AutoBasket

”With AutoBasket, it’s never been easier to get everything you need for a great meal!”

Dhiraj Kumar Sah
Dhru Sanjay Prajapati
Harshil Bhavsar
Keerat Singh

Reporting Date: 8th March 2024

TABLE OF CONTENTS

ABSTRACT 2
INTRODUCTION 3
DATA PREPROCESSING 3
1. DATA GATHERING 3
2. DATA PREPARATION 4
3. DATA CLEANING 4
DATA VISUALIZATION 5
RECOMMENDATION SYSTEM 7
MODIFIED FRONT END IN FLASK 8
CONCLUSION 10
REFRENCES 11

ABSTRACT

This project aimed at giving me a recommendation system for Indian food dishes based on user preferences. Data preprocessing process began initially by mobilizing all the data, carrying out thorough cleaning, transformation, and the usage of a smart feature selection strategy for a trustworthy dataset. The group collected the CosyLab, which was elaborated with improvements of abnormalities into single data analysis. In order to make the visualization more powerful, we used word clouds displaying the ingredient distribution, along with graphs representing the leading ingredients in recipes including their compositions.

A suggest system, focused on the notion of the dimensionality reduction and cosine similarity, was created to propose various options of Indian dishes, taking personal taste into account. Principle of employment of Pythonizer Scikit-learn for optimization was applied. The project ended with a Flask frontend being developed that has enhanced user experience through simple and fast functionality. The interface brought about interaction with the recommendation system, in which users were presented with additional tailored recommendations to further enhance engagement. This abstract will elaborate on the month's objective, main points, and the successfully implementation of innovate procedures in making a better recommendation system.

INTRODUCTION

AutoBasket is a company founded by Larry and Veronica Smiles with the purpose of simplifying the grocery shopping experience for individuals and families. Situated in the heart of Toronto, the company's headquarters pulsate with the energy of a city known for innovation and diversity. The company's app automates grocery and recipe lists, linking recipes to required products, and streamlining the shopping process. Today, AutoBasket assists households across the world to save time and energy on their weekly grocery runs, making it easier for busy families and individuals to get everything they need for a great meal.

AutoBasket has hired us as interns to focus on all areas of development for the organization and to help create solutions to common issues within the industry. Our role as interns is not only focused on the knowledge you have learned in school but also on developing your soft skills, including presentations, teamwork, and leadership. This internship will provide us with valuable real-world experience and an opportunity to contribute to the continued success of AutoBasket.

DATA PREPROCESSING

As we already have dataset containing recipes that are according to the Indian community’s preference due to acquisition and curation of dataset at early stage. The data preprocessing step is like the week 2. Data preprocessing appears as labeling ingredients before cooking a meal-so that data is in order and ready for analysis or modeling smoothly. In essence, it involves several fundamental steps: Initially, data cleaning implies finding out and correcting the mistakes, for instance, missing values or things that are inconsistent so that the dataset becomes the best resource for it to be reliable. In the second place, data transformation contains converting data from a format misfitting to analysis apparently including scaling, binary encoding, and the treatment of outliers. These are the processes that simplify the data. Thereby making it more manageable and less complex for proper future analysis.

Besides that, this preprocessing cut beyond just cleaning; it has to do with data science and getting the relevant features represented in a strategic manner. Feature selection and extraction can help in finding the characteristic variables that are mostly informative to the dataset and hence eliminating the noisy attributes and redundant variables that are not related. Moreover, methods like balancing data help avoid biased results, which usually have a majority class preference. Data preprocessing is the core activity of insightful data analysis and high-quality modeling. The importance of data preprocessing is that it provides a solid foundation needed for finding hidden patterns and building accurate predictive models.

1. DATA GATHERING

This week, our major headings of the report were oriented on the development of the recommendation system for dishes based on user-preferred cuisine, especially in the Indian culinary style. We opted to get this data from CosyLab which is an online platform recognized for working extensively with food related data, computational Gastronomy and making innovative inroads to food science. The dataset we used included a vector of key information like name of the dish, ingredients and associated items.

In a Zip packet which is handy, we have the data set which is composed of various CSV files, with each one specializing in different parameters for our analysis. Those files contain the complete information about the recipes, ingredients, aliases of ingredients and substitutes.

