

# Recipe Recommendation System using Machine Learning Models

Suyash Maheshwari<sup>1</sup>, Manas Chourey<sup>2</sup>

<sup>1</sup>Student (B.E.), Information Technology, Sinhgad College Of Engineering, Pune, India

<sup>2</sup>Student (B.E.), Information Technology, Sinhgad College Of Engineering, Pune, India

\*\*\*

**Abstract** - Completing recipes is a challenging task, as the success of ingredient combinations depends on a multitude of factors such as taste, smell and texture. We can pair various ingredients with each other and generate a dish. However, it is not possible to perform these experiments in kitchen. Hence, science is used to determine various ingredient pairs. Due to the increasing availability of data in online databases, data mining and machine learning methods are starting to play a prominent role in food consumption analysis and food preference modeling. Here, we will use two machine learning models-vector space model and Word2Vec model to find top ingredient pairs from different cuisines and to suggest alternate ingredients. The focus is on the Indian cuisine. Indian cuisine is very vast and diverse and hence it is difficult to find patterns and generate pairs.

**Key Words:** Data analysis, Machine learning, Web scraping.

## 1. INTRODUCTION

Everyday food is an important part of our daily lives and, currently, there are many internet sites that help us plan these meals. Allied to the growth in the amount of food data such as recipes available on the internet there is an increase in the number of studies on these data, such as recipe analysis and recipe search. However, there are few publicly available resources for food research especially for the Indian cuisine; those that do exist do not include a wide range of food data or any meal data that is, likely combinations of recipes. In this project, we will construct a food recommendation system using big data analytics.

### 1.1 Aim And Objectives

The aim is to develop a tool that would help in pairing various ingredients from different cuisines and to suggest an alternate ingredient.

The objective is to help in innovation of new dishes and to help people allergic to certain ingredients by recommending alternate ingredients.

## 2. LITERATURE SURVEY

### 2.1 Data-driven recipe completion using machine learning methods

In this paper Marlies De Clercq, Michiel Stock, Bernard De Baets and Willem Waegeman [1] have proposed that machine learning methods can be used to build models to complete a recipe. The machine learning model suggested is

Non negative matrix factorization. The characteristic of this method is that it gives a linear non negative part based representation of data. But, it does not consider the information about flavor components and hence the results are not accurate.

### 2.2 Alternative-ingredient Recommendation Based on Co-occurrence Relation on Recipe Database

Ryosuke Yamanishia, Naoki Shinoa, Yoko Nishiharaa, Junichi Fukumotoa and Aya Kaizakib[2] have proposed a recommendation method of alternative ingredients based on co-occurrence relation on recipe database. This paper proposes two algorithms to recommend alternative ingredient. Through the cooking and tasting experiments, it was confirmed that the each of the proposed methods were effective for each intended purpose. Through cooking demonstration with the recommended alternative ingredient and subjective evaluation experiments, it was confirmed that the both algorithm recommended acceptable ingredients for over 90 percent. However, algorithm 1 might not recommend alternative-ingredient similar to the exchange ingredient.

## 3. METHODOLOGY

### 3.1 Web Scraping

Web Scraping is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in computer or to a database in table format. Web scraping is used in order to extract the data from Yummly.com for later retrieval and analysis. Yummly is a website that provides recipe recommendations personalized to the individual's taste, semantic recipe search. The data is taken from the Yummly websites which contains various recipes, ingredients and also the flavour components of all the ingredients. This data is then stored into the csv file for further operations. Pkl files are also created. A PKL file is pickled to save space when being stored or transferred over a network and is unpickled and loaded back into program memory during runtime. After the data is scraped, it is cleaned and preprocessed to remove anomalies from the data.

### 3.2 Ingredient Pairing

The ingredient pairing module is used to pair the ingredients of the selected recipe with other ingredients to innovate a new dish. For this, Tf-idf and Cosine similarity are used. Tf-idf (Term frequency - Inverse document frequency) assigns weights to ingredients based on their flavor components.

These Tf-idf scores are then used to calculate cosine similarity for two ingredients. The model used for calculation of cosine similarity is known as vector space model. Based on the similarity, different ingredients from various cuisines are recommended for pairing.

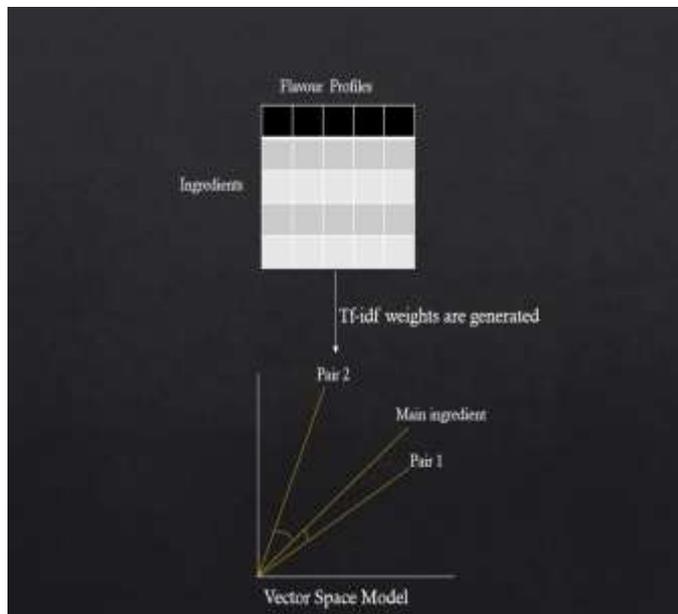


Fig -1: Ingredient Pairing

A matrix is formed for all ingredients present in dataset and all the flavor components. Each ingredient is treated as a single document and flavor components are treated as terms of documents. Tf-idf score is calculated for all ingredients on the basis of this matrix. Then, for the selected ingredient, matching pair is found by calculating the cosine of the angle between two ingredients. The lesser the angle, the better the pairs are. This model is defined as Vector space model.

