CONTENT BASED RESTAURANT RECOMMENDATION SYSTEM

Chandan Dey (21UCS088)
Mainak Saha (21UCS128)
Kabir Prakash Pol (21UCS114)
Dhruv Rohatgi (21UCS182)

COMPUTER SCIENCE & ENGINEERING DEPARTMENT
NATIONAL INSTITUTE OF TECHNOLOGY, AGARTALA

INDIA-799046

November, 2024

CONTENT BASED RESTAURANT RECOMMENDATION SYSTEM

Report submitted to
National Institute of Technology, Agartala
for the award of the degree
of
Bachelor of Technology

by

Chandan Dey (21UCS088)
Mainak Saha (21UCS128)
Kabir Prakash Pol (21UCS114)
Dhruv Rohatgi (21UCS182)

Under the Guidance of

Dr. Swapan Debbarma
Associate Professor

COMPUTER SCIENCE & ENGINEERING DEPARTMENT
NATIONAL INSTITUTE OF TECHNOLOGY AGARTALA

November, 2024

Dedicated To

To our Project Supervisor Dr. Swapan Debbarma, Associate Professor, Computer Science Engineering Department, NIT Agartala for guiding us and sharing his valuable knowledge, encouragement and showing confidence on us all the time. To Head of the Department Prof. Mrinal Kanti Debbarma, Professor, Computer Science and Engineering Department, NIT Agartala for encouraging us in our journey. Each of the faculties of the department to contribute in our development as a professional and help us to achieve this goal. To all those people who have somehow contributed to the creation of this project and who have supported us.

"If I have seen further than others, it is by standing upon the shoulders of giants."

- Isaac Newton

REPORT APPROVAL FOR B.TECH

This report entitled "*Content Based Restaurant Recommendation System*", by Chandan Dey, Mainak Saha, Kabir Prakash Pol, Dhruv Rohatgi, is approved in partial fulfilment of the requirements for the award of ***Bachelor of Technology*** in ***Computer Science & Engineering.***

Dr.Swapan Debbarma	Prof. Mrinal Kanti Debbarma
Supervisor	Head of the Department
Associate Professor	Professor
Computer Science and Engineering Department	Computer Science and Engineering Department
NIT, Agartala	NIT Agartala

Date:_______________
Place:NIT, Agartala

DECLARATION

We declare that the work presented in this report titled "*Content Based Restaurant Recommendation System*", submitted to the Computer Science and Engineering Department, National Institute of Technology, Agartala, for the award of the ***Bachelor of Technology*** degree in ***Computer Science & Engineering***, represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

November, 2024
Agartala

Chandan Dey	Mainak Saha
Kabir Prakash Pol	Dhruv Rohatgi

CERTIFICATE

It is certified that the work contained in the report titled "*Content Based Restaurant Recommendation System*", by Chandan Dey, Mainak Saha, Kabir Prakash Pol, Dhruv Rohatgi, has been carried out under my supervision. This work has not been submitted elsewhere for a degree.

Dr.Swapan Debbarma
Supervisor
Computer Science and Engineering Department
NIT Agartala

# Acknowledgement

I would like to acknowledge my college N.I.T. Agartala for providing a holistic environment that nurtures creativity and research-based activities. I express our sincere thanks to Dr. Swapan Debbarma, Associate Professor CSE Department, N.I.T. Agartala, the supervisor of the project for guiding and correcting throughout the process with attention and care. He has frequently suggested creative ideas and guided through the major hurdles that occurred during the duration of the project. I would also thank Prof. Mrinal Kanti Debbarma, Head of the department and all the faculty members without whom this project would be a distant reality. I also extend my heartfelt thanks to my family and friends who supported me.

- Chandan Dey **- Mainak Saha**

- Kabir Pol **- Dhruv Rohatgi**

List of Figures

List of Tables

Abstract

Recommendation systems are used to provide users with information and services. Generally, users are themselves not aware of the features that affect their choices or likeness. The correct deciding feature and proper weightage is important because they affect the success or failure of the application of these recommendation systems. The small details can have a big impact on user engagement. Analyzing all the possible features and services is essential to understand user experiences and enhance service quality.

In traditional recommendation systems, star ratings provide a basic overview, while delving into content-based features offers valuable insights into the reasons behind the ratings. However, manually processing numerous reviews is impractical. Leveraging machine learning algorithms for similarity analysis becomes a pragmatic solution.

This work presents a recommendation system that utilizes NLP techniques to analyze restaurant tags and provide personalized suggestions. Using CountVectorizer, the system converts textual data into a matrix of token counts, while the Porter Stemmer normalizes words to their root forms. Cosine similarity is then computed to quantify the similarity between restaurants based on their tags. The recommend function identifies the most similar restaurants to a given input, enhancing user experience through tailored recommendations. This approach demonstrates the effectiveness of combining NLP with machine learning to improve the accuracy and relevance of restaurant recommendations.

Contents

CHAPTER 1

Introduction

1.1 Overview

With the exponential growth of social media platforms, the amount of data generated has increased tremendously. In a conventional recommendation system, star ratings provide a summary of the product, but analyzing other restaurant features like cuisine and other facilities provides insightful information about the factors that influence the ratings. This vast amount of unstructured and heterogeneous data can be overwhelming for users and make it difficult to find relevant information. This project aims to address this challenge by developing a personalized, fine-grained user preference-oriented framework that utilizes restaurant features and relevant information to provide personalized recommendations.