2. DATA PREPARATION

These two objectives were met using all dataset files, shoved into the single dataframe that has enough columns to hold the necessary columns. On the data preparation stage, we elaborately checked out each file and fused those datasets in order to have an holistic look at the wealth of information thereat. Yet there were some irregularities we faced in the data; we had to deal with duplicate and scattered data which although happened to be vital for the outputs if not dealt with appropriately can sometimes contribute to the wander of the actual project objectives.

The endurance we are able to take is largely determined by the quality of our data and the streamlining of our workflow. Initially, we obtained complete dataset and after that, we found the redundant features and threw them out from the dataset. Our decision to intentionally do this has a dual benefit: not only does it cut the time savings but also it also ensures that we have the best dataset analyzed for the required purpose. For this week project our aim is to suggest to the user Indian dishes based on the user's dish preferences, In this case, we all decided that ingredients would be the central component which would guide the probability of recommending food items. As such, this method is underpinned by our hypothesis and creates a directed, identifiable framework for carrying quick and consequential analyses that match the Indian recipes.

3. DATA CLEANING

We observed the inconsistency in the data structure and a large presence of empty value with NaN. More on, each dish of ingredients goes in columns, the others row representing the recipe ID. In order to tackle this structural barrier, we went through a revolutionary process that entailed the conversion of these uneven rows of recipe data into more organized columns to present a well-structured overview.

The Recommendation of Indian dishes and their Ingredients raised the need to obtain of data in a specific format. In view of this, we directed a laborious and comprehensive restructuring program, with main goal making it more beneficial. This restructuring, which was based upon our focus on using neural network in its delivery, had the ultimate aim to reveal complex relations between the ingredients. Consequently, the handpicked set of data we collected becomes now arranged in this exact fashion which grants us to present proposals congruent with user preferences.

