

APMA E4990 Final Project: GoHiking NY State

Find the best trail in NY State

Dongxu Zheng, Shijun Zhu

Columbia University in the City of New York

March 21, 2018

Project Goal

- Build up a blog website
- Sort all trails in NY state by area, difficulty level, condition, length, etc.
- Help users find the best trails and campgrounds depending on their needs
- Evaluate the trails and campground



Breakneck Ridge

Data Source & Data Cleansing

- Upload all datasets: NYS state park trails, NYS snowmobile trails, Hudson riverside green land trails, campgrounds in NYS
- Clean the data, select target features

| Unit | Facility | Name | Asset | Surface | Condition | Height | Shape_Leng |
|--------------|------------------|--------------------|--------------|--------------|-----------|--------|-------------|
| Finger Lakes | Fair Haven Beach | | Paved Road | Soil/Natural | Good | 10 | 95.06367418 |
| Finger Lakes | Fair Haven Beach | Bluff Nature Trail | Marked Trail | Soil/Natural | Good | 10 | 161.2065871 |
| Finger Lakes | Fair Haven Beach | Bluff Nature Trail | Marked Trail | Soil/Natural | Good | 10 | 386.1119986 |
| Finger Lakes | Fair Haven Beach | Bluff Nature Trail | Marked Trail | Asphalt | Good | 10 | 29.20668957 |
| Finger Lakes | Fair Haven Beach | Bluff Nature Trail | Marked Trail | Soil/Natural | Good | 10 | 176.3286674 |
| Finger Lakes | Fair Haven Beach | Bluff Nature Trail | Marked Trail | Soil/Natural | Good | 10 | 46.15569908 |
| Finger Lakes | Fair Haven Beach | Bluff Nature Trail | Marked Trail | Soil/Natural | Good | 10 | 29.04865253 |
| Finger Lakes | Fair Haven Beach | Bluff Nature Trail | Marked Trail | Soil/Natural | Good | 10 | 358.8059011 |
| Finger Lakes | Fair Haven Beach | Lakeshore Trail | Marked Trail | Asphalt | Good | 10 | 570.38944 |
| Niagara | Evangola | | Marked Trail | Soil/Natural | Good | 10 | 150.6590958 |
| Finger Lakes | Fair Haven Beach | Woodland Trail | Unpaved Road | Soil/Natural | Good | 10 | 855.7001262 |
| Finger Lakes | Fair Haven Beach | Lakeshore Trail | Marked Trail | Soil/Natural | Good | 10 | 994.0468921 |