1.2 Project Objectives and Significance

The main goal of this project is to develop a framework that can efficiently extract relevant details from restaurant survey data and offer tailored recommendations based on user preferences. This framework aims to:

- Predict likelihood of a user liking a restaurant listed in the Yelp dataset [16] based on analyzing the features, price range, distance and other facilities provided by restaurant for a particular customer, utilizing NLP techniques.

- Improve user experience by offering a graphical user interface that requires two inputs from users or customers to predict the top restaurants in a given city for a specific cuisine that the user provides.
- Determine which services the restaurant offers, including price range, bike parking, and business card acceptance.
- Reduce information overload by filtering out irrelevant recommendations

1.3 Scope

Our focus is on developing a framework that carefully analyzes restaurant data and comments on restaurants to provide personalized recommendations based on city and cuisine. We leverage the power of NLP based models.

To achieve this, we will use a dataset named business dataset from yelp containing information about restaurants and their unique selling points which is responsible for their high or low ratings. This dataset will be preprocessed, tokenized, and subjected to NLP (Stemming and Count Vectorizations) using a regression model. The extracted information will then be used to generate personalized recommendations for each user.

1.4 Motivation

Social media has become an integral part of our daily lives, yet it often presents challenges in extracting relevant information and providing personalized recommendations amidst the vast amount of data generated. Our project aims to bridge this gap by developing a framework that empowers users to navigate the social media landscape with ease, offering tailored recommendations and a delightful user experience.

We believe that everyone should have access to a personalized recommendation system that caters to their unique dining preferences, making every restaurant search a breeze.

CHAPTER 2

Literature Review

2.1 Existing Work

Any system that generates personalized recommendations as an output or directs the user in a tailored manner towards interesting or practical things from a wide range of available possibilities is referred to as a recommendation system [5]. Systems that provide recommendations analyze user preferences and object features of objects and suggest products that users might proactively enjoy. Collaborative filtration, content-based, and hybrid systems of recommendation are the three primary categories under which models of recommendations fall. Collaborative filtering suggests learning from past user interactions, whether explicit (like user ratings) or implicit (like browsing history). Most content-based recommendations are based on comparisons between the auxiliary data of products and users. Text, images, and videos are just a few examples of additional information that could be analyzed. Numerous systems have been proposed in the literature, some of which are detailed below:

Using sentiment analysis, Petrusel, Renata, and Sergiu-George [10]developed a model to recommend restaurants based on positive and negative customer feedback. They combined the outputs of a recommendation system with a sentiment analysis system to create an effective and efficient model for recommending items. The recommendation system utilizes the results of the sentiment analysis to suggest restaurants to other customers through a collaborative filtering method.

Sumedh proposed a Yelp food recommendation system based on customer restaurant reviews [12]. This system retrieved both collaborative and content-based features from the Yelp dataset to identify customer and restaurant profiles. The algorithm developed for recommending restaurants takes into account customer preferences and location, employing K-nearest neighbor clustering.

Additionally, R. Yera et al. [15] introduced a food recommendation system focused on nutrition and calories. This research primarily emphasized users' dietary needs rather than the restaurant's renowned cuisine, flavor, or quality.

Rui Maia and Joao et al. [9] conducted a study to recommend restaurants and food based on reviews from previous users in which when a new user moves by a nearby location, the context-aware method is used to identify the restaurants that are close by and it performs collaborative filtering on food items to predict which items are most popular based on collaborative ratings on food items.

Also, Li Chen et al. proposed restaurant recommendation system [6] that was developed based on the opinions of the visited users. This study developed a review-based restaurant recommendation algorithm using collaborative filtering, preference-based product rating, and content-based approaches. When customers give a preferred rating, all restaurants with that rating are presented. This technique is implemented by using text analysis of customer evaluations.

However, existing systems often struggle to simultaneously sort restaurants based on both location and rating. Saito, Asada, Yoshitomi, et al. [11] proposed a hybrid filtering system that recommends recipes based on the previous night's dinner, user-selected impression words (such as sweet, warm, or spicy), past successful recommendations, and recipes with similar ratings.

M. Gupta et al. [7] aimed to discover user interests and personalize recommendations to match the user's mood using collaborative filtering. Their algorithm can be enhanced to identify facial expressions, predict a user's mood, and make food recommendations based on individual preferences.

Khan et al. [8] introduced a restaurant recommendation system that utilizes opinion mining of unstructured customer feedback, including emojis and GIFs. The study's findings were notable as they transformed each unstructured opinion into quantifiable information that can be used for rating restaurants.

While significant research has been conducted and various models have been developed for recommending food and restaurants based on factors such as past orders (content-based filtering), ratings (collaborative filtering), location, and nutrition, there remains a need for a model that effectively combines content-based and collaborative filtering. Current models exhibit gaps, as they often fail to incorporate diverse features during the prediction process.

2.2 Summary of Existing Work

Table 1: Summary of the existing work

S. No.	Research Paper Title	Authors	Remarks
1	Recommender systems servey	J. Bobadilla, F. Ortega, A. Hernando, and A. Gutierrez	Recommendation systems are the systems that provide recommendations, analyze user preferences and object features and suggest products that user might like. It helps in retrieving the recommended data.
2	A restaurants recommendation system: Improving rating predictions using sentiment analysis	M.-R. Petrusel and S.-G. Limboi	Using sentiment analysis, both positive and negative customer feedbacks are analysed and through collaborative filtering recommendations are provided.
3	Yelp food recommendation system	S. Sawant and G. Pai	Both collaborative and content-based features are used for recommendation based on revies employing K-nearest neighbour clustering.

