

Discovery of Recipes Based on Ingredients using Machine Learning

¹M. Salomi Samsudeen, ²R. Arun Kumar, ³B. Avinash, ⁴A. Maria Santhosh Prabhu

Abstract- *There are millions of recipes which are uploaded on websites. But still, the user cannot find what dishes can be cooked by using the ingredients the user has. A deep learning approach is focused on finding out the cuisine involved in a certain recipe through various ingredients used in it. Some ingredients have a unique feature which will be available only at a certain part of the country. A novel method is used to find out the best-suggested recipe through some available ingredients based on use. A common data set is created which consists of a global recipe and their procedure to cook it. A recipe is suggested as well as a web link to view the recipe procedure. A recommendation method is used in which the ingredients available by the user is taken as input and analyzation process is done with the help of data-set collected, and the appropriate dishes or recipes is recommended to the user by Machine Learning using K-Nearest Neighbors algorithm. The discovered output is visualized in the form of Website which is more convenient and user-friendly.*

Key Words: *Ingredients, Recipes, Machine Learning, K- Nearest Neighbors Algorithm.*

I. INTRODUCTION

In olden days people use books which inform manage the household and prepare food. But nowadays people try new dishes and experiment with food by using popular recipe sharing sites, and cooking websites. Here numerous amounts of recipes are found along side the ingredients and cooking methods. the web enables people to share knowledge. Food recipe sharing platforms are prevalent and receive their due attention from web surfers also as food enthusiasts. Food is an inseparable a part of our lives. it's been observed that ingredients and recipes are often considered when an individual chooses to eat. Influenced by ingredients and style of cooking, a cuisine can have several hundred or thousands of recipes for various dishes. The user currently cooks a dish with regard to recipes on various sorts of information source like book and website. Though many books for cooking has been published, most of the people recently use recipes on Website like "Food network, Jamie Oliver, Food Club". Cooking beginners are encouraged by the convenience to access numerous ingredients and varied recipes. At first, instance when a user must cook a recipe, the recipe name is surfed on the web site. A recipe on Website shows the ingredients that are needed for a dish and therefore the procedure of the cooking. But the matter is, the user cannot identify what are the dishes are

¹ Assistant Professor of English, College of Science and Humanities at Sulail, Prince Sattam Bin Abdulaziz University, KSA

² B.E,Dept.OfComputer Science and Engineering,KPR Institute ofEngineeringand technology,Coimbatore,TamilNadu

³ B.E,Dept.OfComputer Science and Engineering,KPR Institute ofEngineeringand technology,Coimbatore,TamilNadu

⁴ B.E,Dept.OfComputer Science and Engineering,KPR Institute ofEngineeringand technology,Coimbatore,TamilNadu

often cooked by using the ingredients available by the user. To overcome these problems, Machine Learning approach is implemented which enables to suggest the recipes supported the available ingredients by the user. this manner of searching makes the user, the choice of recipes during a smarter way, and makes the household food maker easier. the target is to scale back the choice of recipes in a neater way by using ingredients as input. the most problem is that the user cannot find the recipes by using the available ingredients. It mainly focuses on checking out the cuisine involved during a certain recipe through various ingredients utilized in it. Some ingredients have a singular feature which can be available only at a particular a part of the country. a completely unique method is employed to seek out the best-suggested recipe through a number of the available ingredients with the user. a standard dataset is made which consists of a worldwide recipe and their procedure to cook it. The recipe is usually recommended also as an internet link to ascertain the recipe procedure is provided. The database is made by the scrapping of knowledge from the food websites and data is segregated within the .xls or CSV file format. The obtained data isn't within the proper format and it's preprocessed and converted into a usable form. The ingredients available by the user is taken as an input employing a web search bar and it's compared with a database which is followed by recommendation of the recipe supported the machine learning KNN algorithm.