DATA VISUALIZATION

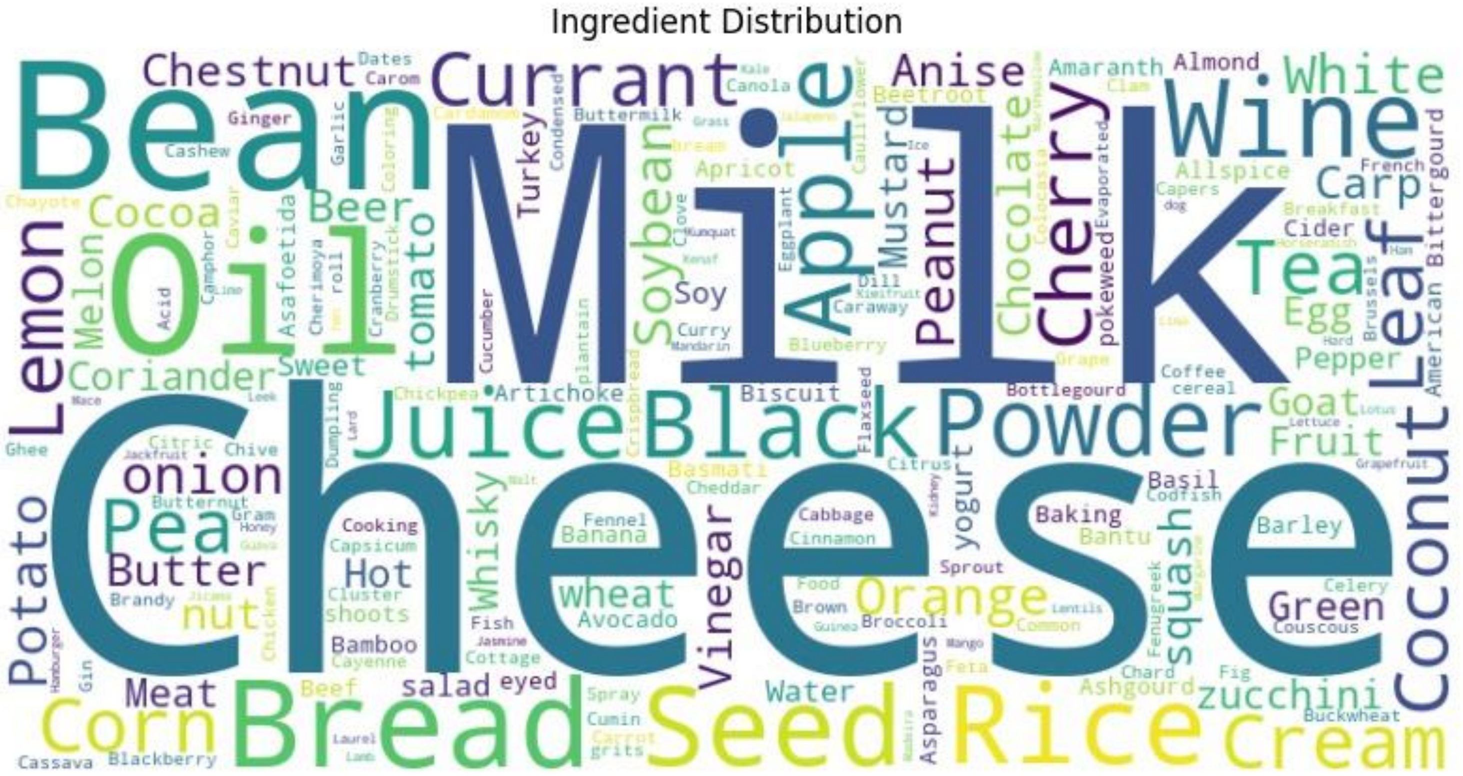


Figure 1 Wordcloud to represent different ingredients

The distribution of ingredients in a dataset of recipes using a word cloud visualization tool. The analysis reveals valuable insights into the most frequently used ingredients in the dataset, which can be useful for identifying patterns, trends, and potential health-conscious recipes. The word cloud visualization tool was used to analyze the distribution of ingredients in the recipe dataset. The tool uses a cloud-like visual representation to display the frequency of each ingredient in the dataset. The size and prominence of each ingredient in the cloud reflect its relative frequency in the data. The word cloud visualization revealed several key insights into the ingredient distribution in the dataset.

1. Milk is the most frequently used ingredient in the dataset, followed closely by cheese and beans. This suggests that dairy products and legumes are important components of many recipes in the dataset.
2. Other frequently used ingredients in the dataset include currant, anise, wine, white rice, cream, coconut, and lemon. These ingredients are well-represented in the dataset and may be important components of many recipes.
3. Some ingredients, such as anise and black powder, are used less frequently in the dataset, but could still be important in certain recipes.
4. Yogurt and juice are also represented in the dataset, which suggests that there may be some health-conscious recipes included.

The word cloud visualization provides a helpful overview of the types of ingredients that are used in the dataset. The analysis reveals that dairy products, legumes, and certain grains and starches are well-represented in the dataset, while some ingredients, such as anise and black powder, may be used less frequently. The presence of yogurt and juice in the dataset suggests that there may be some health-conscious recipes included.