S. No.	Research Paper Title	Authors	Remarks
4	A food recommender system considering nutritional information and user preferences	R. Y. Toledo, A. A. Alzahrani, and L. Martinez	Recommendations are done based on user's dietary needs rather than cusine, flavor or quality.
5	Context-aware food recommendation system	R. Maia and J. C. Ferreira	Context aware method is used for identifying restaurants that are close by and based on collaborative filtering it gives most popular food items based on collaborative ratings.
6	Recommender systems based on user reviews: the state of the art	L. Chen, G. Chen, and F. Wang	Review based recommendation algorithm that uses collaborative filtering, preference-based product rating and context based approaches. It is done by using text analysis of customer evaluations.
7	A recipe decision support system using knowledge information and agent	K. Saito, T. Asada, Y. Yoshitomi, R. Kato, and M. Tabuse	Hybrid system that sorts restaurants based on both location and rating, it recommends recipes based on previous successful recommendations and recipes with similar ratings.

S. No.	Research Paper Title	Authors	Remarks
8	Mood based food recommendation system	M. Gupta, S. Mourila, S. Kotte, and K. B. Chandra	Personalized recommendations based on users mood using collaborative filtering. It can be also used to identify facial expressions, predict mood and make food recommendations based on individual preferences.
9	Mining opinion components from unstructured reviews: A review	K. Khan, B. Baharudin, A. Khan, and A. Ullah	Recommendation system that utilizes opinion mining of unstructured customer feedback, including emojis and GIFs. It transforms unstructured opinion into quantifiable information that can be used for rating.

CHAPTER 3

Methodology

3.1 Dataset Description

The information utilized in this project is derived from the Yelp Data Challenge. The dataset consists of a collection of JSON files that encompass various types of data, including business details, reviews, tips (which are condensed reviews), user information, and check-in records. Each business entry includes key attributes such as its name, address, operating hours, category, average star rating, total number of reviews, and additional characteristics like noise level and service policies. The review entries contain a star rating, the text of the review, the date it was posted, and the total number of votes that the review has received. These two categories of data have been the primary focus of our analysis.

The dataset is organized into six subdatasets, each providing a concise overview of the data it contains. The total size of the dataset is 6.84 GB, which includes the following subfiles:

Business Dataset (139 MB) Check-In Dataset (50.3 MB) Photo Dataset (34.9 MB) Review Dataset (4.39 GB) Tips Dataset (203 MB) User Dataset (2.03 GB) In this project, our main objective is:

Analyzing business attributes, which include the name, address, operating hours, category, average star rating, total number of reviews, and various other features such as noise level and service policies and recommend restaurants on the basis of these features.

3.1.1 Attributes of 'Business.json'

Here's example of how the information presented in a table format:

Table 2: Attributes of "business" dataset [16]

Field	Description	Example
business id		tnhfDv5Il8EaGSXZGiuQGg
name	Business name/restaurant name	Garaje
city, state	Address	4754 rd St, San Francisco, CA, 94106
latitude, longitude	Coordinate	Latitude: 37.7817529521, Longitude: -122.404122
stars	Stars by user	4.5 stars (rounded to half-stars)
review count	Total reviews number	1198 reviews
is-open	Open or closed status	Open (is-open: 1)
attributes	Services offered by business	BusinessParking: Garage: No Street: Yes Validated: No Lot: No Valet: No
categories	categories of business type	-Mexican -Burgers -Gastropub
hours	operating hours of each day of a week	Monday to Friday: 10:00 AM to 9:00 PM Saturday: 10:00 AM to 9:00 PM Sunday: 11:00 AM to 6:00 PM

3.2 Data PreProcessing

3.2.1 Flowchart:

Figure 1 provides a visual representation of the preprocessing steps undertaken in this study. This process is divided into three key stages: Business Dataset Preprocessing, Review Dataset Processing, and the subsequent merging of Business and Review Datasets. Each of these steps plays a crucial role in preparing and organizing the data for further analysis.

