Ingredients Based Food Recommendation System Using Naive Bayes And KNN Algorithm

[1]Mr. Vishal Shinde,[2]Miss. Asmita Bhoir ,[3]Mr.Jaywant Wagh , [4]Miss. Varsha Shinde,
[5]Miss. Vishweta Thakare, [6]Mr. Brijesh Pal

[1]Asst. Professor,[2,3,4,5,6]UG Student,[1,2,3,4,5,6]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

[1]mailme.vishalshinde@gmail.com,[2]asmitabhoir1428@gmail.com, [3] jswagh14@gmail.com, [4]varshshinde1997 @gmail.com, [5]vishwetathakare83830@gmail.com, [6]palbrijesh59@gmail.com

Abstract- Personalized Recipe System is the system that describes/shares the recipes from one user to other on the web in culinary domain. With the help of internet, many users are connected to each other from different places/countries and shares millions of recipes on the internet world wide. In this system, it contains many features/information's related to cuisine/recipe on framework like cooking ingredients, cooking procedures, online video procedures and images, timeline for recipe stories, search bar, users' feedbacks, categories viz. vegetarian or non- vegetarian and so on. The system proposes a hybrid approach of a recipe system that has content information on cuisine with collaborative filtered features. This is a web-based system that is based on Python programming language. From most of the recipe systems, it is more efficient and elegant in performance as well as better in framework. In proposed system, the users can add their own ingredients or procedures for improving the recipe.

Keywords- Python, better framework, hybrid approach, filtered features, recipe system, web-based.

I. INTRODUCTION

Food is one of the most vital elements that human body need. Recommendation system is a type of system that filters the information a present the list of various items. It is a system that learns from the previous preferences of a consumer and prefers them experimental cuisines. The common of recommendation are the ingredients in the recipes which are already liked by the user. As all Indian cuisine are famous for their spice's herbs and its richness of flavors. Numerous cuisines are present in Indian culture but most of the preferred cuisines are the north Indian cuisine and the south Indian cuisine.

India is very admired for its differing multi-food accessible in countless and in resorts, which is comparable with unanimity in assorted variety. The most essential ingredient that is used in the Indian cuisine for nourishment is wheat, rice and Dal. It is the staple food in most of the Indian country. In recent, the Indian sense of taste has experienced a great deal of progress and merging various cuisines with the Indian cuisine is taking the Indian cuisines to the whole new level. Cuisine's in India is famous for its authentic way cooking. Some of them are The Bengali, the Gujarati and the Punjabis, the Bengali cooking is energizing for its distinct aroma due to Panchphoron. It is a colorful combination of five seasonal seeds that is fennel seed, fenugreek seed, cumin seed, nigella seed and mustard seed [1].The North-Indian cuisine has an immense variety of delicious vegetarian as well as non-vegetarian dishes. The system has been provided with a wide-ranging dataset of over 1000 Indian dishes collected from the web, with in-depth data about the ingredient and process of making that recipe. [1].

AIMS AND OBJECTIVE

a) Aim

The aim of the system is to share and recommend the variety of cuisines/recipes across the worldwide with multiple users using internet. The proposed system is a web- based GUI with various feature based on hybrid approach. This system provides a user-interface to the end-users to add and edit their recipes online.[1]

b) Objective

- Helps users find recipes which suit their taste and choice of ingredients.
- Reduces time required for user-item rating.
- Reduce the hassle of looking for recipes.

 DOI : 10.35291/2454-9150.2020.0225

II. LITERATURE SURVEY

Paper 1: Recommendation of Indian Cuisine Recipes based on Ingredients.:

Personalized Recipe System is the system that describes/shares the recipes from one user to other on the web in culinary domain. The system proposes a hybrid approach of a recipe system that has content information on cuisine with collaborative filtered features. This is a web-based system that is based on Python programming language. From most of the recipe systems, it is more efficient and elegant in performance as well as better in framework. [1].

Paper 2: A Stock Secommendation System Using with Distributed Graph Computation and Trust Model-Collaborative Filtering Algorithm:

The system builds a dichotomy model of shareholder-stock relationship based on the distributed graph computing framework Spark Graph X, and using a certain financial theory, transforms the investment behavior of person (shareholders) into the ratings and firm belief of the invested stock. Then the system will calculate the persons similarity graph and the trust graph of shareholders through the parallel graph calculation, and use the improved collaborative filtering algorithm based on trust model to make a recommendation analysis of the stock of A- shares and SME stocks in the merchandise market [5].

