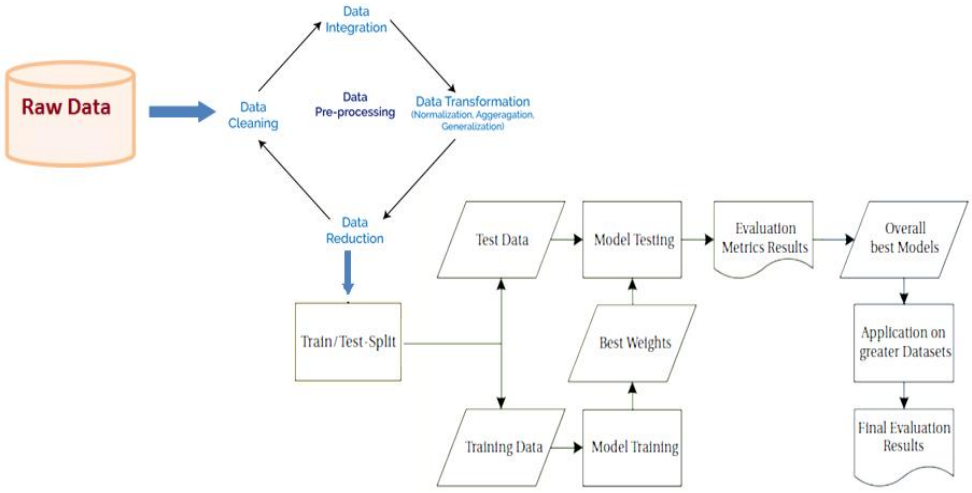

Fig. 1. Steps of Machine Learning Model [10].

4. RESULT AND DISCUSSION

In this study, Indian food dataset is considered. The details of this dataset has discussed in methodology section. The proposed methodologies are evaluated with a dataset obtained from the Kaggle online source.

Before applying the machine learning model, we have analysed the Indian food dataset based on the state, type of food and type of dishes parameters. Before this, we first found the number states for each category of food as shown in Fig. 2. The number of vegetarian and Non-vegetarians states are 226 and 29 respectively.

Fig. 3 and Fig. 4 shows the top 10 state with respect to the type of dishes as well as state wise most liked foods respectively. From the Fig. 3 the top most state is Gujarat to get the various numbers of dishes.

The dataset is split into two sub datasets in the ratio of 80:20, named as training and testing dataset. I.e. 80% of dataset has considered as the training dataset and the rest 20% for test dataset. All machine learning models are trained by training dataset initially. After that test with the new dataset i.e. test dataset.

The outcome of training data is model used for classification to classify the Indian food diet attribute is a Vegetarian or Non Vegetarian. The result of all models with the classification accuracy is shown in Table 1.

To predict and classify the Indian food as Vegetarian or Non Vegetarian, the best model for Random Forest model is no. of estimators=100 with classification accuracy 92.2%, for SVM model at value C=1 with classification accuracy 88.2%, for KNN model is K=3 with 86.3% accuracy and Decision Tree has Max_depth =3 with classification accuracy is 86.3%. The accuracy of KNN and Decision Tree is same, the SVM is little higher than this accuracy and Random Forest model will classify and predict the Indian Food with high accuracy as compared to others. Hence, the Random Forest model is better than the SVM, KNN and Decision Tree model as shown in Fig. 5.

