Samrat Ashok Technological Institute Vidisha (M.P.)
(A Grant-In-Aid Autonomous Engineering College)

RESTAURANT RECOMMENDATION SYSTEM USING KNN CLASSIFIER BASED ON CONTENT-BASED FILTERING

ABSTRACT

In the burgeoning domain of recommendation systems, this research endeavours to develop a robust restaurant recommendation system tailored to individual user preferences. Leveraging data preprocessing techniques, the system effectively filters and cleanses input data, ensuring relevance and accuracy. Upon prompting the user to select a location of interest, the system dynamically curates a dataset specific to the chosen city, subsequently refining it by eliminating extraneous columns. The categorical attributes are meticulously extracted and presented to the user for selection, facilitating a personalized filtering mechanism. Further enriching the user experience, the system incorporates pricing considerations, prompting the user to input a preferred price range for dining. Leveraging this input, along with past user preferences and location-based information, the system applies a K-Nearest Neighbors (KNN) classifier to generate refined restaurant recommendations. This model, trained on historical data, effectively discerns user preferences, and offers tailored suggestions, enhancing user satisfaction and choice convenience.

INTRODUCTION AND HISTORY

In the digital age, navigating the plethora of dining options available can be overwhelming for consumers seeking personalized experiences. Recommender systems have emerged as a solution to this challenge, offering tailored suggestions based on individual preferences. Within this landscape, content-based filtering stands out as a powerful technique for delivering personalized restaurant recommendations.

This paper introduces a novel content-based filtering approach to restaurant recommendation, leveraging a K-Nearest Neighbors (KNN) model. Unlike traditional collaborative filtering methods, which rely on user similarities, content-based filtering focuses on the attributes of the items themselves. In the context of restaurant recommendations, this means considering factors such as cuisine type, restaurant ratings, location, and price range. By harnessing the power of content-based filtering, the proposed system aims to provide users with highly relevant dining suggestions that align with their unique preferences. Through the utilization of a KNN model, the system can effectively identify restaurants that closely match the user's desired criteria, resulting in a more personalized and satisfying dining experience.

METHODOLOGY

A. K-nearest neighbours (KNN)

K-nearest neighbors (KNN) is a simple, instance-based learning algorithm used for classification and regression. In the context of classification, KNN is a non-parametric method that classifies a new data point based on the majority class among its K nearest neighbors. Here's how the KNN classifier works[10]:

Training Phase: In the training phase, the algorithm simply memorizes the feature vectors and corresponding class labels of the training data. No explicit model is built.

Prediction Phase: When a new data point is to be classified, the algorithm calculates the distance between the new data point and all the training data points. The most common distance metrics used are Euclidean distance, Manhattan distance, or Minkowski distance.

Finding Neighbors: The algorithm then selects the K nearest neighbors (data points with the smallest distances) to the new data point.

Majority Voting: For classification, the algorithm assigns the class label that is most common among the K nearest neighbors to the new data point. This is typically done using majority voting.

Regression: For regression, the algorithm assigns the average of the K nearest neighbors' target values to the new data point.

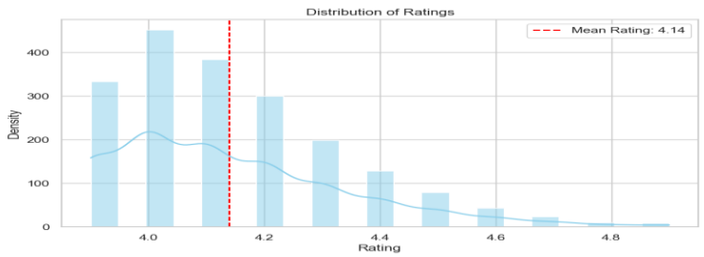
Choosing K: The value of K is a hyperparameter that needs to be specified before training the model. The choice of K can significantly impact the performance of the algorithm. A smaller K value leads to more complex decision boundaries, while a boundaries.

