

Recommendation of Indian Cuisine Recipes based on Ingredients

Nilesh

Computer Science and Engineering
National Institute of Technology
Hamirpur, India
bhartinilesh2@gmail.com

Pritom Hazarika

Computer Science and Engineering
National Institute of Technology
Hamirpur, India
hazarikapritom@yahoo.com

Dr. Madhu Kumari

Computer Science and Engineering
National Institute of Technology
Hamirpur, India
madhu.jaglan@gmail.com

Vishal Raman

Computer Science and Engineering
National Institute of Technology
Hamirpur, India
rmnvishal@gmail.com

Abstract— there are lots of varieties of Indian cuisine available with same ingredients. In India, Traditional cuisines consist of wide varieties due to locally available spices, herbs, vegetables, and fruits. In this paper, we purposed a method that recommends recipes of Indian cuisine on the basis of available ingredients and liked cuisine. For this work, we did web scraping to make a collection of recipes' varieties and after that apply the content-based approach of machine learning to recommend the recipes. This system gives the recommendation of Indian Cuisines based on ingredients.

Keywords— Machine Learning, content-based, bag-of-words, NLP, Indian cuisine, web scraping, Recommendation System

I. INTRODUCTION

Recipe Recommendation System for Indian cuisines is a system that learns from the past preferences of a user's preferred dishes to recommend him/her new, untested cuisines. The basis of recommendation are the ingredients in the recipes already liked by the user.

The conventional food of India has been broadly refreshing for its remarkable utilization of herbs and flavors. Indian food is known for its substantial arrangement of dishes. The cooking style shifts from locale to the district and is generally separated into South Indian and North Indian food. India is very acclaimed for its differing multi-food accessible in countless and inn resorts, which is reminiscent of solidarity in assorted variety. The staple nourishment in India incorporates wheat, rice, and heartbeats with chana (Bengal Gram) being the most vital one. In current occasions, the Indian sense of taste has experienced a great deal of progress. Bengali cooking is refreshing for its astounding utilization of panchphoron, a term used to allude to the five basic flavors, to be specific mustard, fenugreek seed, cumin seed, aniseed, and dark cumin seed.

Conventional Gujarati food is essentially a veggie lover and has a high dietary benefit. The commonplace Gujarati thali comprises of shifted sorts of lip-smacking dishes. Gujarati food has such a great amount to offer and each dish has a totally unique cooking style. The cooking of Punjab has a colossal assortment of mouth-watering vegan just as non-veggie lover dishes. The flavor content reaches from negligible to charming to high. Punjabi nourishment is typically savored by individuals all things considered. In

Punjab, home cooking varies from the eatery cooking style. The food of Rajasthan is principally veggie lover and offers an impressive assortment of divine dishes. The flavor content is very high in contrast with other Indian cooking styles, however the sustenance is completely tasty. Rajasthani use ghee for cooking the greater part of the dishes. Rajasthani nourishment is outstanding for its hot curries and delectable desserts.

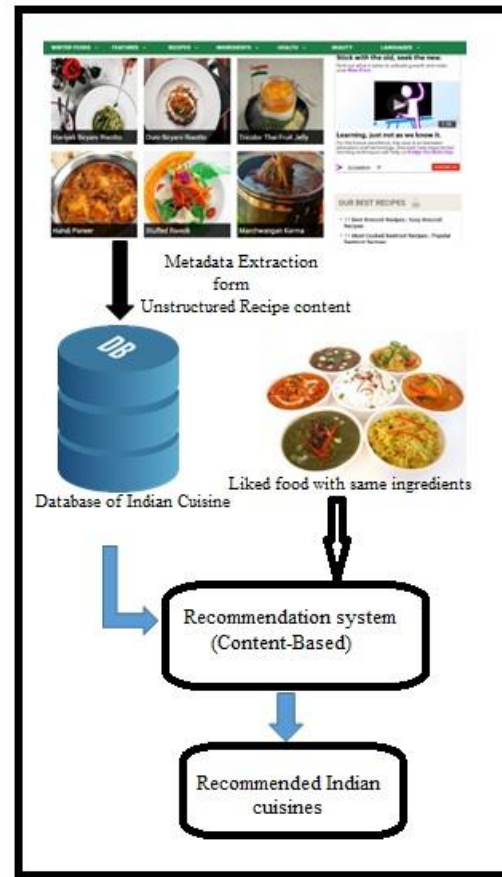


Figure 1: Indian Cuisine Recommender

Figure 1 explains the flow by which the recommender system carries out the process of suggesting new cuisines to the user. The system has been provided with an extensive dataset of over 1000 Indian dishes collected from the web, with exhaustive data about the ingredients and steps of each of the recipes. For each user, the system stores the dishes liked by him/her, and then tries to match their ingredients with the ones in its database. After the search ends, the system suggests the top results to the user.

