A

Mini-Project Report on

Restaurant Recommendation System Using Content Based Filtering

Submitted in partial fulfillment of the requirements

for the degree of

BACHELOR OF ENGINEERING

IN

Computer Science & Engineering

Artificial Intelligence & Machine Learning

By

Aman Thakur (22106014)

Mohit Suthar (22106056)

Vaishnavi Sonawane (22106125)

Disha Suryawanshi (22106112)

Under the guidance of

Prof. Ranjita Asati

Department of Computer Science & Engineering

(Artificial Intelligence & Machine Learning)

A.P. Shah Institute of Technology

G.B.Road, Kasarvadavali , Thane(W)-400615

University Of Mumbai

2023-2024

A.P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled "Restaurant recommendation system using content based filtering" is a bonafide work of Aman Thakur (22106014), Mohit Suthar (22106056), Disha Suryawanshi (22106112) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

Prof. Ranjita Asati	Dr. Jaya Gupta
Mini Project Guide	Head of Department

A.P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval

This Mini project report entitled "**Restaurant Recommendation System using Content Based Filtering ”** by **Aman Thakur (22106014), Mohit Suthar (22106056), Vaishnavi Sonawane (22106125), Disha Suryawanshi (22106112),** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Science &Engineering***, (AIML) ***2023-24***.

External Examiner: _____

Internal Examiner: _____

Place: APSIT,
Thane Date:

DECLARATION

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be 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.

Mohit Suthar	Vaishnavi Sonawane	Disha Suryawanshi	Aman Thakur
(22106056)	(22106125)	(22106112)	(22106014)

ABSTRACT

With the increasing popularity of online restaurant platforms, there is a growing demand for effective restaurant recommendation systems to assist users in finding dining options that suit their preferences. Hence, recommender systems or recommendation systems are simple algorithms that aim to provide the most relevant and accurate items (products, movies, events) to the user (customers, visitors, app users) by filtering useful stuff from a huge pool of information base. Recommendation engines discover data patterns in the data set by learning consumers choices and produces the outcomes that co-relates to their needs and interests.

Our project is to develop similar recommendation system related to restaurants using content based filtering. In this study,we propose a content based filtering restaurant recommendation system which is designed to provide users with tailored restaurant suggestions.

Our aim is to build a restaurant recommendation system that provides personalized restaurant recommendations to users. Since different people have different food preferences and dietary restrictions, we perform careful feature selection to take advantage of the information reflected in a user's reviews. we have generated recommendations not just for an individual user, but also for a group of users.

Keyword: Content based Filtering, Personalized Recommendation, Recommendation system.

INDEX

CHAPTER 1
INTRODUCTION

1.INTRODUCTION

In recent years the lifestyle of human beings is almost changed from the traditional way of living into an easier technological advanced mode. A lot of technological innovations and systems are implemented which can assist daily activities and makes life easier. A recommendation system is the result of evolving technologies through times aimed to assist decision making. Pioneer to it, there are traditional programming which is a bit complex and difficult for decision making due to lack of intelligence and no deepest analysis of the data stored. The solution for such kind of problem is the emergence of a recommendation system. The recommendation system deeply analyzes the data, categorize and arrange them in a well-mannered form, apply some algorithms and make fast and precise decisions. As a result, it will save a great amount of time spent to make a decision, cost and energy lost on a specific course of actions.

The use of recommendation systems will often be encountered in marketing systems, such as online applications and online trading. This application is carried out with the aim of seeing potential customers and providing recommendations for a product related to the common interests of users.. Providing recommendations for certain products or goods will have a certain impact, such as increasing the level of popularity because the product or goods will often be seen by users and increasing the chances of the product or item being sold. A Recommendation System is an information filtering system that seeks to predict the rating a user would give for the item (in this case a restaurant). We can break down the large matrix of ratings from users and items into two smaller matrixes of user-feature and item-feature. Recommender Systems or Recommendation Systems are simple algorithms that aim to provide the most relevant and accurate items (products, movies, events, articles, food, restaurants) to the user (customers, visitors, app users, readers) by filtering useful stuff from a huge pool of information base. Recommendation engines discover data patterns in the data set by learning consumers' choices and produces the outcomes that co-relates to their needs and interests.