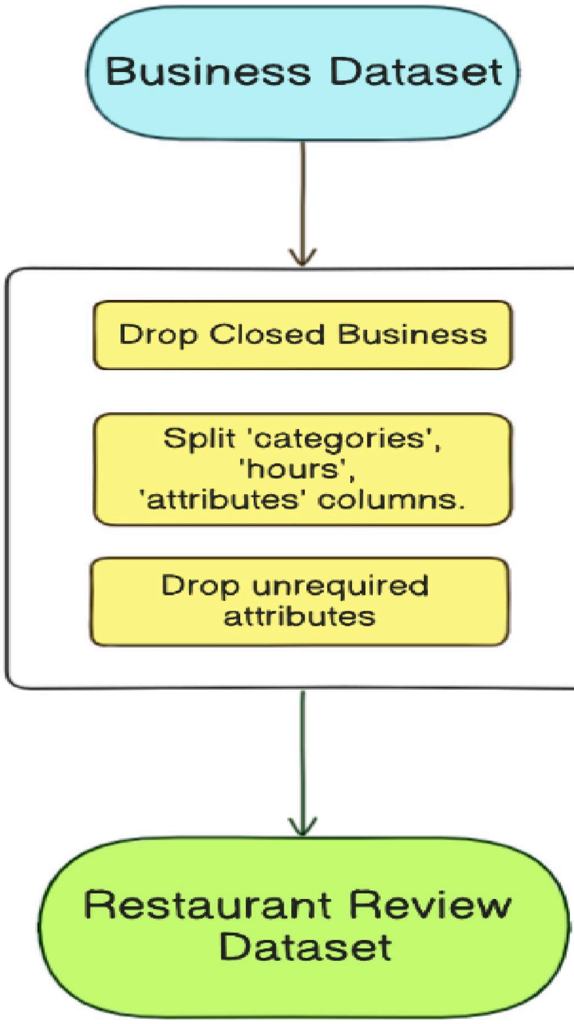

Figure 3.1: Steps of data preprocessing

Here's step-by-step detailed explanation of the process:

3.2.2 Business Dataset Preprocessing:

■ ***Drop Closed Businesses:***

The initial preprocessing step focuses on data relevance by filtering the dataset to include only currently operational businesses. This is achieved by retaining records where the 'is-open' indicator has a value of 1. By excluding closed establishments, we ensure our subsequent analysis captures meaningful patterns from active businesses only. Figure 2 presents a comparative distribution of operational versus non-operational businesses in our dataset. For the purposes of our analysis, we will proceed by removing all closed business entries.

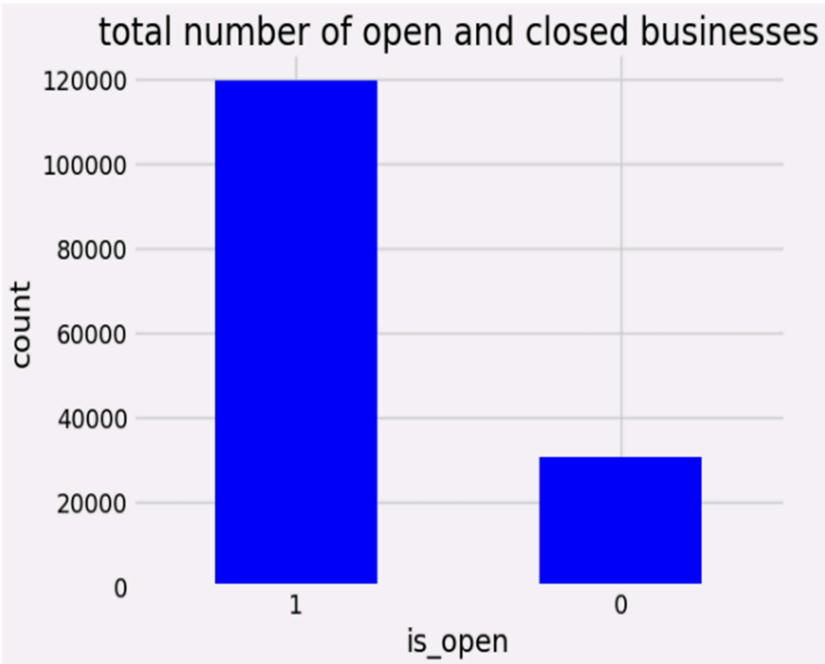

Figure 3.2: Count of open and closed business

■ ***Organize Data for Further Analysis***

The preprocessing likely aims to organize the data for further analysis or visualization. This includes splitting certain columns into separate columns to make the data more structured and easier to work with.

a) Splitting 'categories' Column

The 'categories' column contains various strings representing the business's categories. To make this information more accessible, the column is split into separate columns, using commas as separators. This allows for easier analysis of the different categories associated with each business.

b) Splitting 'hours' Column

The 'hours' column is in nested form, meaning it contains a dictionary within each record. To extract the hour information, the column is split into separate columns, each representing a specific day of the week and its corresponding hours of operation. This makes the hour information more organized and easier to analyze.

c) Splitting 'attributes' Column

The attributes column has values in key value pair format with values being yes in all cases. So for each business we extracted the keys from attributes column and created a new column named features.

d) Dropping Redundant Columns

After splitting the attributes column two redundant columns namely BikeParking and BusinessParking are dropped from the resulting dataFrame. These columns contain overlapping information, and keeping them would introduce unnecessary duplication.

e) Dropping columns that don't contribute

There are some columns that don't contribute in the recommendation process and they are important to remove as it creates the data processing easy.

f) Filtering all the restaurant or food based businesses from the business dataset

The 'business.json' has data of business in various categories. We only need the business that are restaurant or food related but the categories column is a comma separated values datatype so we exploded the column to 10 different columns to filter.

g) Extracting food based businesses

We enlisted all the categories we need and used the 'is in' keyword to extract all the required businesses. This will avoid unnecessary calculation and increase speed of our recommendation system.

3.3 Exploratory Data Analysis

Our analysis primarily focused on two key business characteristics: geographical location (incorporating both state and city data) and business classification. The relationship between these features and customer reviews is visualized in Figure 3.3 , providing insights into regional and category-specific review patterns. This visualization help illuminate the hierarchical structure of business categories and their relative prevalence in the dataset.