B. Content - based Filtering

Content-based filtering (CDF) recommends products based on their similarities, following the principle of "Show me more of what I have liked". It involves creating item profiles by extracting features from items and user profiles from the features of items purchased by users. Similarity scores between user profiles and item profiles are then calculated to recommend items with the highest similarity scores. This method is commonly used to recommend documents like news, websites, movies, and books based on keywords from profiles. It offers personalized recommendations and transparency in its workings. However, it may struggle with recommending items similar to those already purchased by users and could face challenges in generating attributes for items in certain areas

DATA COLLECTION & EDA

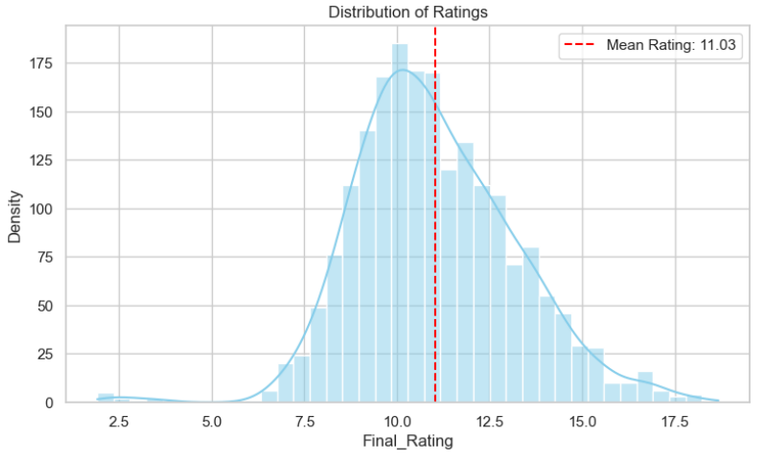
The research uses Zomato Dataset available from the website of Kaggle . The dataset contains information on 1965 restaurants in Delhi NCR as listed on Zomato. The list of restaurants covers the entire NCR Region, last updated on August 30th 2021.
This Data is present in the form of csv file.



Exploratory Data Analysis (EDA)
Exploratory data analysis was conducted to gain insights into the distribution of key variables, such as dining ratings and final ratings. Visualizations, including histograms and bar plots, were utilized to summarize the characteristics of the dataset and identify any trends or patterns.

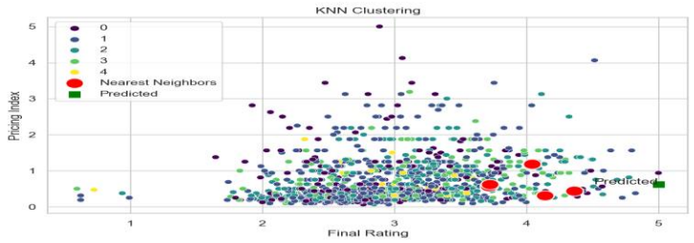
DATA CLEANING & FEATURE ENGINEERING

To prepare the dataset for analysis, several preprocessing steps were performed. Missing values were handled appropriately, and irrelevant columns, such as delivery-related information and geographic coordinates, were removed. Additionally, a new feature called **Final_rating** was engineered based on the dining rating and review count to capture the overall popularity of each restaurant.



RESULT

A. Recommendation and Classification
Finally we got top 5 recommendations from all the available options based on the nearest neighbors.



This is the visual representation how the neighbors are predicted based on the input provided by the user.

CONCLUSION

In conclusion, this paper has presented a novel approach to personalized restaurant recommendations by leveraging content-based filtering and KNN modeling with distance consideration. By integrating key attributes such as cuisine type, restaurant ratings, location, and price range, along with distance from a hypothetical user preference point, the proposed system offers tailored dining suggestions that closely align with individual preferences.

FUTURE WORK

Moving forward, there are several avenues for further exploration and refinement of the personalized restaurant recommendation system. One potential area of focus is enhancing user interaction to gather real-time feedback on recommended restaurants. By implementing interactive features, users can provide valuable insights that can improve the accuracy and relevance of future recommendations. Another aspect that warrants attention is the development of a more dynamic pricing index. By considering factors such as time of day, day of the week, and seasonal variations, the system can offer more precise pricing recommendations tailored to the user's preferences and budget constraints. Furthermore, incorporating contextual information such as weather conditions, special events, or user mood could further refine restaurant recommendations. By leveraging contextual data, the system can adapt recommendations to better suit the user's current situation or preferences. Exploring alternative machine learning algorithms beyond KNN modeling is also worth considering. Neural networks, decision trees, and other techniques may offer advantages in terms of recommendation accuracy and efficiency, warranting further investigation.

REFERENCES

[1] R. J. Mooney and L. Roy. “Content-Based book recommendation using learning for text categorization”. In *Proc. Fifth ACM conference on digital libraries*, 2010, pp. 195-204

[2] S. Sawant, and G. Pai. “Yelp Food Recommendation System.” [ONLINE] Available:http://cs229.stanford.edu/proj2013/SawantPaiYelpFoodRecommendationSystem.pdf

[3] Zomato Restaurants in Delhi NCR
[ONLINE]
Availablhttps://www.kaggle.com/datasets/aestheteaman01/zomato-restaurants-in-delhi-ncr/data |