The next sections of this paper are arranged as follows: Part II describes the underlying techniques and concepts of any Recommendation system. Part III discusses related works and my proposed method of cuisine recommendation and Part IV is about the challenges and issues related to recommendation systems. Part V is the conclusion to the paper with future work suggestions.

II. TECHNIQUES OF RECOMMENDATION SYSTEM

Collaborative filtering [1]: collaborative filtering (CF) is making recommendations based on other comparable customers' profile. It is useful into exceptionally assorted arrangements of things. The resemblances between different things in the dataset are figured by using one of the different closeness measures, and a while later, these comparable qualities are used to foresee examinations for customer thing sets not present in the dataset. The similarity [2] esteems between things are assessed by observing all of the customers who has evaluated both the things. In this, things are suggested by looking at of comparative clients and empower investigation of various substance. It resembles the announcement - Tell me what is well known among my companions. Countless famous recommender frameworks utilizes community oriented separating and functions admirably when clients and things are settled.

Content-Based [3]: The client gives the thing highlight and his inclinations at the season of the review or loved by the client. The prescribed things are favored by the client decision in substance based separating. In this system [4], cold start is an issue for the new client. It can't incline toward class things because of comparable things based technique. It resembles Show me a greater amount of what I loved The Feature representation of items on the basis of content analysis is vector space and topic modeling.

Hybrid [5]: It is a combination of more than one recommendation techniques to predict item for the users. Basically, it is overcome the drawback of limited user-item rating problem of collaborative filtering techniques and cold start problem in content-based techniques. It is possible to combine both techniques to achieve hybrid techniques.

A. Web Scraping

Web Scraping[6] (Scraping or Web Data Extraction or Web Harvesting[7]) is a system utilized to fetch a lot of information from sites whereby the information is extricated and spared to a nearby record in your PC or to a database in the table (spreadsheet) design. Web scraping is the way toward fetching information from sites. All the activity is completed by a bit of code which is known as a "spider". In the first place, it sends a "GET" question to a particular site.

At that point, it parses a HTML record dependent on the got outcome. After it's done, the scrubber looks for the information you need inside the report, and, at long last, changes over it into the predefined format.

B. Data Cleaning

Data cleansing or data cleaning[8] is the way toward distinguishing and adjusting (or expelling) degenerate or off base records from a record set, table, or database and alludes to recognizing fragmented, mistaken, off base or unessential parts of the data and afterward supplanting, altering, or erasing the filthy or undesirable data. Data cleansing might be performed intelligently with data wrangling instruments, or as group preparing through scripting.

C. Bags-of-Words

The bag-of-words model [9] is a way of representing text data when modeling text with machine learning algorithms. The bags-of-words demonstrate is a rearranging portrayal utilized in natural language processing and information retrieval (IR). In this model, a content, (for example, a sentence or an archive) is spoken to as the sack (multi-set) of its words, dismissing syntax and even word request yet keeping variety.

III. PROPOSED METHOD

A. Data Collection:

In this step, we crawl various websites that provide Indian cuisines. Web scraping is done using scrapy1.5.1 framework. Since the collected dataset is not well formatted hence we applied data preprocessing techniques in the collected dataset.

```
In [12]: import pandas as pd
In [13]: data = pd.read_csv("data.csv")
data
Out[13]:
```

	recipe_of	ingredients	steps
0	Vat vake Aloo Recipe	Refined oil, 5-6 Potatoes, Rock Salt, 1/2 lb...	Simply boil some potatoes with vat salt. Heat...
1	Hot Special Matar Kofta Recipe	4 Large Potatoes, boiled, 250 Gram Paneer (c...	Refrigerate the boiled potatoes for 4 to 6 hou...
2	Aula Walnut Cookies Recipe	1/2 Cup Butter, 1/2 Cup Caster Sugar, 1 Cup Aul...	Take butter in a big bowl. Add caster sugar to...
3	Biryani Chicken 65 Recipe	4 Pieces Chicken Breasts (cut into strips), 2 l...	Take freshly cut strips of Chicken Breast in a...
4	Bangra Bangra Masala Recipe	2 tsp Curry, 2 tsp Coriander Seeds, 2 tsp...	Take Curry, coriander and coriander seeds. Dry th...
5	Strawberry Margarita Recipe	40 ml Tequila, 20 ml Triple sec, 100 ml...	Mix all ingredients in a shaker and shake well...
6	Valencia Fizz Recipe	100 ml Valencia orange juice, 75 ml Soda, 1 Lem...	Take a moshel glass and first add valencia o...
7	Crumpet Lotus Root and Truffle Cream Recipe	150 gms thin sliced 50 gms corn flour, to taste...	Peel and cut lotus root into thin slices. dust...
8	Kholi Upside Down Cake Recipe	3 eggs, peeled, 1 Tbsp of oil, 1/2 cup of...	Preheat oven to 180° C. Grease a spring bottom...
9	Mushroom Spring Roll with Black Truffle (DMSU)	15 gm carrot cut into small cubes, 20 gm wh...	Blanch all the vegetables together and keep as...
10	Date And Pistachio Muffin Recipe	1/4 Cup Ghee or white butter, 1 1/4 Khand (A...	Preheat the oven to 180 degrees Celsius. Take...
11	Hibiscus Sherbet Recipe	3-4 dried hibiscus, 2 cups boiling water, 250 g...	Pour boiling water in a bowl of hibiscus. Add...

