

Reporting Date: 8th March 2024

TABLE OF CONTENTS

[ABSTRACT 2](#_Toc160751923)

[INTRODUCTION 3](#_Toc160751924)

[DATA VISUALIZATION 5](#_Toc160751925)

[RECOMMENDATION SYSTEM 5](#_Toc160751926)

[SCIKIT LEARN 5](#_Toc160751927)

[PCA 5](#_Toc160751928)

[NEURAL NETWORK 5](#_Toc160751929)

[GENERATING RECOMMENDATIONS 5](#_Toc160751930)

[ANALYSIS OF RECOMMENDATIONS 5](#_Toc160751931)

[MODIFIED FRONT END IN FLASK 5](#_Toc160751932)

[CONCLUSION 5](#_Toc160751933)

[REFRENCES 5](#_Toc160751934)

# ABSTRACT

# INTRODUCTION

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

# 

# RECOMMENDATION SYSTEM

# GENERATING RECOMMENDATIONS

# MODIFIED FRONT END IN FLASK

# CONCLUSION

# REFRENCES

1. (2024, 2 16). Retrieved from faker: https://faker.readthedocs.io/en/master/
2. *synthetic-data-generation-techniques*. (2024, 2 16). Retrieved from turing: https://www.turing.com/kb/synthetic-data-generation-techniques