By selecting user-related data from a wide range of data, recommender systems are simple algorithms that give users recommendations for products relevant to their needs. This system identifies data patterns in the data set by learning user preferences. It provides them with results relevant to their needs and interests. The internet has now become an irreplaceable part of modern life. Nowadays, users are flooded with options, and much information is available for anything from finding a hotel to making intelligent financial decisions. Hence, companies have created recommendation systems to help users deal with the information explosion. Research regarding these has been going on for many decades because they can be applied in different areas, which is being use there. Hence in order to improve the user experience, personalization is a crucial method to facilitate. Recommendation Systems are prevalent in many web domains, including e-commerce or media sites, and have significantly enhanced business and decision-making in various information access systems.

Our project is to develop similar recommendation system related to restaurants using content based filtering. Restaurant Recommendation System is a recommender system for restaurants based on user preferences. When user or customer requests for a specific type of restaurant, the system automatically recommends the restaurants that meet their requirements. Our aim is to build a restaurant recommendation system that provides personalized restaurant recommendations to users.

Content-based filtering is based on the principle of recommending items that are similar to those that a user has liked or interacted with in the past. In the case of restaurant recommendation systems, this entails analyzing attributes such as cuisine type, price range, location, ambiance, and user preferences to generate personalized recommendations. Unlike collaborative filtering methods, which rely on the collective wisdom of user interactions, content-based filtering does not require explicit user-item ratings or feedback. Instead, it focuses on the inherent characteristics of restaurants and users to infer preferences and make recommendations.

Food is not just a necessity of life. The food we eat represents our culture, tradition, and values. The norms and values of a place can be significantly related to varieties of food available there. Eventually, the highest rated hotel is being recommended to the user by the restaurant recommended system. Our application takes the food preference and ratings into consideration to recommend food the users. The application uses content based filtering method to recommend the food to the users. The application takes user ratings for different food items and stores it into the database. The application then recommends food items to the users on the basis of their ratings. In this project, we developed a web app which recommend the restaurant based on the choice of your interest. This is used for the users to predict the suitable and best restaurant as per their tastes.

We also address common challenges encountered in building restaurant recommendation systems, such as the cold-start problem, where new users or restaurants lack sufficient interaction data for accurate recommendations, and sparsity, where the available data may be sparse or incomplete.

Through empirical evaluation using real-world restaurant review datasets, we aim to assess the performance and effectiveness of collaborative filtering in generating personalized restaurant recommendations. Additionally, we explore techniques to enhance recommendation accuracy by incorporating auxiliary features such as restaurant attributes and user demographics. Hence content based filtering makes the recommendation more efficient so that each user can use this application for their easy prediction of restaurant.

CHAPTER 2
LITERATURE SURVEY

2.LITERATURE SURVEY

2.1-HISTORY

The history of restaurant recommendation systems using collaborative filtering can be traced back to the early days of e-commerce and online platforms. Collaborative filtering, as a concept, gained prominence in the late 20th century with the emergence of recommender systems designed to personalize product recommendations for users. While its roots can be found in various domains, including music and movie recommendations, the application of collaborative filtering to restaurant recommendations has evolved over time.

- A personalized recommender system victimization machine learning based mostly sentiment analysis over social information (2016). during this paper, proposes such a social framework and it provides the user with faster and additional relevant information, so avoiding moot information and providing abundant required personalization.

- Analysis of classification models supported cookery prediction exploitation machine learning (2017). The correlation between numerous recipes and their ingredient sets were investigated with the assistance of common classification techniques. The tests were conducted on the dataset compiled from numerous sources additionally the} accuracy of classifiers want to predict the cuisines were also and compared.

- A hybrid recommendation system considering visual info for predicting favorite restaurants (2017). during this paper, particularly investigate the influence of visual info, i.e., photos taken by customers and placed on blogs, on predicting favorite restaurants for any given user. It offers the visual info effectively aids favorite eating place prediction.