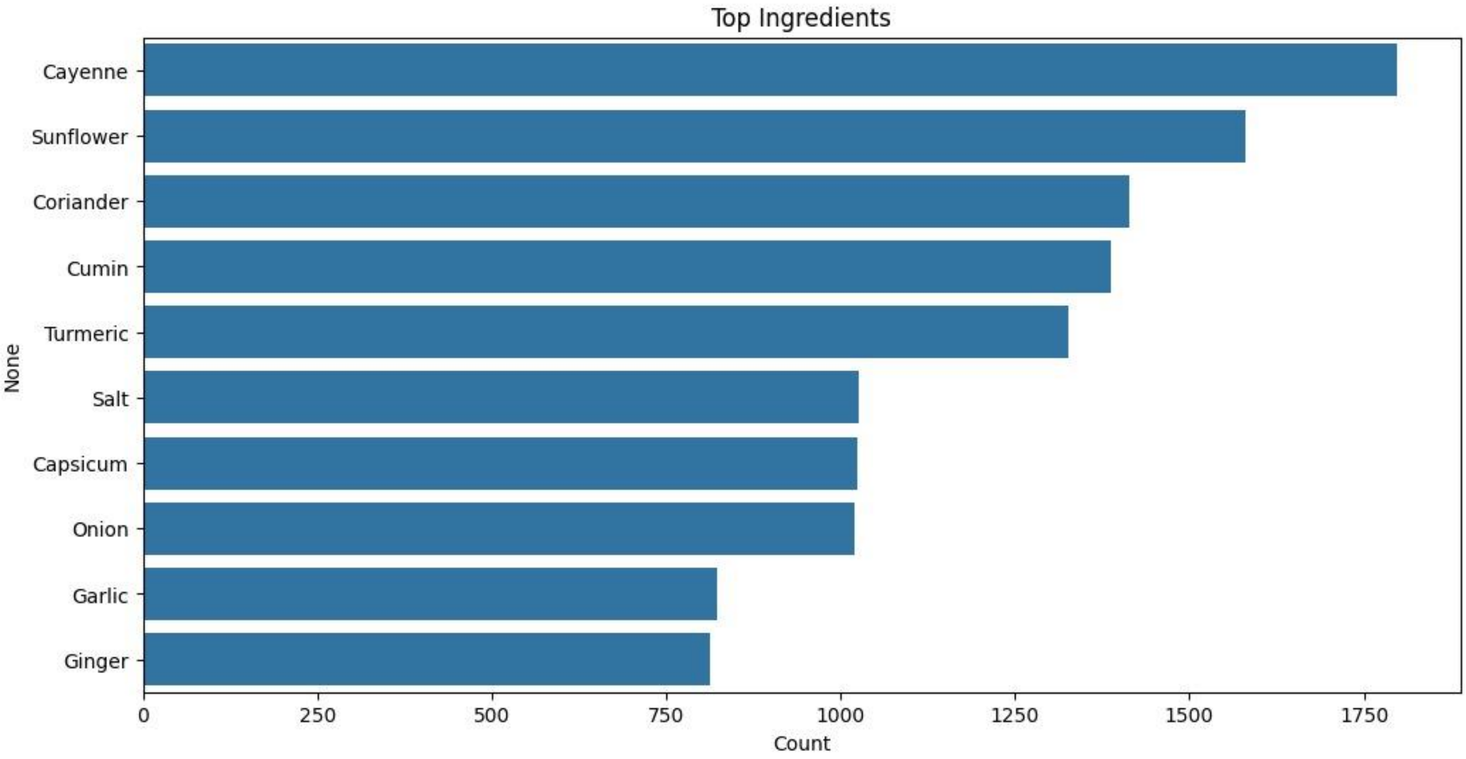


Figure 2 Top 10 ingredients

The column graph titled "Top Ingredients" presents the world of how different ingredients look like inside of the data. Particularly, Cayenne becomes the most frequent ingredient reaching over 1500 times then Sunflower which has the second most of all ingredients and is at 1,430 times. Coriander, cumin, and turmeric are the most occurring spices and they all come at the third place. Meanwhile, our counts for Salt, Capsicum, Onion, and Garlic are seen to be lower than those of Cayenne Pepper and Capsicum, with Garlic and Ginger being the least occurrences among designated ingredients. This graphic portrays to a comprehensive manner, the ingredients, Cayenne being the most abundant one. The information gathered in this way can be very useful with references to flavor profiles and the effective management of food inventories in food businesses. In the graph, the x-axis shows the occurrence of each ingredient facilitating a prompt discernment of ingredient usage frequencies.