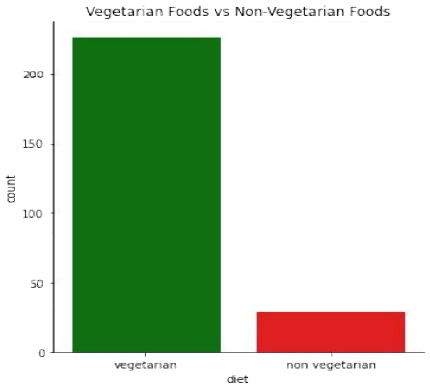

Fig. 2. Number of states with respect to Vegetarians and Non-Vegetarians

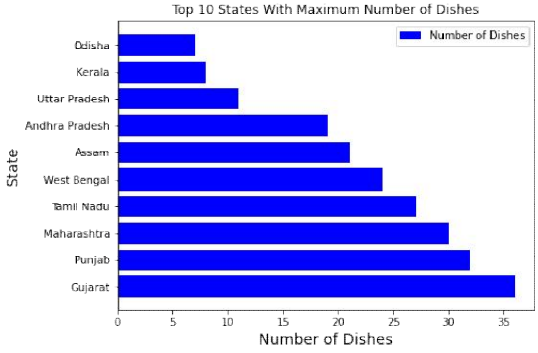

Fig. 3. Top 10 State with Maximum number of Dishes

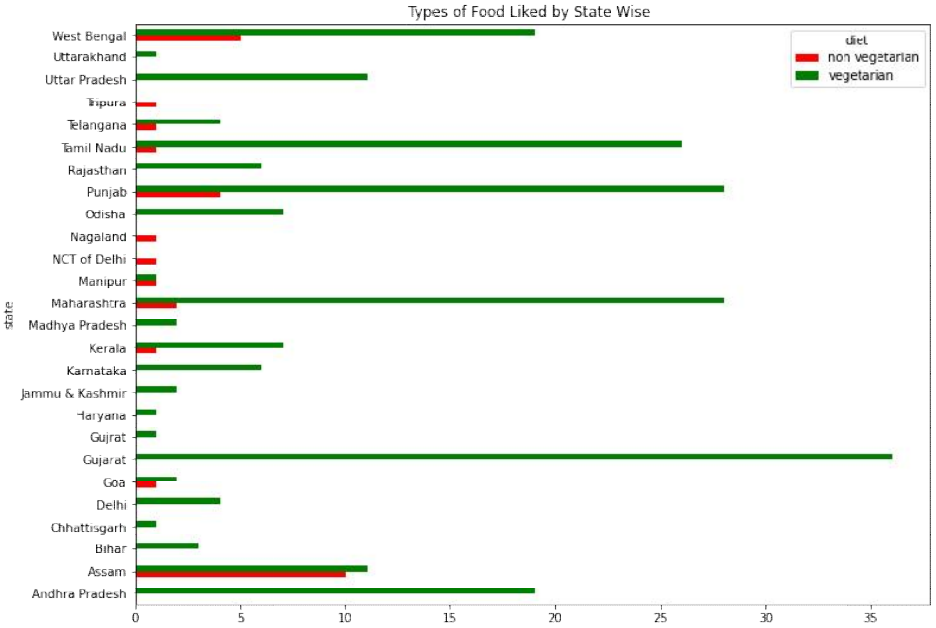

Fig. 4. Most Liked Types of Food by State Wise

Table 1. Result and Accuracy of Machine Learning Model

Model Name	Result	Accuracy Score
Random Forest	No. of estimators=100	92.16%
SVM	C=1	88.24%
KNN	K=3	86.27%
Decision Tree	Max_depth=3	86.27%

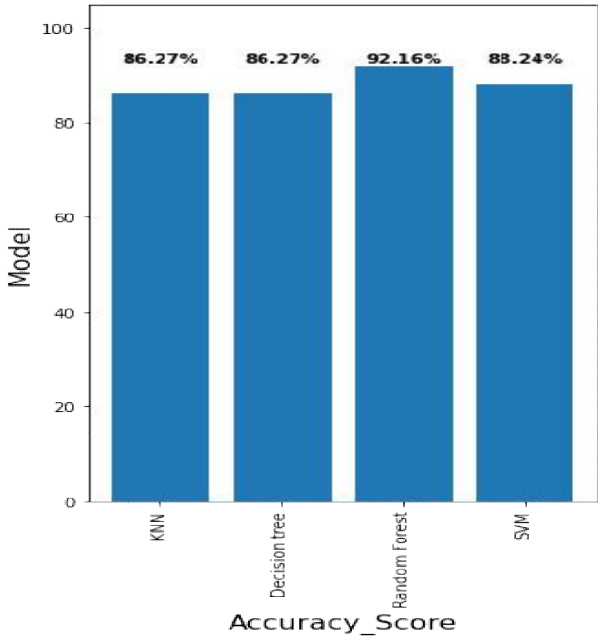

Fig. 5. Comparison of Different Models with Accuracy

5. CONCLUSION

In this article, the experimental results are compared by the most popular machine learning model for classification and prediction of Indian food. All models have their own advantages and limitations, but the toughest work is to choose the best model. From these observations, it concluded that Random Forest works well to predict and classify the Indian food diet with accuracy 92.2%. For this Indian food database may be classified better with more accuracy, if any hybrid model is used. Hence, in future work a hybrid modelwill develop for this database to get the higher and better accuracy.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCE

1. Denise de Ridder, Floor Kroese, Catharine Evers, Marieke Adriaanse & Marleen Gillebaart. Healthy diet: Health impact, prevalence, correlates, and interventions, Psychology & Health, 2017;32(8):907-941. DOI: 10.1080/08870446.2017.1316849
2. Luo C, Wei Q, Zhou L, Zhang J, Sun S. Prediction of vegetable price based on Neural Network and Genetic Algorithm. In International Conference on Computer and Computing Technologies in Agriculture (pp.). Springer, Berlin, Heidelberg. 2010; 672-681.
3. Zhou L, Zhang C, Liu F, Qiu Z, He Y. Application of deep learning in food: A review. Comprehensive Reviews in Food Science and Food Safety, 2019;18(6): 1793-1811.
4. Lule SU, Xia WS. Food phenolics, pros and cons: A review. Food Reviews International, 2005;21(4):367-388. Available:https://doi.org/10.1080/87559120500222862
5. Rajayogi JR, Manjunath G, Shobha G, Indian Food Image Classification with Transfer Learning, 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 2019:1-4. DOI:10.1109/CSITSS47250.2019.9031051
6. Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A Survey on Food Computing. ACM Comput. Surv. 2019;1(1):42. Available:https://doi.org/0000001.0000001
7. Tapas Ranjan Jena, Swati Sucharita Barik, Sasmita Kumarai Nayak, Electricity Consumption & Prediction using Machine Learning Models, Muktshabd. 9(6): 2804-2818.
8. Wiyono, S, Abidin T. Comparative Study of Machine Learning KNN, SVM, and Decision Tree Algorithm to Predict Student's Performance. International Journal of Research-GRANTHAALAYAH. 2019;7(1):190-196.
9. Sripada Swain, Sasmita Kumari Nayak, Swati Sucharita Barik, A Review on Plant Leaf Diseases Detection and Classification Based on Machine Learning Models, Muktshabd 9(6):5195-5205.
10. Sasmita Kumarai Nayak, Swati Sucharita Barik, Mamata Beura, Weather Forecasts Based on Rainfall Prediction Using Machine Learning Methodologies,Adalya Journal. 9(6):72-80.
11. Nayak SK. Analysis and High Accuracy Prediction of Coconut Crop Yield Production Based on Principle Component Analysis with Machine learning Models. International Journal of Modern Agriculture. 2020;9(4):359-369.
12. Sasmita Kumarai Nayak, Swati Sucharita Barik, Mamata Beura, Analysis of Infectious Hepatitis Disease with High Accuracy Using Machine Learning Techniques, TEST Engineering & Management 83. 2020;83:14294-14302.
13. Yiqun H, Kangas LJ, Rasco BA. Applications of artificial neural networks (ANNs) in food science. Critical Reviews in Food Science and Nutrition, 2007;47(2): 113-126. Available:https://doi.org/10.1080/10408390600626453
14. Comparison of Convolutional Neural Network Models for Food Image Classification by Gözde özsert y???T and buse melis özyildirim
15. Cheng JH, Sun DW. Partial least squares regression (PLSR) applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle. Food Engineering Reviews, 2017;9(1): 36-49. Available:https://doi.org/10.1007/s12393-016-9147-1
16. Bossard L, Guillaumin M, Gool LV. Food-101-Mining discriminative components with random forests. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), Computer

Vision - ECCV 2014, Pt Vi. 2014;8694:446-461.
Available:https://doi.org/10.1007/978-3-319-10599-4_29
17. Pouladzadeh P, Villalobos G, Almaghrabi R, Shir mohammadi S. A novel SVM based food recognition method for calorie measurement applications. In 2012 IEEE International Conference on Multi-media and Expo Workshops 2012: 495-498.
Available:https://doi.org/10.1109/ICMEW.2012.92
18. Yordi EG, Koelig R, Mota YC, Matos MJ, Santana L, Uriarte E, Molina E. Application of KNN algorithm in determining the total antioxidant capacity of flavonoid-containing foods. In 19th International Electronic Conference on Synthetic Organic Chemistry; 2015.
Available:https://doi.org/10.3390/ecsoc-19-e002
19. Granato D, Santos JS, Escher GB, Ferreira BL, Maggio RM. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. Trends in Food Science & Technology. 2018;72:83-90.
Available:https://doi.org/10.1016/j.tifs.2017.12.006
20. Ma BQ. Food packaging printing defect detection method based on image wavelet transform. Food Research and Development, 2017;38(5):212-215.
Available:https://doi.org/10.3969/j.issn.1005-6521.2017.05.046
21. Monakhova YB, Tsikin AM, Kuballa T, Lachenmeier DW, Mushtakova SP. Independent component analysis (ICA) algorithms for improved spectral deconvolution of overlapped signals in H-1 NMR analysis: Application to foods and related products. Magnetic Resonance in Chemistry. 2014;52(5):231-240.
Available:https://doi.org/10.1002/mrc.4059
22. Giovany S, Putra A, Hariawan AS, Wulandhari LA. Machine learning and SIFT approach for Indonesian food image recognition. Discovery and Innovation of Computer Science Technology in Artificial Intelligence Era. 2017;116:612-620.
Available:https://doi.org/10.1016/j.procs.2017.10.020
23. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). Computer Vision and Image Understanding. 2008;110(3):346-359.
Available:https://doi.org/10.1016/j.cviu.2007.09.014
24. Ahmed A, Ozeki T. Food image recognition by using Bag-of-SURF features and HOG Features. In Proceedings of the 3rd International Conference on Human-Agent Interaction. 2015;179-180.
Available:https://doi.org/10.1145/2814940.2814968
25. Panesar SS, D'Souza RN, Yeh FC, Fernandez-Miranda JC. Machine learning versus logistic regression methods for 2-year mortality prognostication in a small, heterogeneous glioma database. World neurosurgery: X, 2, 100012; 2019.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/71268