- Machine learning primarily based food direction recommendation system (2018). In this, they use similarity techniques of user primarily based approach and introduce fastened size neighbourhood and threshold-based neighbourhood to identical. The performance for the All recipe knowledge set is found to be higher than the simulated dataset since there are additional range of interactions between users and things.

- Recommendation system supported item and user similarity on restaurants directory on-line (2018). It merges the item similarity and user similarity options to come up with recommendations. analysis shows that the advice system supported item similarity yields higher F1-measure price once examination to user similarity.

- Comparing filtering techniques in restaurant recommendation system (2018). This paper aims to predict restaurants. It results have shown that hybrid 6 filtering outperforms content-based filtering using regression model and collaborative filtering using cluster-based technique.

- Recipe Recommendation System Mistreatment Machine Learning Models (2019). they need used 2 machine learning models-vector house model and Word2Vec model to seek out prime ingredient pairs from completely different cuisines and to counsel alternate ingredients. the main focus is on the Indian culinary art. Indian culinary art is incredibly Brobdingnagian and various and therefore it's tough to seek out patterns and generate pairs.

- Sentiment analysis of eating house reviews mistreatment machine learning techniques (2019). This paper chiefly focuses on the implementation of varied classification algorithms and their performance analysis. The simulation results showed that SVM classifier resulted within the highest accuracy of ninety-four.56% for the given dataset.

- Restaurant Recommendation System for User Preference and Services supported Rating and Amenities (2019). the present paper proposes a machine learning algorithm to resolve the difficulty of customized eating house choice. The results reveal that the advised approach yields high accuracy.

- Restaurant Recommendation System Victimization Machine Learning (2020). during this paper, consistent with the characteristics of eating house recommender systems. It proposes a machine learning algorithm to resolve the problems of customized eating house choice relying upon yelp information.

2.2-LITERATURE REVIEW

[1] T. V. Yadalam et.al.(2020) discussed about the methodologies of a recommendation system. While the paper mentions the chances of hybrid recommendation systems being much more efficient, it is yet to be proven. The content-based methodology given by the paper eliminates the cold start problem of collaborative systems, the work of securing the users' data arises at developers end. Content-based also forms a chain of feedback to sentiment analysis of each content recommendation. Each feedback is analyzed, tokenized and normalized and comes to sentiment by following the cleaning process which in turn provides the input to the collaborative approach. It has established communication between two approaches.

[2] V. Rao et.al.(2017) It uses many machine learning algorithms like random forest classifier, support vector machine and multinomial Naïve Bayes classifier. The dataset under consideration was classified using all of the given three algorithms. The output of the paper proved that random forest classifier is the most efficient one.

[3] Kim, SW. et.al.(2019) discusses the three types of approaches which are about sentimental analysis. The first model is a binary sentiment analysis where the system checks for the presence of specific words to determine the sentiment of that sentence. The second model works by vectoring the words and maintaining scores of each word. The scores determine the term frequency and the inverse document frequency of the word. The third model is based on the next word negation strategy. Here, if a negative word is spotted in a statement, the next word is not considered in the TF-IDF execution. The efficiency of the simple TF-IDF model works well and gives accurate results.

[4] Y. Wang, L. Li et.al.(2015) In this paper it focuses on the k-means clustering algorithm. Various clusters which are created based on the load put on the electricity distribution system. This clustering is done using the k-means algorithm. It has been proved that the k-means algorithm works well on these clusters and gives results of great accuracy.

[5] M. Petrusel et.al.(2019) In this, the system used is TF-IDF algorithm. There are six steps in the sentiment analysis process, viz. data collection, manual reviews annotation, data pre-processing, data representation, classification process and sentiment analysis performance. The output of the sentiment analysis is used as input to the recommendation engine wherein the restaurants are sorted on the basis of their reviews and arranged based on their ratings.

[6] J. Zeng et.al.(2015) It uses a hybrid recommendation system to find the best suitable restaurants for the user. The similarity between a restaurant and the user is calculated. This calculation helps find out the restaurants that are suitable for the user. The recommendation is done on the basis of the distance between the user and the restaurant and the preferences of the user. These two influential factors give the best results for restaurant recommendation.

