International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

RASOI - Recipe Generator

Rabiya Idrishi[1], Suraj Iyer[2], Snehal Waghela[3], Savita Lohiya[4]

[1]*Information Technology Department, SIES Graduate School of Technology*
[2]*Information Technology Department, SIES Graduate School of Technology*
[3]*Information Technology Department, SIES Graduate School of Technology*
[4]*Information Technology Department, SIES Graduate School of Technology*

Abstract - In everyday life, scenarios present themselves wherein one may want to cook something, but due to certain limitations like inaccessibility of ingredients, minimal guidance, lack of inspiration may deter them from accomplishing the task. The solution is a machine learning algorithm that takes into consideration what one has in the fridge or pantry, to generate a recipe utilizing those available ingredients. To be precise, it takes an ingredient in the form of a string, converts it into a vector representation. Further it uses a recurrent neural network to generate a full recipe including a title, a list of ingredients and a set of instructions to be followed. Our aim with this project is to bridge the gap by generating suggestions based on the user input.

Key Words: lstm, natural language processing, react recipe generation, rnn, text generation

1. INTRODUCTION

There may be instances when a person desires unique, delicious, healthy, or even presentable cuisines, and it must be homemade because the options for getting one from outside may be limited, such as during the current pandemic period (Covid-19)[5]. An ideal solution to this may be the development of an application that allows the user to experiment with and prepare a variety of new cuisines that include those elements[11]. As a result, a system could be created that takes ingredients as input and generates the best-matched recipe from a massive list of the most similar recipes. Our project intends to educate users about the numerous recipes that can be prepared using the ingredients that they provide. It will also introduce new ideas to the culinary arts. It will also help control and prevent food-wastage[6].

Furthermore, this initiative would provide an excellent opportunity for people to learn and develop their culinary knowledge[7]. Since technology has progressed tremendously people tend to look for recipes online on the web instead of buying a book. A website like ours will help them get the recipe with the ingredients they wish to incorporate in their dish making it customized to their needs specifically. Online presence of such a food library will help in this age with COVID posing a threat to people who leave their homes. It will save the time spent on referring various websites for the desired recipe.

Moreover, it will aid in the exploration and dissemination of a single region's flavor throughout the world. People can be made aware of the various dishes belonging to various cultures and regions throughout the globe. It will also introduce new ideas to the culinary arts[7].

2. LITERATURE SURVEY

Review Stage Discovery of Recipes Based on Ingredients using Machine Learning[1]:

This paper outlines how a recipe can be found by searching with ingredients. There are millions of recipes available on the internet. However, the user is unable to determine which foods can be prepared with the ingredients at hand. A deep learning technique focuses on determining the cuisine featured in a dish by looking at the numerous elements used. Some ingredients have a distinct trait that will only be available in a specific region of the country. A innovative strategy is employed to determine the best-suggested dish based on the use of some available ingredients. A common dataset is produced that contains a global recipe and their cooking technique. A recipe is recommended, along with a link to a video of the procedure. The user's ingredients are utilized as input, and an analysis process is carried out with the help of the data set acquired, and the right dishes or recipes are offered to the user by Machine Learning using the K-Nearest Neighbors algorithm. The results are presented in the form of a website, which is more convenient and user-friendly. The database is built by scraping specific data like the dish name, ingredients, and a web link to the recipe's method from well-known food websites such as "foodnetwork, Jami Oliver, food club," . In general, scraped data is not in the correct format, thus it is changed into something useable by deleting spaces and stop words. The pre-processed data is saved in.xls or.CSV file formats. The user's accessible ingredients are entered into the web interface as input. The input text is translated into a vector format and classed using the trained data using the KNN Classification algorithm, and a suitable class is defined where the matching recipes are offered to the user. The final product is visualized through user-friendly web pages. This system will aid in the selection of a dish based on the user's available ingredients. The recipe selection is more intelligent and user-friendly. Ingredients can be used to categorize various recipes. The discovery of new recipes using available ingredients by the user can be done more intelligently with advanced machine learning algorithms.