II. RELATEDWORKS

A deep learning approach is concentrated on checking out the cuisine involved during a certain recipe through various ingredients utilized in it. Some ingredients have a singular feature which can be available only at a particular a part of the country. within the analyzation, it's defined that there are 12 major cuisines within the world and also more. Support vector machine and random forest algorithm are used for the classification of cuisine. The linkage between the recipe ingredients and therefore the identification of the cuisine is completed. Then the problem are often tackled by machine learning algorithms like SVM, neural networks and Naïve Bayes. A framework is given for the typicality analysis of the mixture of the ingredients. The list of ingredients are often arranged by adjusting the typicality value by adding or removing ingredients iteratively. The cooking beginners are encouraged by the convenience to access numerous and varied recipes. The recipes on websites show the ingredients, however, a number of the listed ingredients can't be used for the cooking. a completely unique method is meant to seek out the best- suggested recipe through a number of the available ingredients with us. Here we create a standard dataset which consists of a worldwide recipe and their procedure to cook it. within the proposed system we discover out the suggested recipe also as we show you an internet link to ascertain the recipe procedure. The database is made by scraping technique.

III. METHODOLOGY

We are given the master list of all possible dishes, their ingredients, an identifier, and the cuisine for thousands of different dishes. We have to present a display of clustered ingredients and train a classifier to predict the food.

A. *SYSTEM ARCHITECTURE*

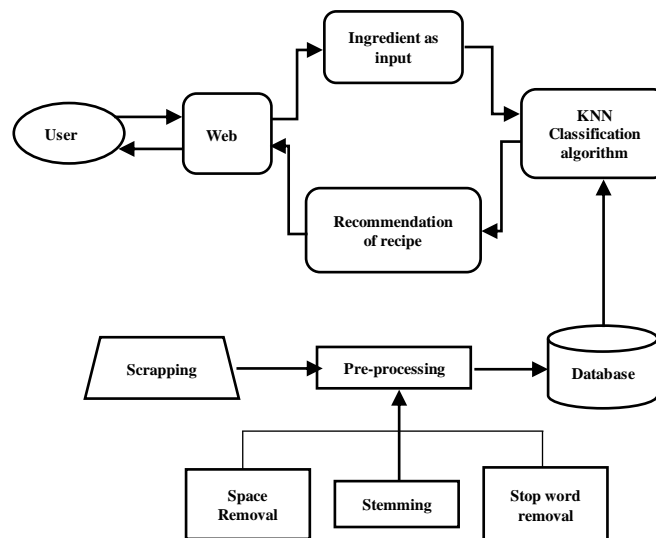


Fig -1: System Architecture Diagram

Fig-1 is the Overall architecture diagram of the Discovery of the Recipe Generation based on machine learning. In the beginning, the database is created by scraping of particular data like Recipe name, ingredients and web link for the recipe's procedure from familiar food websites like "Yummly, Jami Oliver, food club". In general, the scraped data is not in the proper format and it is converted into a usable form by removing space and stop words. The preprocessed data is stored in the .CSV file format. The available ingredients by the user are taken as input through the web interface. Using KNN Classification algorithm the input text is converted into a vector format and classified with the trained data and appropriate class is defined where the matched recipes are recommended to the user. The final output is visualized through websites which is user friendly.

B.DATA PREPROCESSING

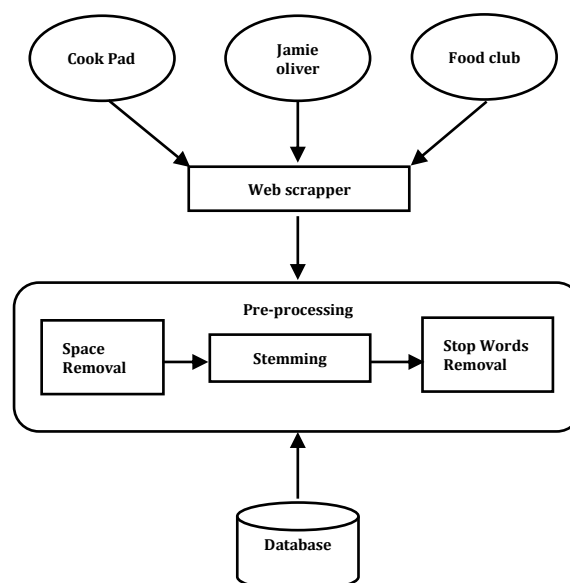


Fig -2: Data Preprocessing Diagram

Fig-2 represents the preprocessing procedures Crawling usually refers to dealing with huge amount of data sets where you develop your own crawlers which crawl to deepest of the web pages. Data scraping on the other side refers to retrieving information or data from any sources. A Web crawler is an online bot which systematically browses the world Wide Web typically for the aim of web indexing. Web search engines and some other sites use web crawling or spidering software to update their web contents or indexes of other sites web content. With Chrome extensions like Agenty Web Scraper, Data Miner, Hence the data is cleaned and converted into a structured format. The preprocessed data is stored in the database in the .xls file or.CSV file format.