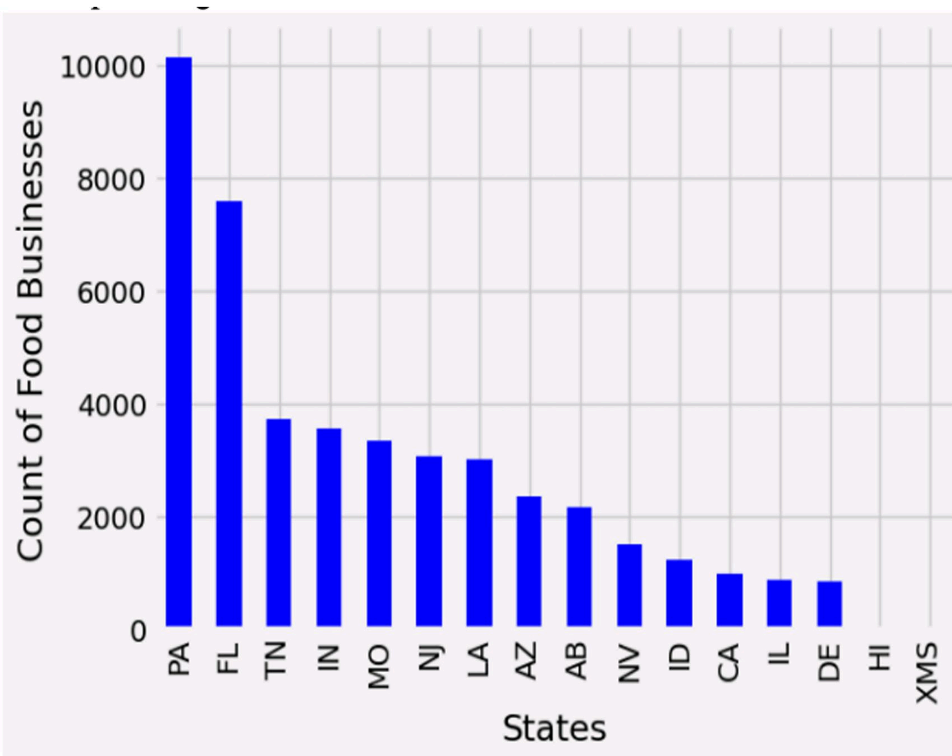

Figure 3.3: Frequency distribution of State v/s Number of food business

3.4 Proposed methodology

3.4.1 Overview

The proposed method integrate results of cosine similarity. Figure 3.4 illustrates the core components and their interactions within the system. The Yelp Restaurant Reviews [16] dataset is

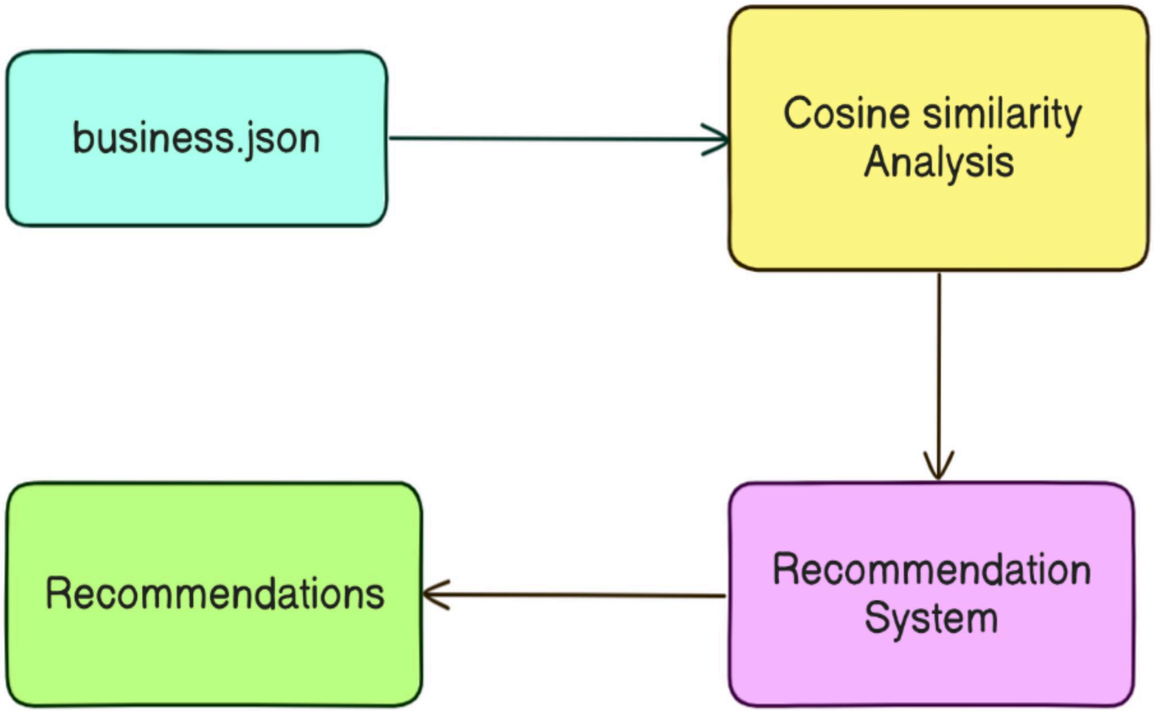

Figure 3.4: Overall System Architecture

utilized as the input for the cosine similarity calculation, which produces a cosine similarity matrix as its output. This score is subsequently passed to the recommendation system to generate a refined and high-quality recommendation list.