Forage: Optimizing Food Use With Machine Learning Generated Recipes[2]:

The algorithms used to generate recipes are the topic of this paper. They took on the challenge of using word2vec and an LSTM to generate new recipes based on available ingredients in this project. Their biggest hurdle was mass evaluation of this generative model, but they overcame it by using k-means clustering and a human survey to comprehend the output. Because their topic of recipe production using a machine learning algorithm is new, there aren't many relevant previous works. Pic2Recipe (Salvador et al., 2017), a recent effort from MIT, included a recipe (ingredient and cooking instruction) module underneath the picture recognition module. The word2vec representation of ingredients words was used in the recipe module, and the primary learning technique was also a recurrent neural network, or LSTM. Pic2Recipe was created with the goal of recognizing a photo of a food whose recipe the algorithm had seen during training and displaying that recipe to the users. Pic2Recipe employed two different LSTM models for ingredients and instructions, therefore no specific relationships were discovered. The representation was on the sentence level in the two-stage LSTM model for instructions, with each sentence encoding serving as a context for decoding or predicting the preceding and next phrases, allowing accurate remembers of trained instruction sentences. In the United States, food waste is a serious problem. The average annual worth of discarded produce for an American family of four is about $1,600. (EPA, 2015). Forage is a machine learning algorithm that takes into account what you have in your fridge or pantry to come up with a unique dish that incorporates those items. Given an input sequence of basic ingredients, a recurrent neural network (RNN) was constructed to create each recipe. The model's intended output is a collection of instructions. Traditional evaluation criteria does not give much meaningful information because of the project's generative nature. K-means clustering was employed to get a feel of how similar recipes should look, and to see if Forage recipes were reasonable, or outliers that didn't correlate with any of the trusted clusters, or categories of recipes from the dataset. Although the output recipes received an average grammar rating, the ingredients, methods, and title logic did not perform well, and readers were unable to prepare the food. When one-line recipes were removed, the scores improved slightly, but they remained below the targeted levels.

Ingredient-driven Recipe Generation Using Neural and Distributional Models[3]:

Cooking instructions are a good example of instructional language. Developing systems that autonomously analyze and develop recipes necessitates investigating and improving various semantically challenging problems. The project was inspired by Dan Jurafsky's book "The Language of Food" (Jurafsky, 2014), particularly after learning about Rozin's flavour principle (Rozin, 1973), which states that western cultures tend to pair up ingredients that share the same flavor, whereas eastern cultures combine ingredients that do not have overlapping compounds, and the developers thought that this difference might characterize a cuisine and also be a good path to creativity by intersecting subsets of ingredients from different cultures. This is an example of how natural language production can be viewed as a subset of the larger topic of computational creativity. The most crucial aspect of creativity is coming up with new ideas that have never been thought of before.

This reveals itself in both the creation of unseen sequences of imperative sentences and actors in instructional language generation. However, this must be weighed against plausibility, so that the generator's output does not deviate too far from what is deemed typical. Most earlier ways to train narrative generators before neural language modelling become the most prevalent method of addressing this job are based on making plans hand-written rules that impose similarity. A data-driven generation system will produce output that closely resembles the corpus on which it was trained. However, in this study, the developers intend to use the semantics of ingredients and cooking events, such that the suggested system can be modified to create both unexpected component combinations and culinary instructional steps. The purpose of this master's thesis is to use massive unannotated corpora to produce chains of cooking events that fit a list of ingredients given ahead of time without direct supervision, as well as to experiment with generating novel chains of events that are nonetheless coherent on the same list.

Web Scrapping[4]:

Web scraping is a technique to extract data from the World Wide Web and save it as a database or file system for later analysis. The process of scraping data can be divided in two steps; acquiring web resources and then extracting the desired information. Web scraping starts by composing a HTTP request to the website we want to scrape the data from. Once the request is received and processed by the website, the requested data can be retrieved from the website. The information retrieved can be in multiple formats such as XML or JSON format, images, audio, video. After the data is downloaded the process continues to reformat, parse the data in a structured format. There are two important modules used for web scraping - a HTTP request like Urllib2 or selenium for composing the data and modules such as Beautiful Soup or Pyquery to parse and extract information from raw HTML codes. The Urllib2 module defines a set of functions to deal with HTTP requests, such as, redirection, authentication, cookies, etc. Selenium is a web browser wrapper that builds a web browser and helps users to automate browsing of a website with the help of programming. BeautifulSoup used for data extraction is designed to scrap HTML and XML pages. It provides python functions for searching, modifying and navigating a parse tree. It also has a toolkit to decompose HTML file and extract required information via lxml or html5lib.