WEBSITE CRAWLING

Website Crawling is the automated fetching of web pages by a software process, the purpose of which is to index the content of websites so they can be searched. The crawler analyzes the content of a page looking for links to the next pages to fetch and index Web scraping, often called web crawling or web spidering, or “programmatically going over a collection of web pages and extracting data,” is a powerful tool for working with data on the web. With a web scraper, you can mine data about a set of products, get a large corpus of text or quantitative data to play around with, get data from a site without an official API, or just satisfy your own personal curiosity. You can build a scraper from scratch using modules or libraries provided by your programming language, but then you have to deal with some potential headaches as your scraper grows more complex. For example, you’ll need to handle concurrency so you can crawl more than one page at a time. You’ll probably want to figure out how to transform your scraped data into different formats like CSV, XML, or JSON.

C. CUISINE CLUSTERING



Fig- 3 Cuisine Clustering Example

Fig-3 represents cuisine clustering. Clustering is an unsupervised machine learning approach, but can it be used to improve the accuracy of supervised machine learning algorithms as well by clustering the data points into similar groups and using these cluster labels as independent variables in the supervised machine learning algorithm.

Supervised Machine Learning is used instead of a unsupervised ML like clustering. Examples of Clustering Algorithms 1. Affinity Propagation, 2. Agglomerative Clustering, 3. BIRCH, 4. DBSCAN, 5. K-Means, 6. Mini-Batch K-Means, 7. Mean Shift, 8. OPTICS, 9. Spectral Clustering, 10. Gaussian Mixture Model. Each algorithm offers a different approach to the challenge of discovering natural groups in data. There is no best clustering algorithm, and no easy way to find the best algorithm for your data without using controlled experiments. example K means clustering: K-Means Clustering may be the most widely known clustering algorithm and involves assigning examples to clusters in an effort to minimize the variance within each cluster. it is the process for partitioning an N-dimensional population into k sets on the basis of a sample.

D. KNN ALGORITHM

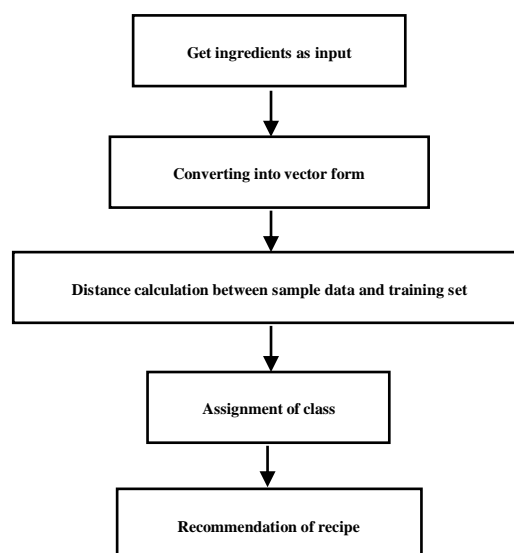


Fig -4: KNN algorithm Diagram

Fig-4 represents that the KNN classification algorithm is divided into three parts namely Sampling, Training, and Class. In sampling, the ingredients are taken as input in the text format and converted into vector form. In training, the predefined vector values are assigned to the ingredients. Based on the weightage of ingredients of the recipes the class is assigned. The nearest distance between the sampling and training data is compared and the obtained value is assigned to the appropriate class. The assigned class recipes are recommended to the user. Classification is done by a majority vote to its neighbors.

Determine parameter K = number of nearest neighbors.

Calculate the distance between the query-instance and all the training samples. Sort the distance and determine nearest neighbors based on the K -th minimum distance. Gathered the category y to the nearest neighbors. Use simple majority of the category of nearest neighbors as the prediction value of the query instance.