Paper 3: Combining user-based and item- based collaborative filtering techniques to improve recommendation diversity: Collaborative filtering (CF) is basically used for recommending those items to a user which other like-minded users preferred in the past. User-based collaborative filtering (UbCF) and item-based collaborative filtering (IbCF) are two types of CF with a similar objective of estimating target user's rating for the target item. It is an approach which is proposed for merging predictions from UbCF and IbCF through numerous linear regression (MLR) and hold up vector regression (SVR) [7].

III. EXISTING SYSTEM

Using a Secommendation systems it's an efficacious way to solve the problem of information overload and help users to discover extremely useful information. This system builds a dichotomy model of shareholder-stock relationship based on the distributed graph computing framework Spark Graph X, and using a certain financial theory, transforms the investment behavior of person (shareholders) into the ratings and firm belief of the invested stock. Then the system will calculate the persons similarity graph and the trust graph of shareholders through the parallel graph calculation, and use the improved collaborative filtering algorithm based on trust model to make a recommendation analysis of the stock of A- shares and SME stocks in the merchandise market. Finally, the comparison of common collaborative filtering algorithm in the experimental environment shows that the system has better algorithm scalability and accuracy. The point of this work was to progress and differentiate recommendation systems which use the item-based collaborative filtering algorithm, based on Hadoop and Spark [5].

IV. COMPARTIVE STUDY

SR NO.	PAPER TITLE	AUTHOR NAME	TECHNOLOGY	ADVANTAGE	DISADVAN TAGE
1.	Recommendation of Indian Cuisine based on Ingredients.	Madhu Kumari	Django Framework, Postgre SQL and Nginx.	No costing issue, a person will not force to grocery shopping certain ingredients because they can search based on what is already available.	Absence of detail, diversty of user preferences.
2.	Secommendation System Using Distributed Graph Computation and Trust Model Collaborative Filtering Algorithm	Wang, H., Sun,Y., Li, X., Xie, Y., & Qi, Y	Distributed Graph Computation, Trust Model- Collaborative Filtering Algorithm	Scalability and accuracy.	Do not allow point to point communication.
3.	Combining user- based and item- based collaborative filtering using machine Learning.	Thakkar, Priyanka / IEEE (2013)	Collaborative Filtering (CF)	Minimizing overall prediction error.	Insufficiency of information, co-occurrence of item selection of different user.
4.	Data engineered content extraction studies for web pages.	Kolla, Bhanu Prakash and Arun Raja Ramanhua/S pringer (2019)	Data mining, A basic pixel-based approach.	It is used to generate content in a short period, so that further developments can be easier.	Limited presentation capacity, such as mobile phones and speech readers.

Table no: 1

V. PROBLEM STATEMENT

System is unsuitable for implementing complex algorithms, do not allow point to point communication. Insufficiency of information, co-occurrence of item selection of different user.

 DOI : 10.35291/2454-9150.2020.0225

The existing techniques fail to abolish the drawbacks such as information sparsity, new

user cold start problem, new item cold start problem, overspecialization, and shilling attacks.

Limited presentation capacity, such as mobile phones and speech readers.

The limitation is that the users must be sound enough to give searching keywords in order to search relevant information.

VI. PROPOSED SYSTEM

The recommendation system crawl various websites that provide Indian cuisines. Web scraping is done. Since the collected dataset is not well formatted hence, the recommendation system applied data pre- processing techniques in the collected dataset. Content based recommendation system recommends based on contents of the matching profile. The collected dataset has a lot of features like ingredient, steps, time to prepare etc. but it need only a few features to recommend similar recipes. The system selects the column based on which it will perform the recommendation and drop the other features.