3. OBJECTIVES

a. To help users cook great food with simplified recipes based on the ingredients available with him/her which are easy to understand and even easier to make.
b. To develop a searching algorithm for the web app that takes the input from the user i.e., an ingredient and sorts the recipes in real time.
c. Today, users are increasingly drawn towards the idea of cooking meals from the comfort of their homes. Finding the proper source for recipes to cook for a beginner is difficult, that’s where our web app comes into the picture.

4. MODELLING

As the Ingredients-to-recipe dataset isn't available online, it will have to be scraped, then pre-processed to make it training-ready. Filtering out unfinished recipes and inserting stop words after the title, ingredients, and instructions will all be part of the dataset's pre-processing. Since the application necessitates the creation of a model that can assist in the generation of recipes, deep learning understanding will be necessary.

In the next phase, the dataset has to be trained. Given an ingredient as an input, a recurrent neural network(RNN) must be created to curate each recipe. For this, LSTM RNN will be used because of its capacity to remember data sequences while capturing remote dependencies and generating new data. A character-based LSTM has to be created that predicts the next character in a sequence based on the previous characters sampled. The model is expected to generate whole recipes, word by word, from an initial single input ingredient provided by the user.

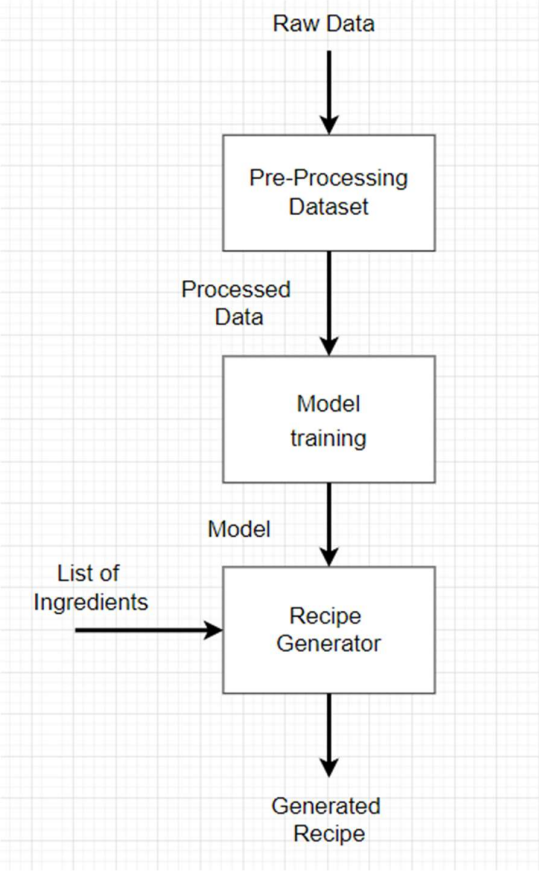

Fig -1: Fetching Recipes

Another duty is to create a user-friendly UI for the website. React.js, JavaScript, and CSS will be used to do this. The benefit of such an application is that it can be accessible from anywhere and that it is a 24/7 application, meaning that it is always available, regardless of the user's location or other considerations. A web application can be used to deliver functionality to the end user efficiently and without putting a strain on their systems. To avoid data and memory leaks, processing can be done on the server.

Once the model is complete, we can add the Ingredients-to-Recipe features and also other interesting socializing features that allow people to share recipes made by them online.

5. WORKFLOW

In this project we are going to generate recipes by entering the ingredients. We have divided our workflow into 3 different phases i.e. Data Scraping, Model Training and Front-end Development.