CHAPTER 3
PROBLEM STATEMENT

3.Problem Statement

The problem we're addressing is the difficulty people face in choosing a restaurant from the overwhelming number of options available. Traditional methods like asking friends for recommendations or relying on generic online reviews often don't provide satisfactory guidance because they're not personalized to individual preferences. This lack of personalization can make it challenging for individuals to find restaurants that align with their specific tastes and dietary restrictions. To solve this problem, we aim to develop a restaurant recommendation system using collaborative filtering. Content based filtering is a technique that analyzes patterns in user preferences and behaviour to make personalized recommendations.

By leveraging content based filtering, our system will consider factors such as the types of cuisine you enjoy, your budget preferences, and any dietary restrictions you may have. Through this system, users will be able to receive tailored restaurant suggestions that match their unique preferences and needs. By simplifying the process of finding a restaurant and offering personalized recommendations, we aim to enhance the dining experience for individuals and make it easier for them to make informed decisions when choosing where to eat.

CHAPTER 4
EXPERIMENTAL SETUP

4.1 Hardware Setup

Server/Cloud Infrastructure:

- Utilize cloud services (AWS, GCP, Azure) or dedicated servers with ample computing resources.
- Choose instances with sufficient CPU, RAM, and storage based on the expected workload and dataset size.

Processor (CPU/GPU):

- Select a multi-core CPU or GPU for parallel processing tasks.
- CPUs with high clock speeds and multiple cores are preferable for handling computational tasks efficiently.

Memory (RAM):

- Allocate enough RAM to store and process the dataset efficiently.
- Depending on the dataset size and computational complexity, aim for at least 16GB or more of RAM.

Storage:

- Use SSDs for faster data retrieval and processing.
- Ensure sufficient storage capacity to accommodate the dataset and any intermediate data.

Networking:

- Ensure high-speed internet connectivity for accessing cloud resources or external data sources.
- If deploying on a local network, ensure reliable and high-bandwidth connectivity.

4.2 Software Setup

Programming Language:

Choose a programming language suitable for implementing the recommendation system. Python is commonly used due to its rich libraries for data processing and machine learning.

Database Management System (DBMS):

Set up a DBMS (MySQL, PostgreSQL) to store and manage restaurant data efficiently.

Design database schemas to represent restaurant attributes and user preferences.

APIs and Microservices:

Design RESTful APIs or microservices to handle user requests and serve recommendations.

Implement API endpoints for user interactions, such as search queries and feedback submission.

Recommendation Algorithm Implementation:

Implement content-based filtering algorithms for generating restaurant recommendations.

Develop functions for feature extraction, similarity calculation, and recommendation generation.

User Interface:

Develop a user interface (web or mobile) for users to interact with the recommendation system.

Design an intuitive interface for searching, browsing, and receiving recommendations.

Testing and Evaluation:

Set up testing frameworks to evaluate the performance of the recommendation system.

Implement unit tests, integration tests, and performance tests to ensure reliability and accuracy.