### 3.2 Alternate Ingredient Recommendation

Alternative Ingredient recommendation is used to recommend an alternative ingredient if a particular ingredient is not present or cannot be used in the recipe. Word2vec model is created to perform this function and recommend alternative ingredients. Word2vec takes as its input as large list of ingredients and produces a vector space, typically of several hundred dimensions, with each unique ingredient in the corpus being assigned a corresponding vector in the space. Vectors are positioned in the vector space such that ingredients that share common contexts in the corpus are located in close proximity to one another in the space. Gensim library is used to automatically extract semantic topics from documents, as efficiently as possible. Downsampling is also done to help reducing data size. At last, most similar() method is used to find top-n most similar ingredients. This method computes cosine similarity between a simple mean of the projection weight vectors of the given words and the vectors for each word in the model. This method returns the sequence of ingredients and their similarity value to the ingredient that is to be exchanged.

## 4. FOOD PAIRING HYPOTHESIS

Food pairing hypothesis states that two ingredients sharing common flavor profile taste well when used together in a recipe. In the Indian cuisine vegetables, fruits, cereals, pulses, nuts support the food pairing hypothesis whereas spices and dairy products do not support this hypothesis as shown in the figure below. An analysis was done to find out the relationship between ingredients and their flavor profiles. It showed a linear dependency.

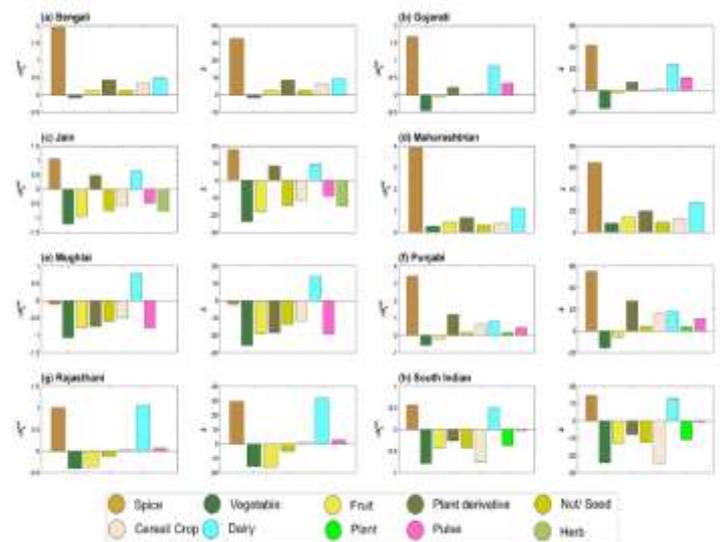


Fig -2: Analysis of Indian Cuisine

## 5. STATE DIAGRAM

A state diagram is used to represent the condition of the system or part of the system at finite instances of time. It's a behavioral diagram and it represents the behavior using finite state transitions. A state diagram is used to model the dynamic behavior of a class in response to time and changing external stimuli.

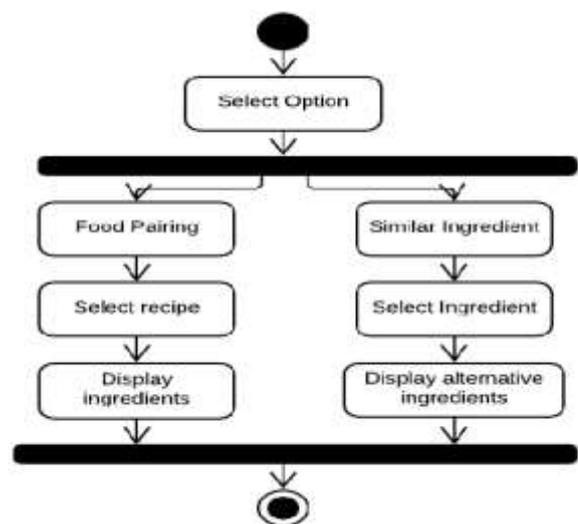


Fig -3: state diagram



[2] Ryosuke Yamanishia, Naoki Shinoa, Yoko Nishiharaa, Junichi Fukumotoa, Aya Kaizakib, "Alternative ingredient Recommendation Based on Co-occurrence Relation on Recipe Database". *Procedia Computer Science* 60 (2015) 986-993

[3] Jain, Anupam, N. K. Rakhi, and Ganesh Bagler. "Analysis of food pairing in regional cuisines of India." *PloS one* 10.10 (2015): e0139539.