3.4.2 Business.json

”Business.json” is our main dataset taken from kaggle[16]. It includes the details of businesses and has attributes like business features, category and location which is crucial for our system. This dataset includes businesses of various categories but during the pre processing we extracted the data of food related businesses. Moreover we filtered and restructured the data into a tokenizable format for applying operations like stemming and vectorization.

3.4.3 Cosine Similarity Analysis

Before putting the dataframe for cosine similarity analysis we first use CountVectorizer to covert textual data into a matrix of token counts on the basis of the ’tags’ column in our dataframe. After that we user Porter Stemmer to normalize the words to their root format to avoid conflict in context and then we do tokenization and lowercasing of the data in the ’tags column’ in our dataframe. So that we can find the similarity values for the restaurants. To define Cosine Similarity in technical terms:

Cosine Similarity is the measure of similarity between two non-zero vectors widely applied in many machine learning and data analysis applications. It actually measures the cosine of the angle between two vectors (in our case the vectors representing two restaurants).

Cosine similarity basically calculates the similarity between two different data objects spread across given data sets. In our project, cosine similarity is basically used to calculate similarity between two restaurants based on user rating data, price range , location, cuisine and the features of the restaurants etc. Each restaurant acts as a vector and cosine angle is calculated among them to aggregate the similarity amongst them. Formula for cosine similarity analysis is given below:-

$$\text{Cosine Similarity (C.S)} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|}$$

where,

- $\mathbf{A} \cdot \mathbf{B}$ = the dot product of the vectors $\mathbf{A}$ and $\mathbf{B}$.
- $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ = the magnitudes (lengths) of the vectors $\mathbf{A}$ and $\mathbf{B}$.
- $\|\mathbf{A}\| \cdot \|\mathbf{B}\|$ = the product of the magnitudes of $\mathbf{A}$ and $\mathbf{B}$.

3.4.4 Recommendation System

A recommendation system is a machine learning algorithm that uses data to suggest items or content to users based on their preferences. Recommendation systems process information about users' online activity, such as their interests, preferences, and purchases. They then use this information to order search results or generate personalized recommendations. They are common in today's digital scene, serving an important role in online shopping, streaming services, social networking, and other platforms where personalization and user experience are critical.

3.4.5 Recommendations

In machine learning, ”recommendations” refers to a system that uses algorithms to suggest items, content, or products to users based on their past behavior, preferences, and data analysis, essentially predicting what a user might like or find interesting, often seen in social media applications, where personalized recommendations are displayed to enhance user engagement and experience.

CHAPTER 4

Experiment Environment

4.1 System Configuration

1. Natural Language Toolkit (NLTK):

A comprehensive Python library for NLP[14] , the NLTK [2]. It provides an extensive array of tools for text processing tasks, such as segmentation by sentence, tokenization, and part-of-speech tagging. These features were employed during the prototyping phase of the NLP[14] pipeline.

Version: 3.8.1

2. Matplotlib:

A versatile Python library for creating static, animated, and interactive visualisations is Matplotlib[1]. It offers a variety of plot types and data visualisation tools, making it possible to visualise model performance metrics and insights obtained from the NLP[14] models.

Version: 3.7.3

3. NumPy:

NumPy is a fundamental Python library for scientific computing, developed by NumPy[3]. Effective numerical operations on arrays, matrices, and other multidimensional data structures are offered by it. It was employed in a number of NLP[14] pipeline functions, such as performance evaluation, model training, and data processing.

Version: 1.23.5

4. Pandas:

Pandas [4] is a robust Python library for data analysis and manipulation. It provides tools and structures for processing tabular data, such as DataFrames, Series, and Index objects. In the NLP [14] experiments, Pandas was utilized to load, clean, and organize the datasets.

Version: 1.5.2

4.2 Experimental Setup

The experimental setup begins with cleaning and organizing the data, merging information from businesses, and reviewing datasets to create a comprehensive dataset. Sentiment scores are then calculated for each review, and these scores are used to predict ratings. The accuracy of these predictions is evaluated by comparing them to actual user ratings. In the final step, personalized restaurant recommendations are generated based on user-provided city information, leveraging sentiment analysis to suggest the top 10 positively reviewed establishments in the specified city. This comprehensive approach aims to offer tailored and positively rated restaurant suggestions aligned with user preferences and location choices.

4.3 Comparison with existing models

To evaluate the performance of the proposed methodology, comparisons were made with existing models and techniques commonly used in Natural Language Processing (NLP) and recommendation systems. The comparison metrics included prediction accuracy, precision, recall, F1-score, and computational efficiency.

4.3.1 Baseline Models

- **Bag of Words:** It is a traditional feature extraction technique in text processing. The Bag of Words model represents text data as a collection of word occurrences. It does not take the grammar and the word order into consideration while processing. Being simplistic, Bag of Words served as a baseline for assessing the effectiveness of modern NLP techniques that are used in this study.

- **TF-IDF (Term Frequency - Inverse Document Frequency):** It is a statistical measure which is used to evaluate the importance of a word in a document relative to a collection/set of documents. TF-IDF provides weighted features that helps this model by making it a stronger baseline for comparison than Bag of Words model.

