TRAIL RECOMMENDER

CS 5800 FINAL PROJECT

WUDASSIE WALLE, BRIGGS TWITCHELL, MATTHEW QUAGLIA, FABIAN GAZIANO


INTRODUCTION

INTRODUCTION

- This project aims to develop a trail recommender system for hikers considering factors such as:
  - trail popularity, difficulty level, elevation gain,
  - trail features(dog and children friendliness, scenic views)
  - activities (birding, camping, hiking, etc.)to provide personalized and efficient trail recommendations for hikers.
- Design a trail recommendation system that employs a sorting algorithm and distance metric to suggest trails that align with the user's preferences and restrictions.

QUESTION

How can we develop an effective trail recommender system that provides personalized recommendations for hiking trails based on user preference and interests?

- *What are the relevant data sources that we can leverage to build a trail recommender system?*
- *How can we preprocess and analyze this data to extract relevant features and insights that will inform the development of the recommendation engine?*
- *What algorithms and techniques can we use to build an accurate and effective recommendation engine?*
- *How can we continually improve the accuracy and effectiveness of the recommendation engine over time?*

PROJECT MOTIVATION

Wudassie Walle

- Provide a valuable resource for fellow parents seeking safe and enjoyable hiking trails for young children
- Foster a love for the outdoors and promote healthy development in young children

Matthew Quaglia

- Easing the stress I, and many other hikers, face in determining what trail to check out next
- Help avid-hikers find new areas to explore with less time wasted doing the research
- New hikers will hopefully feel a little less intimidated about choosing the right places to go for their skill levels with a system like this

PROJECT MOTIVATION

Briggs Twitchell

- Professional perspective: Applying recommendation algorithms to a new domain
- Personal perspective: Creating tools that make outdoor exploration more seamless & enjoyable


ANALYSIS

DATA COLLECTION & PREPARATION

DataSet Source & Data Format

- *Dataset Source:* Kaggle's National Parks Dataset
- *Data Format:* CSV
- Data processing with Python and Pandas

DATA COLLECTION & PREPARATION

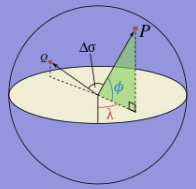
Preprocessing Steps:

1. Converting Lists- Pandas is used to convert strings containing lists into actual lists for easier data manipulation
2. Binarizing Categorical Data- MultiLabelBinarizer and OneHotEncoder from scikit-learn used to transform categorical data into binary format
3. Handling Duplicates and Null Values- Pandas is used to drop duplicate rows and null values for clean and reliable data
4. Adjusting Average Ratings - Accounted for bias in average ratings by adjusting based on the number of reviews

STEPS

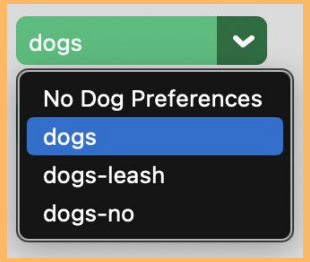
1) READ IN TRAIL DATABASE
2) FILTER TRAILS BASED ON DISTANCE
3) TAKE IN USER TRAIL PREFERENCES
4) CALCULATE DEGREE OF SIMILARITY FOR EACH TRAIL
5) SORT THE TRAILS BY SIMILARITY

GeoPy



$$S_{\text{Gower}}(x_i, x_j) = \frac{\sum_{k=1}^{p} s_{ijk}\delta_{ijk}}{\sum_{k=1}^{p} \delta_{ijk}}$$

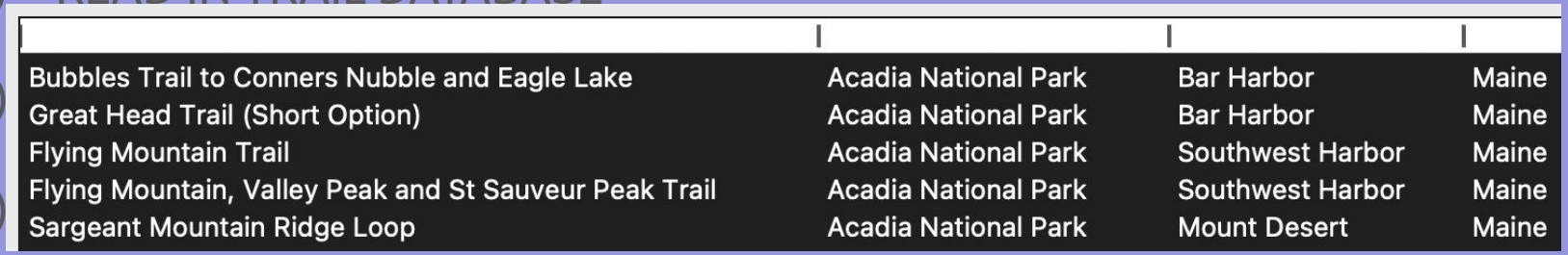

3
Any Difficulty
1
2
3
4
5


dogs
No Dog Preferences
dogs
dogs-leash
dogs-no


Trail Preferences
Features
Activities
beach
cave
city-walk
forest
historic-site
hot-springs
lake
partially-paved
paved
rails-trails
river
views
waterfall
wild-flowers
wildlife
Search
Trail Preferences
Features
Activities
cross-country-skiing
mountain-biking
trail-running
snowboarding
road-biking
fishing
walking
whitewater-kayaking
scenic-driving
ice-climbing
backpacking
canoeing
birding
hiking
horseback-riding
paddle-sports
nature-trips
bike-touring
camping
surfing
skiing
off-road-driving
rock-climbing
sea-kayaking
fly-fishing
Search