CHAPTER 5

PROPOSED SYSTEM AND IMPLEMENTATION

5.1 Block Diagram

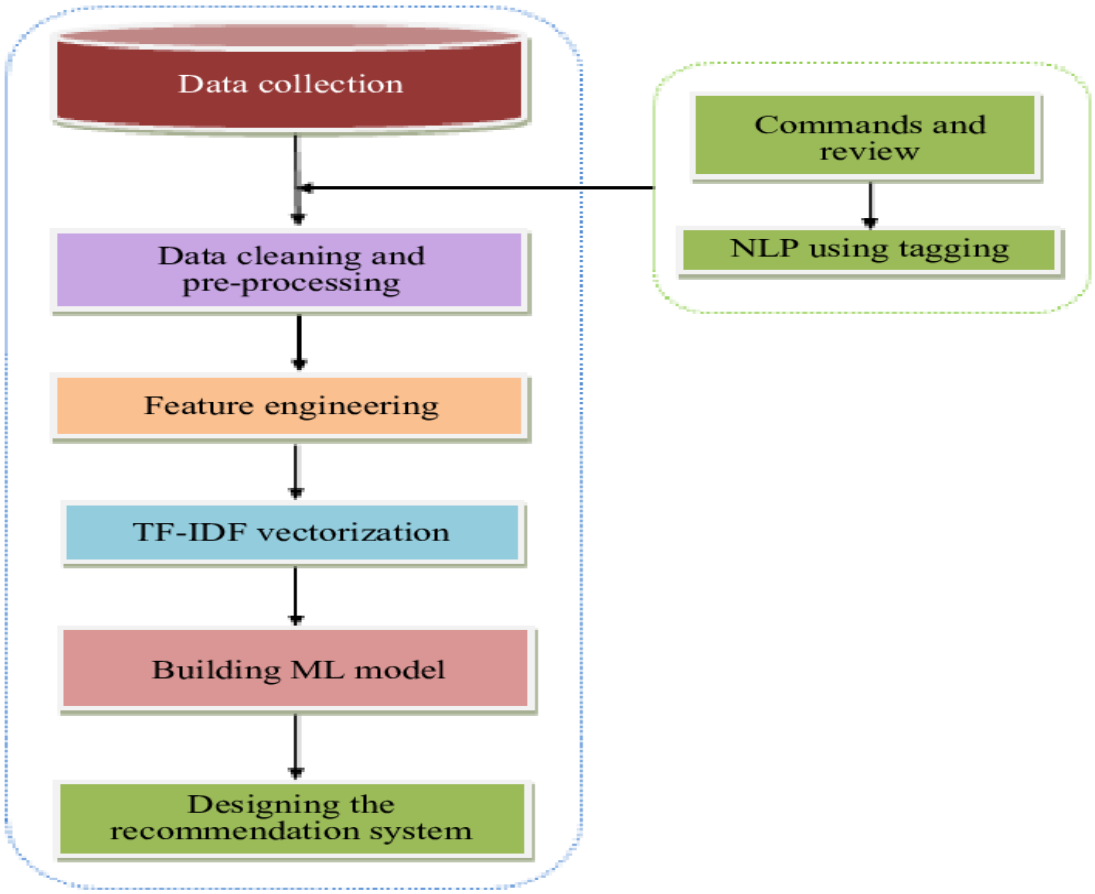

Fig 5.1: Control flow diagram of proposed system

The workflow of the current study is illustrated as follows:

1. Data collection: The dataset employed for this research work has been fetched from Kaggle. The dataset comprises of columns like URL, address, name, rate, phone, location, cuisines, rest_type, dish_liked, menu_item, approx._cost(for two people), etc.

2.Data cleaning and pre-processing: The subsequent stage involves data cleaning and feature engineering, which entails performing various tasks on the data, including:

- Removing unnecessary columns from the dataset.
- Eliminating duplicate entries.
- . Modifying column names to ensure clarity and consistency.
- Conducting data transformations as required.
- Adjusting the column names to align with the desired format.
- Handling any missing values by removing or imputing them.

3.Feature engineering: This refers to the process of analyzing and visualizing data to gain insights, discover patterns, and understand the underlying structure of the dataset. It involves examining the data from various angles, summarizing its main characteristics, and identifying any relationships or trends present in the data. It is typically conducted before performing more advanced statistical modeling or machine learning tasks to represent the underlying information in the data and improve model performance.

4.TF-IDF vectorization: TF-IDF vectorization is a technique used to represent text documents as numerical feature vectors. It is widely employed in NLP and information retrieval tasks. TF-IDF vectorization aims to capture the importance of a term within a document and across a collection of documents. It consists of two main components:

- **TF:** The term frequency of a term (word) within a document is calculated as the number of times the term appears divided by the total number of terms in the document. It reflects the relative importance of the term within the specific document.
- **IDF:** The inverse document frequency of a term is calculated as the logarithm of the total number of documents divided by the number of documents containing the term. It measures how rare or unique a term is across the entire document collection.