Figure 2: Snapshot of Indian Recipe metadata set collected through web scraping

In preprocessing we remove numbers, punctuations and other elements which is not useful for our purpose.

B. Features extraction for Recommender:

Content based recommendation system recommends based on contents of the matching profile. Our collected dataset has a lot of features like ingredient, steps, time to prepare etc. but we need only a few features to recommend similar recipes. In this step we select the column based on which we will perform the recommendation and drop the other features.

```
In [5]: df = data[['ingredient']]
df.head()

Out[5]:
```

recipes_name	ingredient
Chewy Peanut Butter Cookies Recipe	Butter Peanut butter Castor sugar Brown sugar ...
Escabeche Recipe	babycorn carrot water chestnuts coconut milk p...
Filo Pizza Recipe	filo pastry himalayan salt black pepper saite...
Spider Cake Recipe	Butter, unsalted Cocoa Powder Chocolate Stout...
Tricolor Seekh Kebab Recipe	Yogurt Onions Green chillies Salt Cheese Cori...

Figure 3: Feature extraction

C. Bags-of-Words:

Our recommendation system recommends based on ingredients of a specific recipe. So we select the ingredient column of our dataset and create bags of word for each recipe. Bags of word contain the keywords for each recipe and based on the similarity of those keywords we rank other recipes in decreasing order of their similarity.

```
In [9]: df.head()

Out[9]:
```

recipes_name	bag_of_words
Chewy Peanut Butter Cookies Recipe	dark sesame seeds butter powder peanut eggs ba...
Escabeche Recipe	welsh himalayan green carrot coconut pineapple...
Filo Pizza Recipe	red capsicum salt green salted butter filo bro...
Spider Cake Recipe	soft dark soda powder extract eggs stout butte...
Tricolor Seekh Kebab Recipe	red cheese seeds chilli powder chillies onions...

Figure 4: Bag of words

and generating the cosine similarity matrix

```
In [11]: # generating the cosine similarity matrix
from sklearn.metrics.pairwise import cosine_similarity
cosine_sim = cosine_similarity(count_matrix, count_matrix)
cosine_sim

Out[11]: array([[1.          , 0.          , 0.05407381, 0.46169026, 0.16222142],
 [0.          , 1.          , 0.30151134, 0.          , 0.35176324],
 [0.05407381, 0.30151134, 1.          , 0.05270463, 0.27777778],
 [0.46169026, 0.          , 0.05270463, 1.          , 0.10540926],
 [0.16222142, 0.35176324, 0.27777778, 0.10540926, 1.          ]])
```

Figure 5: Cosine Similarity matrix

D. Recommender:

In the recommender part we took recipe name as input and output similar recipes. First, we find the index of the recipe which is imputed by user and then we create a series with similarity score using cosine similarity matrix. Then we get the index of the top 2 most similar recipes and recommend those recipes to the user.

E. Result:

We input "Mix Veg" as input and our recommender system recommends "Samosa" and "Rajma Chawal".

```
In [60]: def recommendations(title, cosine_sim = cosine_sim):
    recommended_recipe = []
    idx = indices[indices == title].index[0]

    score_series = pd.Series(cosine_sim[idx]).sort_values(ascending = False)

    top_2_indexes = list(score_series.iloc[1:3].index)
    for i in top_2_indexes:
        recommended_recipe.append(list(df.index)[i])
    return recommended_recipe

In [61]: recommendations('Mix Veg')
Out[61]: ['Samosa', 'Rajma Chawal']
```

Figure 6: Recommender

F. Performance

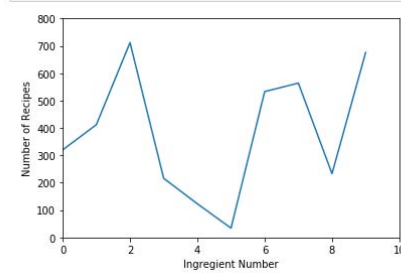


Figure 7: Graph

The above graph shows the comparison between the numbers of recipes suggested by the recommender system with respect to a particular main ingredient. Some ingredient like rice appears in a huge number of results due to its ubiquitous presence in Indian cuisine.