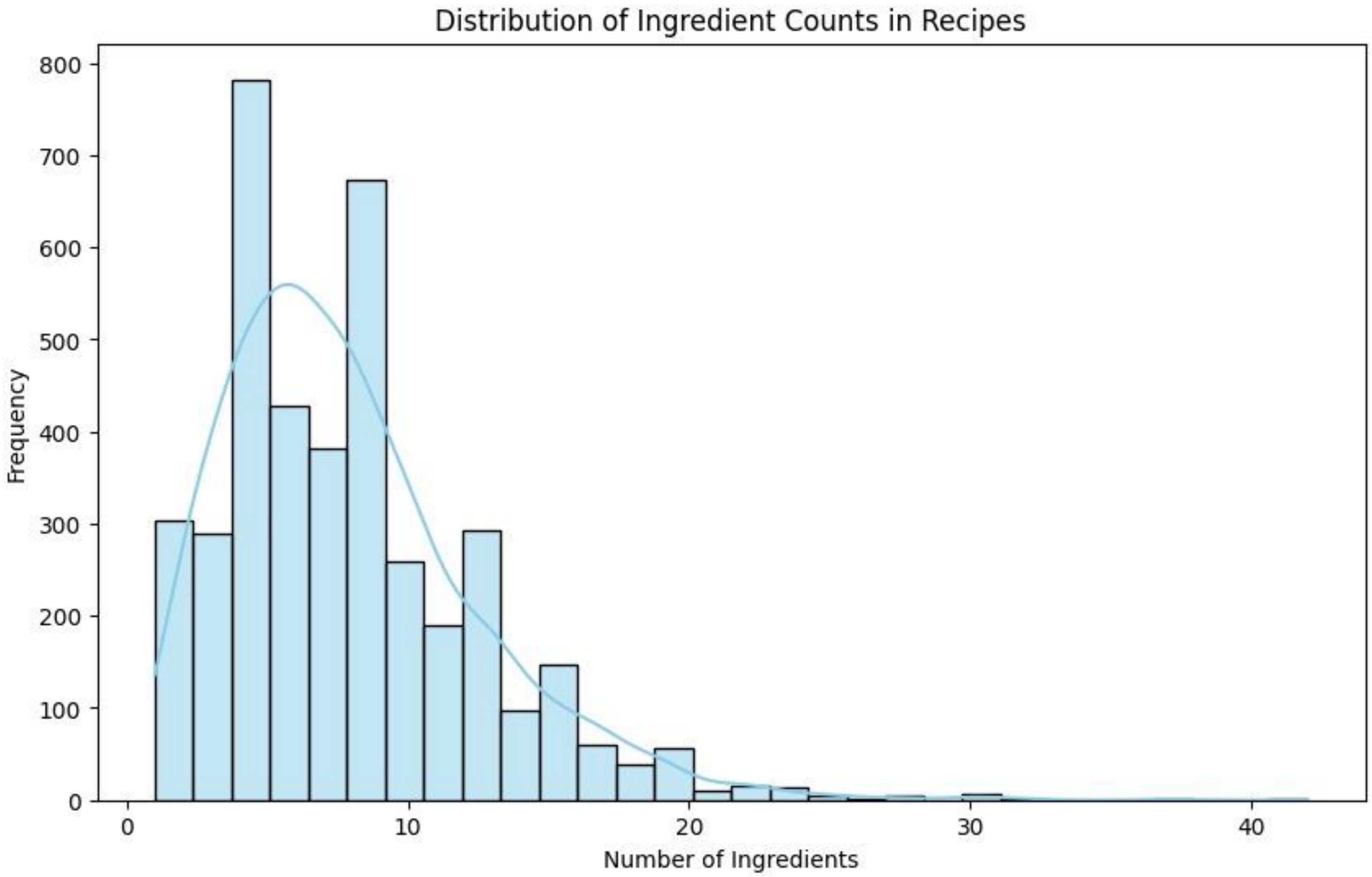


Figure 3 Distribution of Ingredients

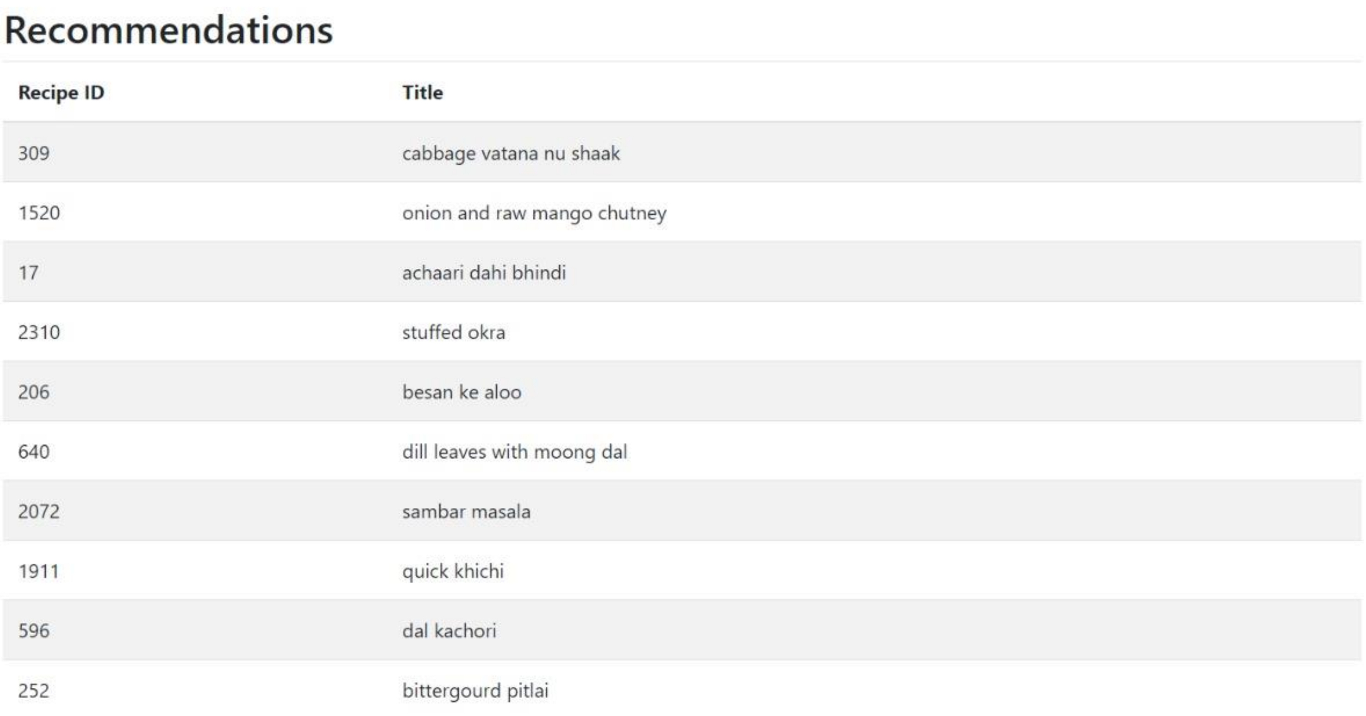
An eye-catching bar graph titled, "The Number of Ingredients in each Recipe", shows how frequently recipes use the ingredients that are required. The horizontal axis shows the numbers of ingredients, starting from 0 and till 40, while the vertical axis represents the frequency of recipes with the lowest value at 0 and the highest value at 800. Whereas having 10 being most frequent ingredients count, we have noted nearly 1000 recipes with this quantity. As shown by above chart the chances stroke down with the increase of ingredients indicating the right-skewed curve. Here, the picture becomes the necessary factor for comprehension of recipes in terms of ingredients number and the search for components quantity patterns.

