Project Progress 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:

The MapMingle project is currently at the halfway mark in its development timeline. Our aim to deliver a dynamic, user-friendly recommender system for nearby places is well on its way to realization. We have successfully completed the initial phases involving data gathering, system design, and the establishment of core functionalities. The subsequent phases, which involve the refinement of algorithms, backend, and frontend integration, and comprehensive testing, are in progress. This report outlines our accomplishments to date and delineates the roadmap for the remaining tasks.

Introduction:

MapMingle is envisioned as a breakthrough in personalizing user experiences in real-time. We have committed to a user-centered design approach, focusing on delivering tailored suggestions based on user location and preferences. The project's first half saw the laying down of a robust foundation, including data acquisition, system architecture design, and the initiation of the development phase.

Completed Tasks:

Ideation and Data Gathering

- Scope and Goals Definition: Our team has established a clear project scope and defined measurable goals.
- Data Source Acquisition: We explored different datasets which also had latitude and longitude coordinates. We finally settled on a comprehensive Yelp dataset that serves as our primary data source.

Initial Data Filtering

- Spatial Filtering Research: We have researched applicable spatial filtering algorithms.
- Algorithm Implementation: A basic version of the spatial filtering algorithm has been implemented.

Preliminary Ranking Algorithm

- Ranking Criteria Research: We have laid the groundwork for the ranking algorithm by selecting appropriate ranking criteria.

- One of the key components of MapMingle is its ranking algorithm. We are planning to use the BM25Okapi model from the Python library *'rank_bm25'*. This model will be employed to rank places of interest based on the relevance of their categories to the user's query. The algorithm begins by loading and preprocessing data from the dataset, including cleaning and tokenizing text in the 'categories' column.

- To tailor recommendations based on location, the algorithm includes a function to filter businesses within a specified radius from the user's location. This is achieved using the *'geopy'* library to calculate distances between coordinates. After filtering, the BM25Okapi model is applied to the tokenized data to score and rank places of interest based on query relevance.

Sample result:

```
# Example query to test
test_query = "coffee tea the of"
test_results = bm25_query(test_query)

# Print the results in the specified format
for result in test_results:
    print(f"Place Name: {result['Place Name']}")
    print(f"Stars and Review counts: {result['Stars and Review counts']}")
    print(f"Distance from user's lat and longitude: {result['Distance']}")
    print(f"Categories: {result['Categories']}")
    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
```

Tasks in Progress

1. **Backend Framework Setup**
2. **Data Filtering Optimization:** We are currently refining our spatial filtering algorithm to improve performance.
3. **Ranking Algorithm Development**: The implementation of a sophisticated ranking algorithm is underway.
4. **Frontend Development**: The development of the frontend UI components is ongoing and post that the integration process between the backend services and frontend components will begin.

Plan for Remaining Tasks

1. **Backend and Frontend Integration**
   - We plan to fully integrate the frontend with the backend, allowing for real-time data exchange and user interaction.

2. **Advanced Data Filtering**
   - The spatial filtering algorithm will be optimized for better accuracy and efficiency.

3. **Completion of Ranking Algorithm**
   - The ranking algorithm will be completed to ensure the relevance of the recommendations.

4. **Final Documentation and Reporting**
   - Complete documentation of the system and project outcomes will be prepared.

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.

Conclusion

MapMingle is on track to meet its project goals with significant progress made in the first half of the development timeline. The team remains committed to overcoming the challenges faced and is confident in delivering a system that will enhance user engagement and satisfaction.

Next Steps

In the coming weeks, our focus will be on completing the integration of system components and launching into a full testing phase to ensure the reliability and efficiency of MapMingle.