VII. ALGORITHM

NAIVE BAYES ALGORITHM FUNCTION:

1. **procedure** $R_{ECOMMEND}$ ($\hat{P}_{ICF}$, $\hat{P}_{NB}$, L) SS
2. **if** ($\hat{P}_{ICF}$ == 0) then
   a. $\hat{P}_{Final}$l ← $\hat{P}_{NB}$
   b. **return** $\hat{P}_{Final}$
3. **end if**
4. Sort the list L in ascending order, so that L (1) contains the lowest value and L(S) contains the highest value.
5. **if** (L(S) 6= L (S − 1)) then
   a. **if** d (S, S − 1) > _ then
6. $\hat{P}_{Final}$← $\hat{P}_{NB}$
7. **return** $\hat{P}_{Final}$
   a. **else**
   b. **if** ($|\hat{P}_{NB} - \hat{P}_{ICF}| <$ _) **then**
8. $\hat{P}_{Final}$ ← $\hat{P}_{NB}$
9. **return** $\hat{P}_{Final}$
   a. end if
10. **end if**
11. **else** (i.e. L(S) = L (S − 1))
12. **for** t ← S − 1, 1 do
13. **if** (L(S) == L(t)) then
    a. **if** ($|\hat{P}_{ICF}$ − t| < _) then
14. $\hat{P}_{Final}$ ← t
15. **return** $\hat{P}_{Final}$
    a. **end if**
16. **else**
    a. Break **for**
    b. end if
17. **end for**
18. **end if**
19. $\hat{P}_{Final}$ ← $\hat{P}_{ICF}$
20. **return** $\hat{P}_{Fina}$
21. end procedure

KNN ALGORITHM FUNCTION:

1. Procedure KNN FUNCTION
   a. Read the value of '*k'
   SET '*k*'
   b. Set paths for training and testing data directories
   SET *trainFile*
   SET *testFile*
   c. Create new JOB
   d. SET MAPPER to map class defined
   e. SET REDUCER to reduce class define
   f. Set paths for output directory
   g. SUBMIT JOB
2. End procedure = 0

IX. MATHEMATICAL MODEL

Naive Bayes:

The Naive Bayes classifier is based on the *Bayes theorem* with strong (Naive) independence assumption, and is suitable for the cases having high input dimensions. Using the Bayes theorem, the probability of a document d being in class C_j is calculated as follows:

$$P(C_j|d) = \frac{P(C_j)P(d|C_j)}{P(d)},$$

where P (C_j |d), P(C_j), P (d| C_j), and P(d) are called the *posterior*, *prior*, *likelihood*, and *evidence* respectively. The Naive assumption is that features are conditionality independent, for instance in a document the occurrence of words (features) do not depend upon each other16 [29].

Formally, if a document has a set of features F_1, ・ ・ ・, F_h then express equation 3 as follows:

$$P(CP(C_j|d) = \frac{P(C_j)\prod_{i=1}^{h} P(F_1|C_j)}{P(F_1, \dots, F_h)}$$

An estimate b P(Cj) for P(Cj) can be calculated as:

$$\hat{P}(C_j) = \frac{A_j}{A}$$

where A_j is the total number of training documents that belongs to category C_j and A is the total number of training documents. To classify a new document, the posterior probability of a movie, n_y, is calculated as follows:

$$P(C_j|n_y) = \frac{P(C_j)\prod_{t=1}^{T}\prod_{t=1}^{|d_t|} P(\omega_{tt}K_j, T_t)}{Pn_y}$$

 DOI : 10.35291/2454-9150.2020.0225

KNN:

There are numerous partitions works anyway Euclidean is the most commonly used measure. It is generally used when information is nonstop. Manhattan separate is likewise basic for nonstop factors.

EUCLIDEAN:

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}$$

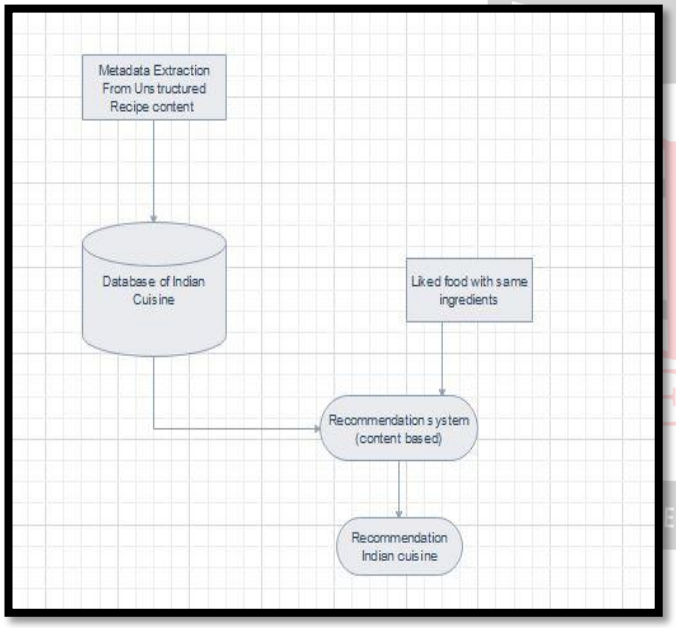
MANHATTAN / CITY-BLOCK:

$$d(x,y) = \sum_{i=1}^{m}|x_i - y_i|$$

K-Nearest Neighbor: The k-Nearest Neighbor algorithm is used to test the degree of law of similarity between documents and k training data.