RECOMMENDATION SYSTEM

Upon finishing data collecting, cleaning, and analysis phases, our team eventually developed the individualized advisory system. Given the existence of our dataset which hold a lot of ingredient features for each recipe, we realized the need to use a dimensionality reduction method on this column (ingredients). This choice was made deliberately to decrease difficulties posed by the fact of dimensionality decrease for our recommendation model performance.

Additionally, we applied the cosine similarity to identify commonalities among recipes in the process of our recommendation refinement. By means of this technique, we could easily modify our recommendations with account for the latent relationships among our data. Using the machine learning library called scikit-learn from Python, we incorporated these complex methods on our recommendation engine for better optimizations.

After spending the initial two weeks of the project with these tasks, now we have a great opportunity to reinforce and refine them to make our work more effective. Dimensionality reduction and cosine similarity being integrated into our work flow, we can boldly proclaim the effectiveness and accuracy of our recommendation system.



Recommendations

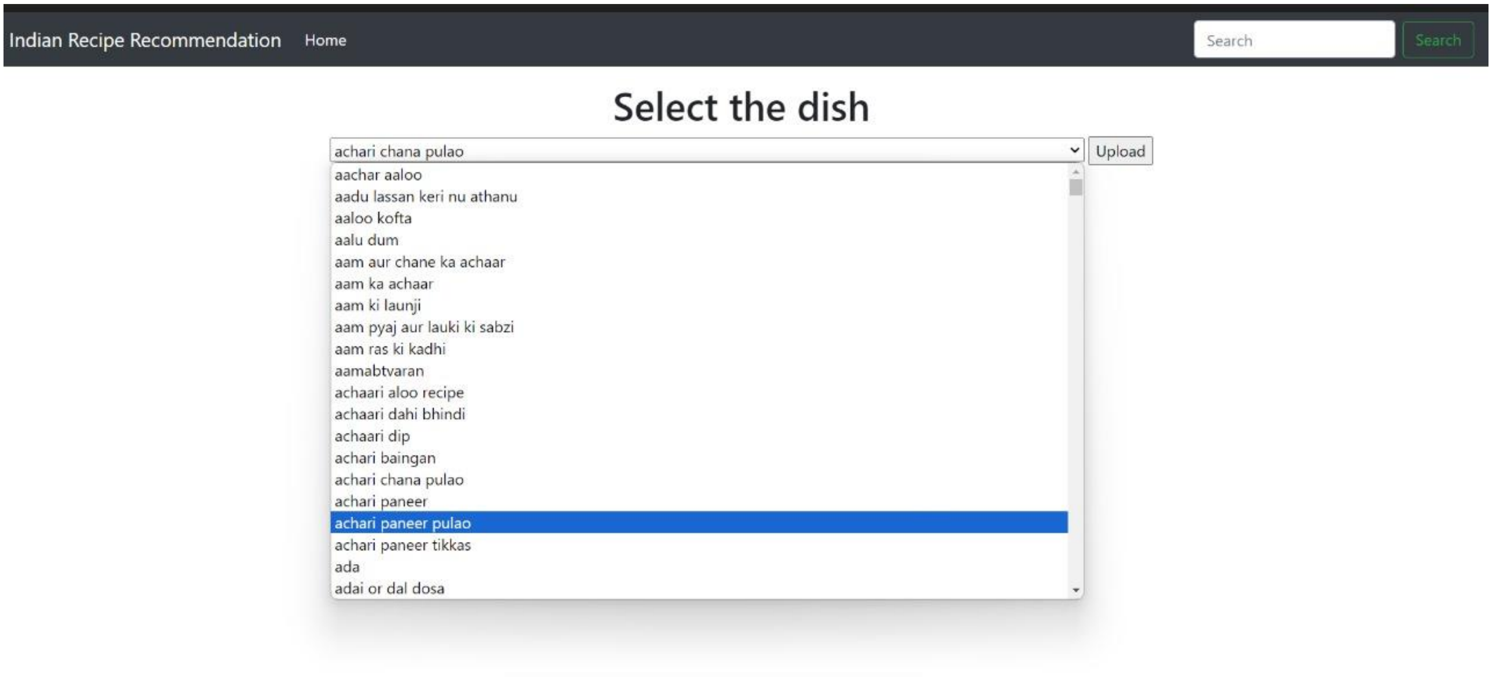
Recipe ID	Title
309	cabbage vatana nu shaak
1520	onion and raw mango chutney
17	achaari dahi bhindi
2310	stuffed okra
206	besan ke aloo
640	dill leaves with moong dal
2072	sambar masala
1911	quick khichi
596	dal kachori
252	bittergourd pitlai

