Project Report: MapMingle

Dynamic Nearby Place Recommendations System

Team Members:
- Rohit Pandey (Captain) (NetID - rp23)
- Prasad Gole (NetID - gole2)
- Asmita Chihnara (NetID - asmitac2)
- Avinash Pathak (NetID - apathak3)
- Surya Sindhu Mallimanugula (NetID - ssm13)

Executive Summary:

MapMingle, a dynamic, user-friendly recommender system for nearby places, has successfully achieved its development goals. The project journey involved comprehensive phases, including data gathering, system design, and establishing core functionalities, followed by refining algorithms, integrating backend and frontend components, and extensive testing. This documentation highlights our journey from inception to completion, showcasing the robust and innovative features of MapMingle, and our commitment to enhancing user experience through technology.

Introduction:

MapMingle was developed to revolutionize real-time user experiences by providing personalized suggestions based on location and preferences. The project combined advanced spatial filtering and ranking algorithms to deliver tailored recommendations.

Addressing a Need

The inception of MapMingle was driven by the growing demand for personalized and dynamic location-based services. This project aims to provide a solution that is both convenient and customizable, meeting the modern requirements of users.

Enhancing User Experience

Traditional location-based services often offer static and irrelevant suggestions. MapMingle revolutionizes this approach by delivering real-time, tailored recommendations, thereby enhancing user engagement and satisfaction.

Leveraging Technology

Utilizing advancements in spatial filtering, ranking algorithms, and mobile technology, MapMingle represents a step forward in creating an intelligent and user-friendly recommender system.

Filling the Gap

Despite the abundance of location-based services, there exists a gap in the market for a system that intelligently combines user preferences, location data, and real-time adaptability. MapMingle aims to bridge this gap.

Impact and Innovation
MapMingle has the potential to significantly impact how individuals interact with their surroundings, making the discovery of new places more intuitive, personalized, and enjoyable. It represents an opportunity for our team to apply cutting-edge technologies in data filtering and ranking algorithms to address real-world challenges.

Completed Tasks:

Ideation, Data Gathering, and Initial Filtering

- Defined project scope and goals.
- Acquired and utilized a comprehensive Yelp dataset.
- Implemented spatial filtering algorithms.

Ranking Algorithm Development

- Developed a ranking algorithm using the BM25Okapi model, focusing on relevance to user queries.

Backend and Frontend Development

- Established a robust backend framework.
- Created an intuitive and interactive frontend user interface.

Integration and Testing

- Successfully integrated frontend and backend components.
- Conducted extensive testing to ensure system reliability and efficiency.
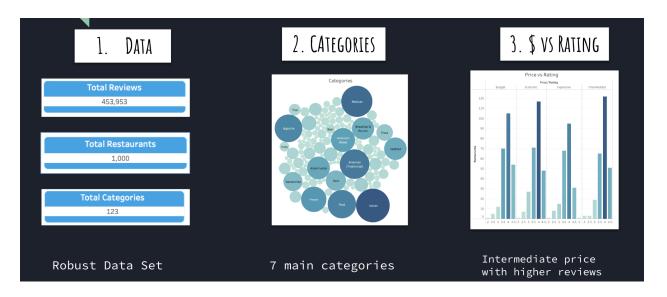
Dataset and Ranking Algorithm

Yelp Academic Dataset

MapMingle utilizes the Yelp Academic Dataset, which is a comprehensive collection of data from Yelp, a well-known user-generated review platform. This dataset is particularly valuable for location-based services and recommendation systems because of its depth and breadth. Key components of the Yelp dataset include:

- Business Information: It contains detailed information about local businesses, including their location, categories, and various attributes.
- Reviews and Ratings: The dataset includes millions of user-generated reviews and ratings, providing deep insights into customer preferences and experiences.
- User Data: Information about the users who submit reviews and ratings is also part of the dataset, offering a window into demographics and user behaviors.
- Check-in Information: This aspect of the dataset covers data about user check-ins at various businesses, which can indicate popularity and customer visit patterns.

This dataset forms the backbone of MapMingle's recommendation system, enabling a sophisticated understanding of both user preferences and business characteristics.

Ranking Algorithm - BM25

For ranking and recommendation purposes, MapMingle employs the BM25 (Best Matching 25) ranking algorithm. BM25 is a well-regarded probabilistic information retrieval model, known for its effectiveness in ranking documents based on their relevance to a search query. Here's how BM25 is utilized in MapMingle:

- Relevance Scoring: BM25 computes a relevance score for each business in the dataset, based on the extent to which it matches the user's search query. This scoring considers the frequency of query terms in the business descriptions and categories.
- Term Frequency-Document Frequency: The algorithm factors in both the frequency of the query term in each document (term frequency) and the number of documents containing the term (document frequency). This helps in balancing the significance of common versus rare terms.
- Tuning Parameters: BM25 comes with tuning parameters like 'k1' and 'b', which can be adjusted to modify the algorithm's sensitivity to term frequency and document length. This allows for customization to suit specific needs and characteristics of the dataset.
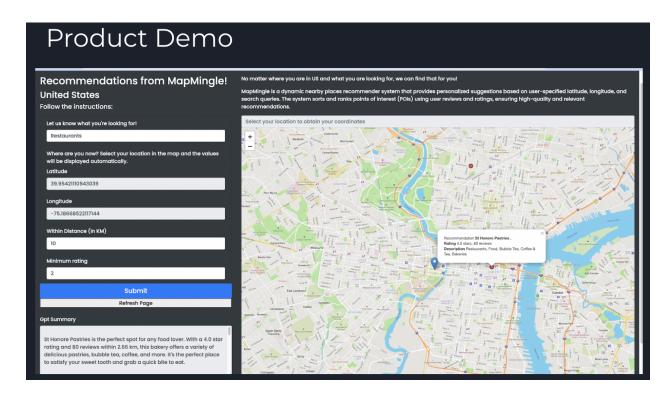
By integrating the BM25 algorithm, MapMingle can efficiently rank businesses based on user preferences and search inputs, ensuring that the recommendations are relevant and personalized.

The combination of the Yelp dataset and the BM25 ranking algorithm provides a solid foundation for MapMingle to deliver personalized, relevant, and real-time recommendations to users.

Sample result:

```
1 # Example query to test
2 test_query = "coffee tea the of"
3 test_results = bm25_query(test_query)
4
5 # Print the results in the specified format
6 for result in test_results:
7     print(f"Place Name: {result['Place Name']}")
8     print(f"Stars and Review counts: {result['Stars and Review counts']}")
9     print(f"Distance from user's lat and longitude: {result['Distance']}")
10    print(f"Categories: {result['Categories']}")
11    print(f"BM25 Score: {result['BM25 Score']:.2f}\n")
```

```
Place Name: Starbucks
Stars and Review counts: 4.5 stars, 7 reviews
Distance from user's lat and longitude: 9.19 km
Categories: Coffee & Tea, Food
BM25 Score: 7.06

Place Name: Starbucks
Stars and Review counts: 4.0 stars, 7 reviews
Distance from user's lat and longitude: 5.96 km
Categories: Food, Coffee & Tea
BM25 Score: 7.06

Place Name: Double Shots Espresso Bar
Stars and Review counts: 3.5 stars, 46 reviews
Distance from user's lat and longitude: 9.71 km
Categories: Coffee & Tea, Food
BM25 Score: 7.06

Place Name: Cosi
Stars and Review counts: 3.5 stars, 5 reviews
Distance from user's lat and longitude: 9.66 km
Categories: Food, Coffee & Tea
BM25 Score: 7.06

Place Name: The Lola Bean
Stars and Review counts: 4.5 stars, 46 reviews
Distance from user's lat and longitude: 9.86 km
Categories: Food, Coffee & Tea
BM25 Score: 7.06

Place Name: ReAnimator Coffee
Stars and Review counts: 4.0 stars, 105 reviews
Distance from user's lat and longitude: 9.66 km
Categories: Food, Coffee & Tea
BM25 Score: 7.06

Place Name: Starbucks
Stars and Review counts: 3.5 stars, 20 reviews
Distance from user's lat and longitude: 1.47 km
Categories: Food, Coffee & Tea
BM25 Score: 7.06

Place Name: Starbucks
Stars and Review counts: 3.5 stars, 63 reviews
Distance from user's lat and longitude: 9.86 km
Categories: Coffee & Tea, Food
BM25 Score: 7.06

Place Name: Dunkin'
Stars and Review counts: 1.5 stars, 6 reviews
Distance from user's lat and longitude: 2.75 km
Categories: Food, Donuts, Coffee & Tea
BM25 Score: 6.57

Place Name: OCF Coffee House
Stars and Review counts: 3.5 stars, 50 reviews
Distance from user's lat and longitude: 9.85 km
Categories: Coffee & Tea, Bagels, Food
```

Product Demo:

Challenges and Mitigation Strategies

We faced challenges getting a good dataset which also had latitude and longitude coordinates. We settled on a dataset from Kaggle which was contributed by Yelp.

We have encountered challenges in data filtering accuracy and frontend-backend connectivity. We are implementing a more robust filtering algorithm and considering the use of additional tools for backend-frontend communication.

We faced challenges converging on a ranking algorithm. We tested with BERT, DistilBERT, etc but they turned out to be very slow. Finally, we opted for BM25 because of its speed and accuracy.

How to Run

1. Clone the code from https://github.com/Prasad-py/CS410_NLP_Ninjas.git and navigate to src folder

2. copy the dataset in the "data" folder. the expected file name is:
yelp_academic_dataset_business.json

3. run -- pip install pandas numpy nltk geopy geojson rank_bm25 matplotlib seaborn scikit-learn -- to install relevant packages

4. run -- pip install flask. -- to host the site

5. once this is done run -- python app.py

6. you can access the site at --http://127.0.0.1:5000

Conclusion

MapMingle has successfully achieved its project goals, with impressive progress and milestones reached throughout its development timeline. The team, having adeptly overcome various challenges, is proud to deliver a system that significantly enhances user engagement and satisfaction. This accomplishment reflects the team's dedication and skill in bringing the project to a successful conclusion.

Enjoy exploring with MapMingle!