$$R^x \leq R_{kNN} \leq R^x\left(2 - \frac{MR^x}{M-1}\right)$$

X. SYSTEM ARCHITECTURE

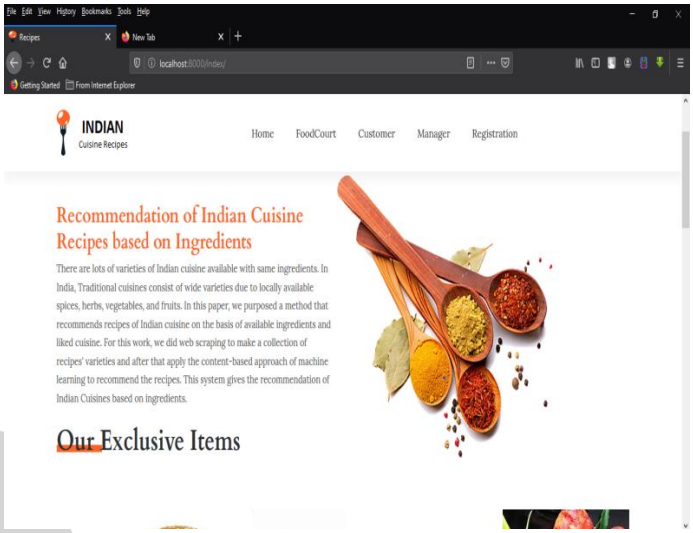


Above figure explains the system architecture that how the flow of data helps the recommender system to carry out the process of suggesting new cuisines to the users. The proposed system has been provided with a large amount of dataset about 10000 Indian dishes which has been scraped from the web, with comprehensive data of the ingredients with steps of each of the recipes. For each end-user, the system stores the dishes which was liked by him/her, and then the system tries to match their ingredients with the ones that are present in its database. After the search ends in the system, it tries to suggests the top results to the end-users.

XI. ADVANTAGES

- It helps user to take correct decisions in their choice of food.
- It helps to increase in sales and helps user get a better experience of choices in cuisines.
- Recommendation engine provides personalization.
- It helps users by exploiting the user preferences and priorities the food items based on the user's past behavior.

XII. DESIGN DETAILS



XIII. CONCLUSION

Thus, we have tried to implement the paper by

"Dr. Madhu Kumari, Vishal Raman, Pritom Hazarika, Nilesh", IEEE (2019), ICDEW, "Recommendation of Indian Cuisine Recipes based on Ingredients". In this paper , method for Indian cuisine recommendation using ingredients matching of cuisine and liked food by users is presented. This personalized recipe system presents cooking ingredients, procedures, cuisines etc. It exploits user preferences and helps them get a better experience in choices of cuisines. Recommendation approach in terms of accuracy, and coverage and is more scalable has been proved.

XIV. REFERENCES

[1] Hazarika, Pritom, Madhu Kumari, and Vishal Raman. "Recommendation of Indian Cuisine Recipes based on Ingredients."(2019)

[2] Pereira, Nymphia, and Satish Kumar L. Varma. "Financial Planning Recommendation System Using Content-Based Collaborative and Demographic Filtering". Smart Innovations in Communication and Computational Sciences. Springer, Singapore, 2019. 141-151.

[3] Kolla, BhanuPrakash, and Arun Raja Raman. "Data Engineered Content Extraction Studies for Indian Web Pages." Computational Intelligence in Data Mining. Springer, Singapore, 2019. 505-512.

[4] Shah, Lipi, Hetal Gaudani, and Prem Balani. "Personalized Recipe Recommendation System using Hybrid Approach"(2016).

[5] Ghazanfar, Mustansar, and Adam Prugel-Bennett. "An improved switching hybrid recommender system using naive bayes classifier and collaborative filtering." (2010).

 DOI : 10.35291/2454-9150.2020.0225