Figure 4 Recommendations based upon the selected recipe

MODIFIED FRONT END IN FLASK

We have developed a frontend user interface for a recommendation system using Flask, a Python microweb framework. The interface prioritizes user experience by focusing on simplicity, responsiveness, and clear presentation of recommendations. Leveraging Flask's integration capabilities, the frontend seamlessly communicates with the backend recommendation system, powered by machine learning algorithms. Users can access personalized recommendations, search for specific items, provide feedback, and customize their preferences. The interface ensures secure user authentication and delivers recommendations in a visually appealing format, fostering engagement and satisfaction.

The front-end design for our recommendation system offers an intuitive and feature-rich experience. By integrating with Flask backend machine learning algorithms, we've created a responsive and personalized interface that enhances and user engagement and satisfaction. Moving forward, we aim to continuously refine and optimize the front end to further improve the effectiveness and usability of our recommendation system.

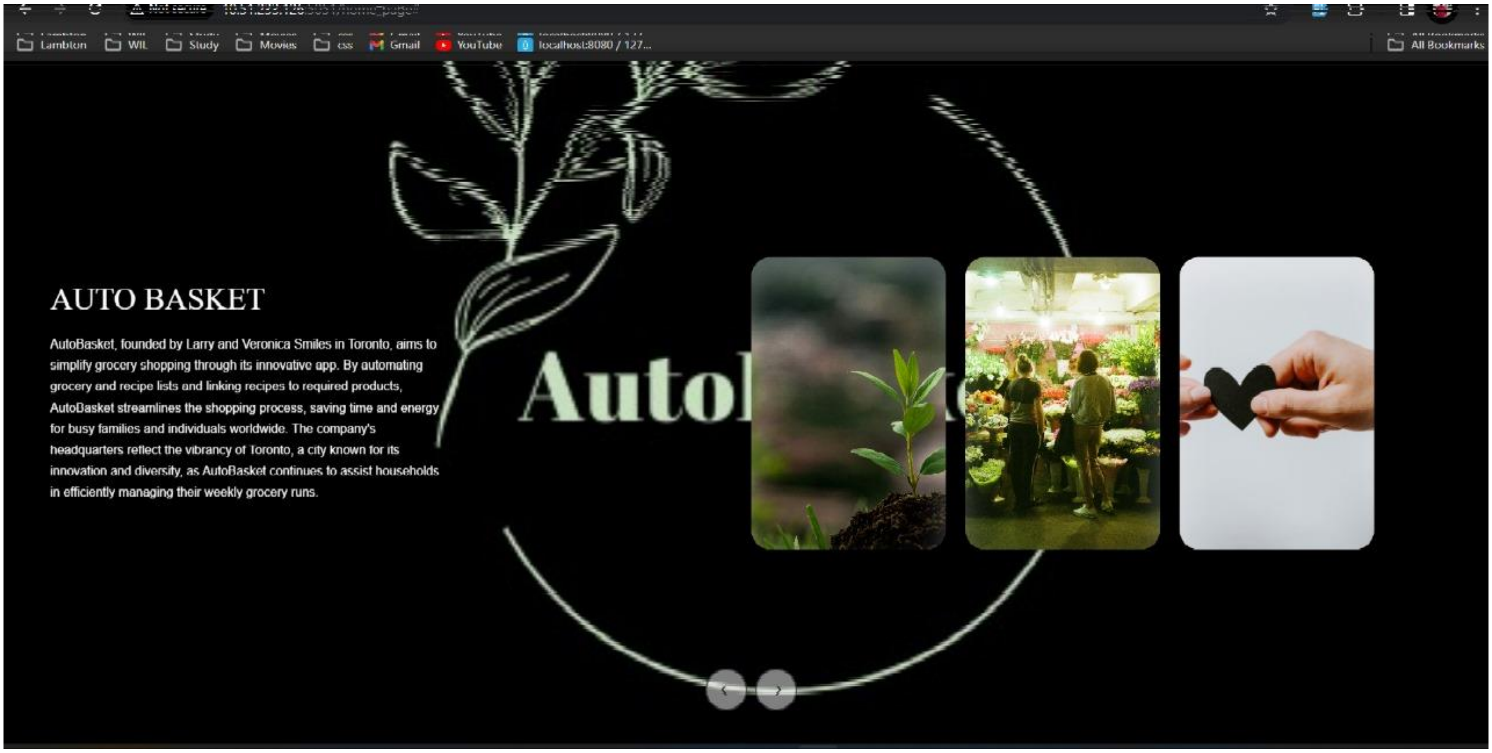


Select the dish

achari chana pulao

aachar aaloo
aadu lassan keri nu athanu
aaloo kofta
aalu dum
aam aur chane ka achaar
aam ka achaar
aam ki launji
aam pyaj aur lauki ki sabzi
aam ras ki kadhi
aamabtvaran
achaari aloo recipe
achaari dahi bhindi
achaari dip
achari baingan
achari chana pulao
achari paneer
achari paneer pulao
achari paneer tikkas
ada
adai or dal dosa

Figure 5 Recommendation system home page with dropdown menu



AUTO BASKET

AutoBasket, founded by Larry and Veronica Smiles in Toronto, aims to simplify grocery shopping through its innovative app. By automating grocery and recipe lists and linking recipes to required products, AutoBasket streamlines the shopping process, saving time and energy for busy families and individuals worldwide. The company's headquarters reflect the vibrancy of Toronto, a city known for its innovation and diversity, as AutoBasket continues to assist households in efficiently managing their weekly grocery runs.

Figure 6 AutoBasket Homepage POC

CONCLUSION

To sum up, the crusade from the data collection to the creation of an Indian meal recommendation system was characterized by the exploitation of abstruse data processing techniques and profound mode of system analysis. Based on the initial information provided by our partner CosyLab, we took extra prettiness efforts to work with the initial data, correcting distortions, checking its dependencies and arranging it for better usability. The past data visualization was a useful tool that showed us percentage of certain products by a supplier and formed the basis of our recommendation system development.

The recommendation system with the dimension reduction and cosine similarity method, spotlighting on similar relationships of recipes, provides more refined and tailored suggestions to users. Through scikit-learn, we used the engine for optimization of strong and accurate results. The last piece of the puzzle was developing a user-friendly GUI powered by Flask. The purpose was to magically talk to the recommender system and display beautiful and engaging personalized recipes via the interface. Concludingly, the undertaken project integrates the data science principles (visualization and U.I design) to output a comprehensive and powerful recommender system suitable for Indian dishes.

REFRENCES

[1] */over-half-of-consumers-place-an-online-grocery-order-once-a-week*. (2024, 01 26). Retrieved from euroshop: https://mag.euroshop.de/en/2021/02/over-half-of-consumers-place-an-online-grocery-order-once-a-week/

[2] *build-a-recipe-recommender-system-using-python*. (2024, 01 22). Retrieved from javatpoint: https://www.javatpoint.com/build-a-recipe-recommender-system-using-python

[3] *candinavia-food-python-recommendation-systems*. (2024, 1 26). Retrieved from duarteocarmo: https://duarteocarmo.com/blog/scandinavia-food-python-recommendation-systems