In the first phase the important thing is making a good database with all the recipes. We have scraped 3 websites i.e. Allrecipes, Epicurious and Food Network. We basically wanted all the data from these websites so we have curated a scraping algorithm using the BeautifulSoup module in python that works for these websites. In scraping an http request is continuously sent to the website that is being scraped so we require fast Internet and it is also very time consuming. In all we have scraped 1,25,000 recipes approximately. We have pre-processed this data and removed all the unwanted/incomplete data that was scraped.

For the next phase we have trained the dataset. Given an input sequence of basic ingredients, we created a recurrent neural network (RNN) to create each recipe. The model's intended output is a collection of instructions. We used k-means clustering to get a sense of how similar recipes should look, and we were able to determine whether our recipes were reasonable, outliers that did not correlate with any of the trusted clusters, or as one large corpus of recipes, where one training instance is a fixed-length text sequence (25 or 50 words). We were able to generate our training batch as a single tensor of form batch size sequence length features because the input had a fixed size.

We investigated two techniques of choosing recipes from the trained model, beam search and greedy search, in our models that generated the next most likely word in a sequence based on prior words. Beam search differs from greedy search in that it expands the x most promising nodes and maintains track of probabilities for each beam, whereas greedy search predicts the next word based on the single most probable word in the probability distribution at time t. We discovered that beam search with a width greater than 1 produces lower quality recipes, thus the findings are limited to greedy sampling.

The LSTM RNN was chosen because of its capacity to "remember" data sequences while capturing remote dependencies and generating new data. We created a word-based LSTM that generates words in a sequence based on the previous words sampled. The model generated whole recipes, word for word, from an initial feeder sequence of ingredients. A vector of memory cells was used to achieve "long term" memory. LSTMs can determine whether to overwrite a memory cell during training, use it for the current prediction, or leave it alone until the next time step.

Another task was to design a user-friendly website frontend. This was done using React.js, JavaScript, and CSS. The advantage of such an application is that it can be accessed from anywhere and is an anytime application- i.e., it is always available, regardless of the location of the user or other factors. A web application can be used to efficiently deliver functionality to the end user without burdening their systems. Processing can be done on the server to prevent data and

memory leaks. We can add the ingredients-to-recipe functionality to it once the model is ready.

6. RESULT

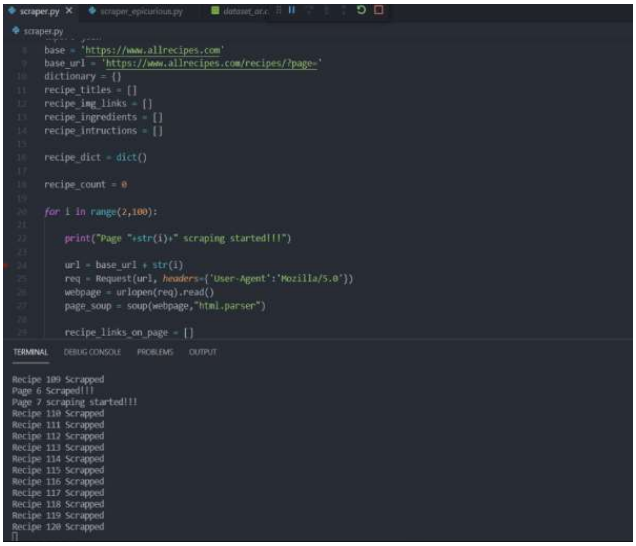

Fig -2: Web Scrapping

The BeautifulSoup module in python was used to develop a scraping module in python to scrape the recipes.

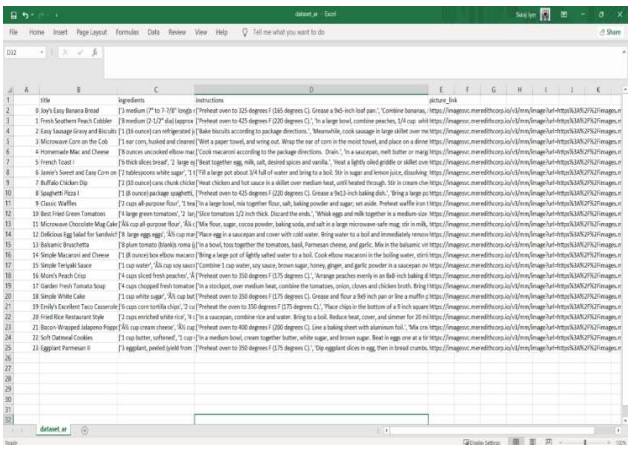

Fig -3: Scrapping Output

We scraped nearly 1,25,000 recipes from the three mentioned website.