Performance of our system depends on the quality of dataset used to train it. Preprocessing of dataset consumes the most of the time. With the help of more structured dataset our system will able to recommend more relevant recipes to the user.

Our work is a demo and has not been yet opened to user evaluation. However, beta users have tried the system and have reacted positively to its results, as well as suggested improvements, which we have been working on.

IV. CHALLENGES & ISSUES

- Cold Start [10]: This issue happens when new clients enter the system or new things are added to the index. In such cases, neither the ingredients of the new clients can be anticipated.
- The client or guest cold start essentially implies that a recommender meets another guest which has no liked food. Since there is no client history about the user, the recommender doesn't know the individual preferences of the user.
- Heterogeneity [11] of website structure: In spite of the fact that websites are progressively getting to be intelligent and easy to use, this has the turnaround impact on web crawlers. Truth be told, websites with a great deal of dynamic coding rehearses are not in any way crawler well disposed.

V. CONCLUSION AND FUTURE WORK

The conventional food of India has been broadly refreshing for its remarkable utilization of herbs and flavors. Indian food is known for its substantial arrangement of dishes. In this paper, we presented a method for Indian cuisine recommendation using ingredients matching of cuisine and liked food. For this, we did web scraping to make the database of Indian cuisine and collect information all about the all cuisine recipes and used ingredients.

The above issues such as cold start need to be addressed. One of the ways in which we can do this is by linking each user to

their social network profiles and suggest recipes liked by their friends. Heterogeneity can be addressed by building better, more dynamic crawlers.

So, it will be possible in future that enhance the food recommendation by using hybrid approach [12] and web crawling methods where the extracted meta-data is more. [13] Future improvements could include making suggestions based on the geographical location where the cuisine originated, or based on the particular chef whose dishes the user likes. The system could also leverage the user's location to suggest specialty dishes found in nearby restaurants.

REFERENCES

- [1] Wang, Haoyu, et al. (2018) "A Stock Recommendation System Using with Distributed Graph Computation and Trust Model-Collaborative Filtering Algorithm." 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). IEEE)
- [2] Thakkar, Priyank, et al. "Combining User-Based and Item-Based Collaborative Filtering Using Machine Learning." *Information and Communication Technology for Intelligent Systems*. Springer, Singapore, 2019. 173-180.
- [3] Pereira, Nymphia, and Satishkumar 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.
- [4] Pereira, Nymphia, and Satishkumar 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.
- [5] Qian, Yongfeng, et al. "EARS: Emotion-aware recommender system based on hybrid information fusion." *Information Fusion* 46 (2019): 141-146.
- [6] Kolla, Bhanu Prakash, and Arun Raja Raman. "Data Engineered Content Extraction Studies for Indian Web Pages." *Computational Intelligence in Data Mining*. Springer, Singapore, 2019. 505-512
- [7] Patel, Ankit Dilip, and Yogesh Kumar Sharma. "Web Page Classification on News Feeds Using Hybrid Technique for Extraction." *Information and Communication Technology for Intelligent Systems*. Springer, Singapore, 2019. 399-405.
- [8] Goswami Santarsi et al. "A review on application of data mining techniques to combat natural disasters." *Ain Shams Engineering Journal* 9.3 (2018): 365-378.
- [9] Zhao, Rui, and Kezhi Mao. "Fuzzy bag-of-words model for document representation." *IEEE Transactions on Fuzzy Systems* 26.2 (2018): 794-804.
- [10] Sang, Jitao, Ming Yan, and Changsheng Xu. "Understanding Dynamic Cross-OSN Associations for Cold-start Recommendation." *IEEE Transactions on Multimedia* (2018).
- [11] Khan, Sadik, Yashpal Singh, and Kalpana Sharma. "Role of Web Usage Mining Technique for Website Structure Redesign." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 3.1 (2018).
- [12] Logesh, R., and V. Subramaniaswamy. "Exploring Hybrid Recommender Systems for Personalized Travel Applications." *Cognitive Informatics and Soft Computing*. Springer, Singapore, 2019. 535-544.
- [13] Sun, Y., Fang, M. and Wang, X., 2018. A novel stock recommendation system using Guba sentiment analysis. *Personal and Ubiquitous Computing*, 22(3), pp.575-587.