NY state trails datasets come from data.gov: www.data.gov

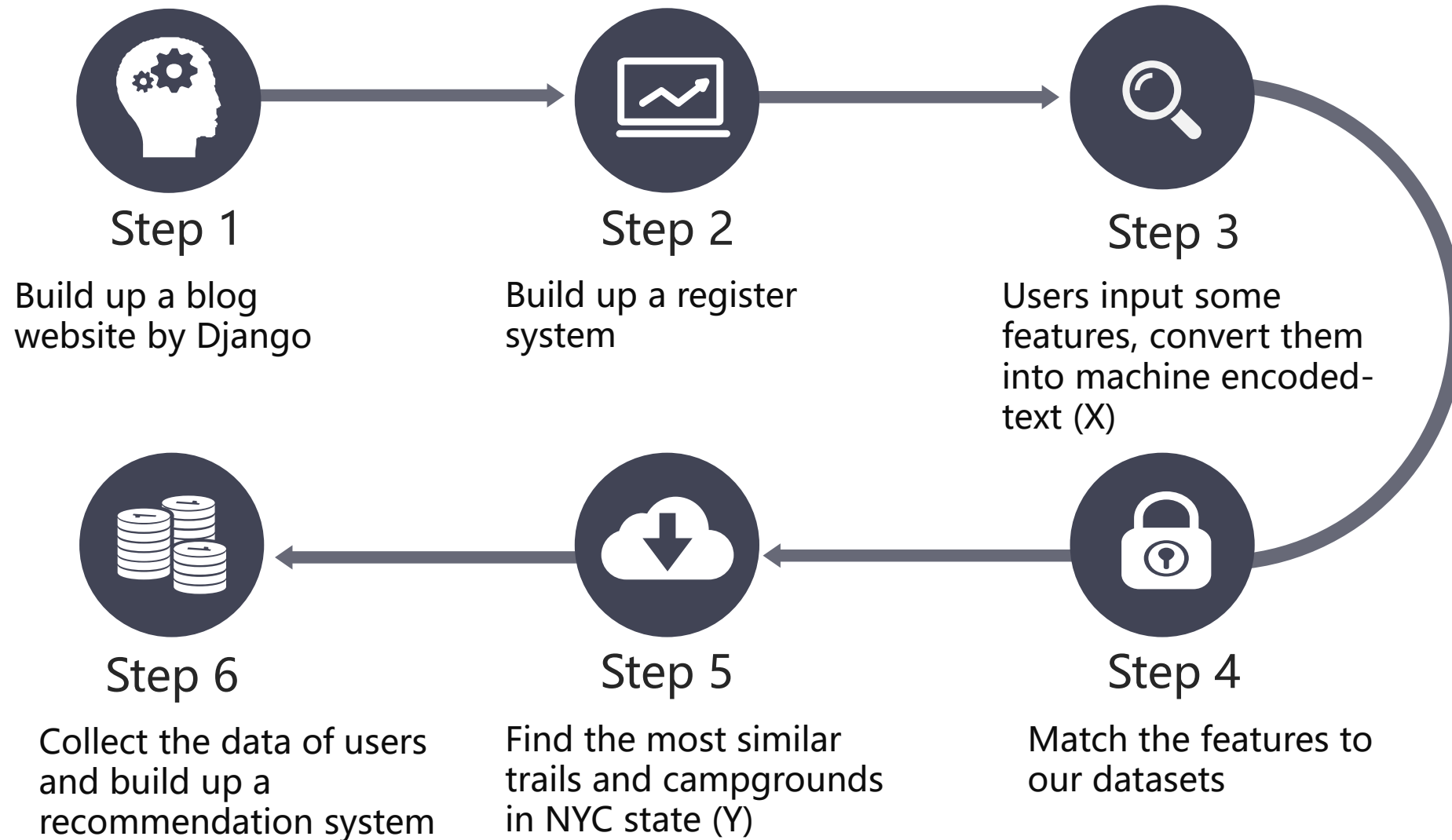
Data Source & Data Cleansing

- Method to calculate the difficulty level of each trail

| | | | |
|---|----------------------|--------------------|-----------------------|
| Length of Hike (Round Trip in miles)= | <input type="text"/> | Score Level | Difficulty |
| Total Elevation Gain (Feet)= | <input type="text"/> | 0 - 7 | <u>EASY</u> |
| <div>Calculate Hike Difficulty</div> <div><input type="text"/> Score</div> | | 8 - 13 | <u>MODERATE</u> |
| | | 14 - 18 | <u>CHALLENGING</u> |
| | | 19 - 23 | <u>DIFFICULT</u> |
| | | 24 - 27 | <u>VERY DIFFICULT</u> |
| | | 28 + | <u>EXTREME</u> |
| <div>Reset</div> <div>Back</div> | | | |

Method comes from: <http://www.nwhiker.com/HikeEval.html>

Method



Technique

- Content-based recommendation engine
- TF-IDF(Term Frequency – Inverse Document Frequency)
- Similarity is calculated by Cosine Similarity

TF-IDF

TF-IDF is a measure of originality of a word by comparing the number of times a word appears in a doc with the number of docs the word appears in.

$$\text{TF-IDF} = \text{TF}(t, d) \times \text{IDF}(t)$$

Term frequency

Number of times term t appears in a doc, d

Inverse document frequency

$\log \frac{1 + \overset{\text{\# of documents}}{n}}{1 + \underset{\text{Document frequency of the term } t}{df(d, t)}}$

Ideal Outcome

- Using our web app, user is able to find top-N trails and campgrounds which satisfy his needs
- All of the trails and campgrounds will gain a chance to be evaluated by hikers
- Users can get recommendations from our web app