Fig -2: Design of Web Application - 1

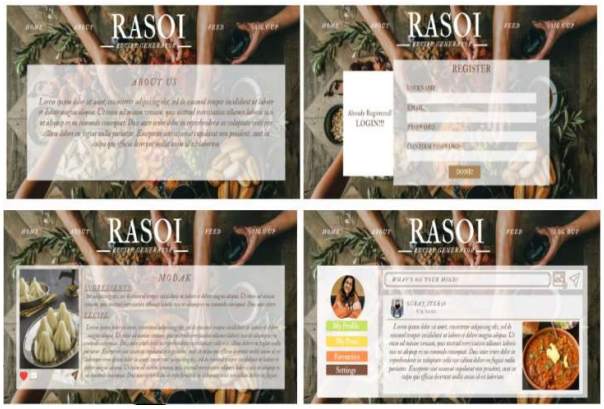

Fig -2: Design of Web Application - 2

We developed prototypes of the end designs.

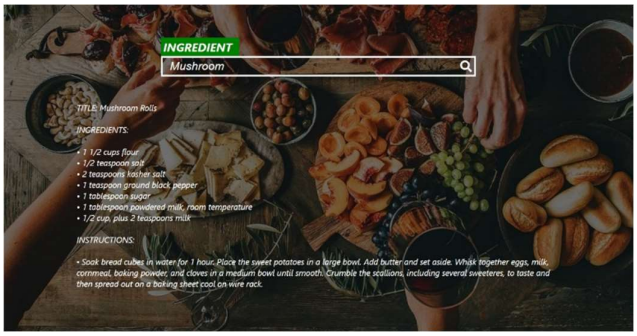

Fig -2: Output of generated recipe

7. CONCLUSION

Our project Rasoi-Recipe Generator is a dynamic web application that incorporates deep learning techniques to search for recipes. For which a lot of recipes were scrapped and processed. We have trained the model in such a way that when we input ingredients, it will generate a recipe. Developing a user-friendly website and having good accuracy were few other objectives achieved from this project.

ACKNOWLEDGEMENT

We wish to express our deep sense of gratitude to thank our project guide prof. Savita Lohiya for providing timely assistance to our query and guidance.

REFERENCES

[1] S. Praveen, M.V. Prithivi Raj, R. Poovarasan, V. Thiruvenkadam, M. Kavinkumar - Discovery of Recipes Based on Ingredients using Machine Learning - https://www.irjet.net/archives/V6/i3/IRJET-V6I328.pdf

[2] Angelica Willis, Elbert Lin, Brian Zhang - Forage: Optimizing Food Use With Machine Learning Generated Recipes – http://cs229.stanford.edu/proj2017/final-reports/5244233.pdf

[3] https://lct-master.org/getfile.php?id=2194&n=1&dt=TH&ft=pdf&type=TH

[4] Bo Zhao- Web Scraping https://www.researchgate.net/publication/317177787_Web_Scraping

[5] https://www.worldbank.org/en/topic/agriculture/brief/food-security-and-covid-19

[6] https://www.fao.org/3/i2697e/i2697e.pdf

[7] https://www.humanium.org/en/indian-food-diversity/#:~:text=In%20those%20regions%2C%20food%20is,culinary%20subtle%20dishes%20are%20created.

[8] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5348196/#:~:text=Nutritional%20awareness%20is%20also%20related,reasons%20for%20eating%20such%20foods

[9] https://bbsrc.ukri.org/research/briefings/food-nutrition-health/

[10] https://www.breastcancer.org/tips/nutrition/healthy_eat/nutrients

[11] https://hdsr.mitpress.mit.edu/pub/wjrl1qsq/release/2

[12] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4259868/