The TF-IDF value for a term within a document is obtained by multiplying its TF by its IDF. A higher TF-IDF value indicates that a term is more important within a specific document compared to the overall collection. TF-IDF vectorization helps capture the uniqueness and significance of terms in individual documents while considering their frequency and distribution across the document collection. It is commonly used for tasks such as text classification, information retrieval, text clustering, and document similarity analysis.

5.Building ML model: Building the ML model mainly consists of choosing appropriate ML algorithms or models based on the problem type (e.g., classification, regression, clustering) and the characteristics of the data. Consider factors such as model complexity, interpretability, and scalability. The chosen models are then trained with the training data by optimizing their parameters, and finally, their performance is evaluated through various evaluation metrics.

6.Designing the recommendation system: The suggested hybrid recommendation system combines CBF and item-to-item CF techniques. It analyzes user-written comments to extract preferences and recommends restaurants based on the similarity between their menus and the user's extracted preferences. Additionally, by considering the user's past food choices, the system suggests restaurants with similar menus.

Cosine Similarity:

Cosine similarity is used in content based filtering to measures the cosine of the angle between two vectors, representing the similarity between their directions or orientations in a high-dimensional space.

Cosine similarity is typically applied in recommendation systems for the following reasons:

- **Representing user preferences and item features:** In content-based recommendation systems, user preferences and item features are often represented as high-dimensional vectors. Each dimension corresponds to a specific attribute or characteristic (e.g., genre, price, location) that describes the items. Similarly, user preferences are represented as vectors, where each dimension represents the user's preference for a particular attribute.
- **Calculating cosine similarity:** To determine the similarity between a user's preferences and an item's features, the cosine similarity is computed. This involves taking the dot product of the user preference vector and the item feature vector and dividing it by the product of their magnitudes.
- **Ranking and recommendations:** Once the cosine similarity values are calculated between user preferences and item features, the items are ranked based on their similarity scores. Higher cosine similarity values indicate greater similarity between the user's preferences and the item's features. The top-ranked items are then recommended to the user.

5.2 Implementation

CHAPTER 6
CONCLUSION

6.CONCLUSION

In this project, we have developed a web app which recommend the restaurant based on the choice of your interest. This is used for the users to predict the suitable and best restaurant as per their tastes. The content based filtering makes the recommendation more efficient so that each user can use this application for their easy prediction of restaurant. In this project, we have implemented a model that could recommend the restaurant based on the choice of interest. This application is used for the users to predict the suitable restaurant and find out which dish is famous in region wise and in person. This application ensures the availability of ratings to the customers. The popularity based filtering makes the recommendation more efficient so that each user can use this application for their easy prediction of restaurants been widely applied in making recommendation systems.

REFERENCES

[1] T. V. Yadalam, V. M. Gowda, V. S. Kumar, D. Girish and N. M., "Career Recommendation Systems using Content based Filtering," 2020 5th International Conference on Communication and Electronics Systems (ICCES), COIMBATORE, India, 2020, pp. 660-665, doi: 10.1109/ICCES48766.2020.9137992.

[2] V. Rao and J. Sachdev, "A machine learning approach to classify news articles based on location," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp. 863-867, doi: 10.1109/ISS1.2017.8389300.

[3] Kim, SW., Gil, JM. Research paper classification systems based on TF-IDF and LDA schemes. Hum. Cent. Comput. Inf. Sci. 9, 30 (2019). https://doi.org/10.1186/s13673-019-0192-7

[4] Y. Wang, L. Li and Q. Yang, "Application of clustering technique to electricity customer classification for load forecasting," 2015 IEEE International Conference on Information and Automation, Lijiang, 2015, pp. 1425-1430, doi: 10.1109/ICInfA.2015.7279510.

[5] M. Petrusel and S. Limboi, "A Restaurants Recommendation System: Improving Rating Predictions Using Sentiment Analysis," 2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 2019, pp. 190-197, doi: 10.1109/SYNASC49474.2019.00034.

[6] J. Zeng, F. Li, H. Liu, J. Wen and S. Hirokawa, "A Restaurant Recommender System Based on User Preference and Location in Mobile Environment," 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Kumamoto, 2016, pp. 55-60, doi: 10.1109/IIAIAAI.2016.126.