E. WEB DEVELOPMENT (UI)

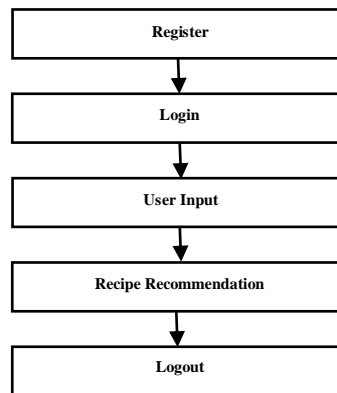


Fig -5: User Interface Diagram

Fig-5 Represents When the user access the website for the first time the registration is done by filling the basic details and login username and password is created. There will be three text box the user can type the three type of ingredients that user have while clicking the find button the appropriate recipes will be recommended the user through web pages.

IV. CONCLUSION

This system will help to select the recipe by using the available ingredients by the user. The selection of the recipe is in a smarter way and user-friendly. Various recipes can be classified according to the ingredients. In future enhancement, User can give input also in image format where the Image Processing is applied in detection of Ingredients. With the advanced machine learning techniques, the discovery of new recipe using available ingredients by the user can be applied in a smarter way.

REFERENCES

- [1] Li, Y., Meng, X., Wang, L., Li, Q. Recipecrawler: Collecting recipe data from www incrementally. In: *Proc. of the International Conference Web-Age Information Management*. 2006, p.263–274.
- [2] M.Ueda, M.Takahata, and S.Nakajima, "User's food preference extraction for personalized cooking recipe recommendation," *Proc. 2nd Workshop on Semantic Personalized Information Management: Retrieval and Recommendation*, pp. 98–105, Oct. 2011.
- [3] Lam, W. and Ho, C.Y. (1998). Using a Generalized Instance Set for Automatic Text Categorization. *SIGIR'98*, pages 81-89.
- [4] Li, H. and Yamanishi, K. (1999). Text Classification Using ESC-based Stochastic Decision Lists. In *Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management*, pages 122-130.

- [5] Joachims, T. (2001). A Statistical Learning Model of Text Classification for Support Vector Machines. In Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval, pages 128-136.
- [6] X. Meng, D. Hu, and C. Li. Schema-guided wrapper maintenance for web-data extraction. In the 5 th ACM CIKM International Workshop on Web Information and Data Management, pages 1–8, 2003.
- [7] Teng, C., Lin, Y.-R., and Adamic, L. A. Recipe recommendation using ingredient networks. CoRR abs/1111.3919.
- [8] Yang, Y. and Liu, X. (1999). A Re-examination of Text Categorization Methods. In Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval, pages 42-49.
- [9] van Pinxteren, Y., Geleijnse, G., and Kamsteeg, P. Deriving a recipe similarity measure for recommending healthful meals. In *Proceedings of IUI '11*, ACM (New York, NY, USA, 2011), 105--114.
- [10] Seki, Y., Kouta, O.. Discriminating practical recipes based on content characteristics in popular social recipes. In: Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014, p. 487–496
- [11] Hamada, R., Okabe, J., Ide, I., Satoh, S., Sakai, S., Tanaka, H.. Cooking navi: Assistant for daily cooking in kitchen. In: Proc. of the 13th annual ACM international conference on Multimedia. 2005, p. 371–374.
- [12] Blansch'é, A., Cojan, J., Dufour-Lussier, V., Lieber, J., Molli, P., Nauer, E., et al. Taaable 3: Adaptation of ingredient quantities and of textual preparations. In: Proc. of the 18th International Conference on Case-Based Reasoning. 2010, p. 189–198.
- [13] Using kNN Model-based Approach for Automatic Text Categorization Gongde Guo¹, Hui Wang¹, David Bell ², Yaxin Bi ², and Kieran Greer ¹ School of Computing and Mathematics, University of Ulster Newtownabbey, BT37 0QB, Northern Ireland, UK ¹{G.Guo, H.Wang, Krc.Greer@ulst.ac.uk} School of Computer Science, Queen's University Belfast Belfast, BT7 1NN, UK ²{DA.Bell, Y.Bi}@qub.ac.uk
- [14] Melville, Prem, Raymond J. Mooney, and Ramadass Nagarajan. "Content-boosted collaborative filtering for improved recommendation" In *Aaai/iaai*, pp. 187-192. 2002.
- [15] 19th International Conference on Knowledge-based and Intelligent Information and Engineering Systems Alternative-ingredient Recommendation Based on Co-occurrence Relation on Recipe Database