STEPS

1) READ IN TRAIL DATABASE



Bubbles Trail to Conners Nubble and Eagle Lake	Acadia National Park	Bar Harbor	Maine
Great Head Trail (Short Option)	Acadia National Park	Bar Harbor	Maine
Flying Mountain Trail	Acadia National Park	Southwest Harbor	Maine
Flying Mountain, Valley Peak and St Sauveur Peak Trail	Acadia National Park	Southwest Harbor	Maine
Sargeant Mountain Ridge Loop	Acadia National Park	Mount Desert	Maine

2)

3)

4) CALCULATE DEGREE OF SIMILARITY FOR EACH TRAIL

5) SORT THE TRAILS BY SIMILARITY

COURSE MATERIALS REFERENCED

Merge Sort

- Merge Sort is a popular divide-and-conquer algorithm for sorting a list or an array.
- It follows the divide-and-conquer paradigm by recursively dividing the input list into smaller sublists, sorting them, and then merging them back to obtain the final sorted output.
- It has a time complexity of O(n log n), making it an efficient sorting algorithm for large datasets.

COURSE MATERIALS REFERENCED

Algorithm Steps:

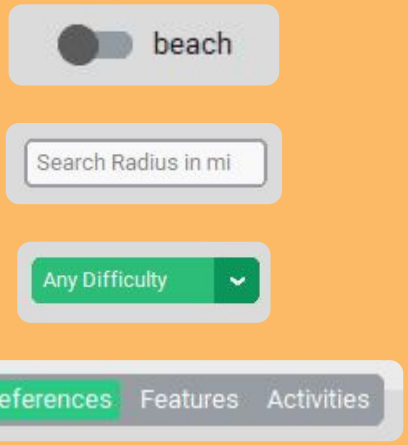
1. Divide:
    - The input list is divided into two equal halves.
    - This step is performed recursively until the sublists contain only one element each.
2. Conquer:
    - The sublists are sorted using the same Merge Sort algorithm recursively.
3. Merge:
    - The sorted sublists are merged back together to obtain the final sorted output.
    - The merging process involves comparing elements from the two sublists and placing them in the correct order in the merged output list.


CONCLUSION

ANSWER TO QUESTION

- Created recommendation system using merge sort and Gower distance
  - User-friendly interface
    - Straight-forward and efficient
    - Tkinter and Customer Tkinter

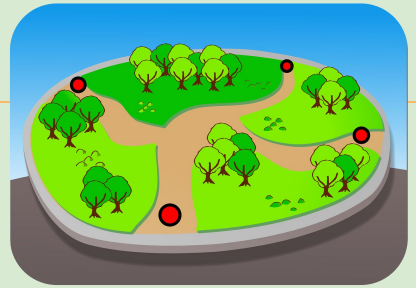
Trail Name	Area	City	State
Eben's Head Trail	Acadia National Park	Isle Au Haut	Maine
Median Ridge and Nat Merchant Trails	Acadia National Park	Isle Au Haut	Maine
Birch Spring Trail	Acadia National Park	Mount Desert	Maine
Western Head via Goat Trail	Acadia National Park	Isle Au Haut	Maine
Jordan Cliffs, South Bubble and Pemetic Mountain	Acadia National Park	Seal Harbor	Maine
Valley Trail and Canada Cliff Loop	Acadia National Park	Southwest Harbor	Maine
Bubble Divide, Sargent East Cliffs, Penobscot Moun	Acadia National Park	Mount Desert	Maine
South Face Trail	Acadia National Park	Seal Cove	Maine
Pretty Marsh Trail	Acadia National Park	Mount Desert	Maine
Redfield Hill Loop	Acadia National Park	Seal Harbor	Maine
Bowditch and Long Pond Trails via Duck Harbor Tra	Acadia National Park	Isle Au Haut	Maine
Ledge Trail	Acadia National Park	Mount Desert	Maine
Mansel, Bernard and Beach Mountain Loop	Acadia National Park	Southwest Harbor	Maine
Sargeant Mountain Ridge Loop	Acadia National Park	Mount Desert	Maine
Great Long Pond Trail from Great Notch	Acadia National Park	Mount Desert	Maine
Beech Mountain West Ridge and South Ridge Trails	Acadia National Park	Southwest Harbor	Maine
Duck Harbor and Deep Cove Trail	Acadia National Park	Isle Au Haut	Maine
Spring Trail and Amphitheater Trail	Acadia National Park	Mount Desert	Maine



WEAKNESSES & LIMITATIONS

- Data only pulled from one dataset affiliated with National Parks
- Data set includes a limited number of columns and features
- Time-constraint

FURTHER RESEARCH



- Implement datasets with a wider variety of trails
- Create system to identify dataset-tidying needs and organize and clean as needed
- Consider user feedback
  - Additional features to include in system
  - Inaccuracies in data

MEMBER TAKEAWAYS

BRIGGS TWITCHELL

- Great teamwork
- Python practice
- Real-world implementation of Gower distance

MATTHEW QUAGLIA

- Implementing sorting algorithms in real-world context
- Gained knowledge of UI
- Teamwork and collaboration

WUDASSIE WALLE

- Collaboration and teamwork
- Project management
- gain hands-on experience in algorithm implementation.

FABIAN GAZIANO

- Python UI experience
- Enjoyable group experience
- User input to Algorithm to result workflow.

TRAIL RECOMMENDER

CS 5800 FINAL PROJECT

WUDASSIE WALLE, BRIGGS TWITCHELL, MATTHEW QUAGLIA, FABIAN GAZIANO