- **Word2Vec:** It is a neural embedding technique that converts the given words into continuous vector representations by capturing semantic relationships between words. Models using Word2Vec embeddings were compared to assess the improvements in the sentiment analysis accuracy.

4.3.2 Strengths and Limitations of the Proposed Model

Strengths:

- Enhanced sentiment analysis using fine-tuned models.
- Robust integration of user preferences and location data for recommendations.
- High scalability for larger datasets and multiple cities.

Limitations:

- Dependence on the quality and quantity of review data.
- Sensitivity to outliers or mislabeled sentiment data.
- Computational overhead when scaling to very large datasets.

CHAPTER 5

Deployment

5.1 Introduction

This document describes the deployment process for the **Restaurant Recommendation System**, which recommends restaurants based on user preferences. The system consists of:

- **Frontend:** Built using React with Vanilla CSS for styling.
- **Backend:** Developed with Python's Django framework.
- **Machine Learning Model:** Designed using a restaurant dataset from Kaggle, implemented in Jupyter Notebook, and integrated with the backend.

5.2 Deployment Steps

5.2.1 Frontend Deployment (React)

- Build the React application using `npm run build`.
- Use **Netlify** for deployment:
  1. Create a Netlify account and set up a new project.

2. Upload the `build` folder or link the GitHub repository for automated deployment.
3. Configure environment variables, such as the API base URL for backend communication.

- Verify the deployment by accessing the hosted site and testing the UI functionalities.

5.2.2 Backend Deployment (Django)

- Push the Django backend code to a GitHub repository.
- Use **Koyeb** for deployment:
  1. Create a Koyeb account and set up a new service.
  2. Link the GitHub repository to deploy the backend application.
  3. Add necessary environment variables, such as:
     - Secret Key
     - Database Credentials (if using an external database)
     - Path to the ML model
- If the database is external, configure the connection (e.g., PostgreSQL or SQLite).

5.2.3 Machine Learning Integration

- The recommendation model was trained on the Kaggle restaurant dataset using Jupyter Notebook.
- Save the trained model as a serialized file (`.pkl`) using libraries like `joblib` or `pickle`.
- Integrate the ML model into the Django backend by:
  1. Loading the model in the appropriate views or API endpoint.
  2. Designing API endpoints that process frontend requests, analyze user preferences, and fetch recommendations using the ML model.

5.2.4 Integration of Frontend and Backend

- Update the React frontend to communicate with the Django backend using the correct API endpoint provided by Koyeb.
- Ensure CORS is enabled in the Django backend to allow cross-origin requests.
- Test the end-to-end flow by submitting choices in the frontend and verifying the recommendations from the backend.

5.2.5 Security Measures

- Store sensitive keys and credentials as environment variables.
- Use HTTPS for secure data transmission.
- Implement authentication if required for user-specific recommendations.

5.2.6 Testing and Monitoring

- Test the entire application locally and post-deployment to ensure all functionalities work as expected.
- Use tools like **Netlify Analytics** for monitoring frontend performance and Koyeb's monitoring tools for backend health.

5.2.7 Backup and Recovery

- Periodically back up the database to secure cloud storage or a local server.
- Maintain version control for both frontend and backend codebases using Git.

CHAPTER 6

CONCLUSION & FUTURE WORK

6.1 Conclusion

After reading the existing text around our topic we observed that most of the recommendation systems were based on food, reviews and tips dataset by using sentiment analysis and KNN algorithms but none of them addressed the fact that any restaurant distinguishes itself from the others due to it's ambiance and facilities mostly. Today's consumer likes to experience new things. With our Content Based restaurant recommendation system we provided a tool so that they can get recommendations by putting basic information and not worrying about the details. User input on city and cuisine preferences serves as a crucial filtration step, tailoring our analyses to specific geographic locations and culinary tastes, amplifying the relevance of our recommendations and catering to diverse user preferences. Strategic result filtering based on user-specified criteria, combined with top similarity values, elevates our recommendation system's effectiveness. Its role in extracting features from restaurants, paired with user-focused filtering and recommendation strategies, emphasizes our dedication to a tailored and enjoyable dining experience.

6.2 Future Work

As we think about the future, there are several areas worth exploring in content based analysis. Taking a closer look at larger and more diverse datasets could reveal additional insights into how well models can adapt to different languages and subjects. Using multiple criteria to filter information holds promise for improving accuracy and strength. Additionally, continuously refining and adapting recommendation models to changes in language patterns and user behaviors is crucial for ensuring their effectiveness in dynamic real-world situations. We hope that our new path may contribute in creating a just and more accurate models.

References

[1] Matplotlib — visualization with python. https://matplotlib.org/. Retrieved December 1, 2023.

[2] Nltk :: Search. https://www.nltk.org/search.html?q=stopwords. Retrieved December 1, 2023.

[3] Numpy. https://numpy.org/. Retrieved December 1, 2023.

[4] pandas. python data analysis library. https://pandas.pydata.org/. Retrieved December 1, 2023.

[5] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.

[6] L. Chen, G. Chen, and F. Wang. Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 25:99–154, 2015.

[7] M. Gupta, S. Mourila, S. Kotte, and K. B. Chandra. Mood based food recommendation system. In *2021 Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–6. IEEE, 2021.

[8] K. Khan, B. Baharudin, A. Khan, and A. Ullah. Mining opinion components from unstructured reviews: A review. *Journal of King Saud University-Computer and Information Sciences*, 26(3):258–275, 2014.

[9] R. Maia and J. C. Ferreira. Context-aware food recommendation system. In *Proceedings of the Conference on Context-aware Food Recommendation*, pages 349–356, 2018.

[10] M.-R. Petrusel and S.-G. Limboi. A restaurants recommendation system: Improving rating predictions using sentiment analysis. In *2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 190–197. IEEE, 2019.

[11] K. Saito, T. Asada, Y. Yoshitomi, R. Kato, and M. Tabuse. A recipe decision support system using

knowledge information and agent. *Journal of Robotics, Networking and Artificial Life*, 5(3):204–207, 2018.

[12] S. Sawant and G. Pai. Yelp food recommendation system, 2013.

[13] R. Singh and P. Dwivedi. Food recommendation systems based on content-based and collaborative filtering techniques. 10 2023.

[14] C. to Wikimedia projects. Natural language processing. https://en.wikipedia.org/wiki/Natural_language_processing, November 2023. [Online; accessed 2023].

[15] R. Y. Toledo, A. A. Alzahrani, and L. Martinez. A food recommender system considering nutritional information and user preferences. *IEEE Access*, 7:96 695–96 711, 2019.

[16] I. Yelp. Yelp dataset. https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset, 2021. [Online; accessed 2021].

APPENDIX A

Publications related to report work

1. ”Food Recommendation Systems Based On Content-based and Collaborative Filtering Techniques”. 14th ICCCNT IEEE Conference July 6-8, 2023 IIT -Delhi, Delhi, India.[13]

2. S. Sawant and G. Pai, “Yelp food recommendation system,” 2013. [12]

APPENDIX B

Biographical sketch

Chandan Dey

Sarkar Colony Jamshedpur,Jharkhand PIN-831013,
E-Mail: chandandey1357@gmail.com, Contact. No. +91-8298020458

- Pursuing B.Tech. in Computer Sc. & Engg. branch from N.I.T,Agartala with CGPA of 8.61.
- Intermediate from D.A.V Public School N.I.T Campus , Adityapur. Jamshedpur under (C.B.S.E), Jharkhand with 94.8 % in 2020.
- Matriculation from D.A.V Public School N.I.T Campus , Adityapur. Jamshedpur under (C.B.S.E), Jharkhand with 96.0 % in 2018.

Mainak Saha

Gandhighat, Hospital road extension, Agartala, Tripura(W), 799001
E-Mail: mksnitacse25@gmail.com, Contact. No. +91-6033241409

- Pursuing B.Tech. in Computer Sc. & Engg. branch from N.I.T, Agartala with CGPA of 8.41.
- Graduated High School from Bhavan's Tripura Vidya Mandir, Narsingarh under C.B.S.E, Tripura 91.6% in 2020.
- Intermediate from Bhavan's Tripura Vidya Mandir, Narsingarh under C.B.S.E, Tripura 93.2% in 2018.

Kabir Prakash Pol

Samarat Ashok Nagar, Kurduwadi, Maharashtra PIN-413208
E-Mail: polkabirkwd@gmail.com, Contact. No. +91-8446541207

- Pursuing B.Tech. in Computer Sc. & Engg. branch from N.I.T,Agartala with CGPA of 8.82.
- Intermediate from Narayana Junior College, Hyderabad under (T. S. B. I. E.), Telangana with 95.8% in 2021.
- Matriculation from Jawahar Navodaya Vidyalaya, Solapur under (C.B.S.E), Maharshtra with 93.4% in 2019.

Dhruv Rohatgi

D-107, Aparna Prem Apartment, Shastripuram, Agra, Uttar Pradesh PIN:282007
E-Mail: rastogidhruv28@gmail.com, Contact. No. +91-7248650548

- Pursuing B.Tech. in Computer Sc. & Engg. branch from N.I.T,Agartala with CGPA of 9.07.
- Intermediate from Simpkins School, Agra under (C.B.S.E), Uttar Pradesh with 95.4% in 2021.
- Matriculation from Simpkins School, Agra under (C.B.S.E), Uttar Pradesh with 97.2% in 2019.

BTech project

ORIGINALITY REPORT

11%	9%	3%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	github.com Internet Source	3%
2	www.geeksforgeeks.org Internet Source	2%
3	Submitted to Queen Mary and Westfield College Student Paper	1%
4	ceur-ws.org Internet Source	1%
5	www.disneyholidays.com Internet Source	1%
6	Di Wu. "Data Mining with Python - Theory, Application, and Case Studies", CRC Press, 2024 Publication	1%
7	Submitted to The University of Texas at Arlington Student Paper	1%
8	fastercapital.com Internet Source	<1%

9	Ravi. M, Parvesh Ahmed. K. M, Sengottaiyan. P, Karthikeyan. T. "IoT Based Manhole Status Monitoring and Alert System", 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023 Publication	<1%
10	midra.uni-miskolc.hu Internet Source	<1%
11	www.coursehero.com Internet Source	<1%
12	www.mdpi.com Internet Source	<1%
13	Azizian, Sasan. "A Data-Driven Discovery System for Studying Extracellular Micro RNA Sorting and RNA-Protein Interactions", The University of Nebraska - Lincoln, 2024 Publication	<1%
14	datasciencedojo.com Internet Source	<1%
15	towardsdatascience.com Internet Source	<1%
16	1library.net Internet Source	<1%