
Tourist Attractions
Recommendation &
Information System
Indiehunt

Problem Statement

1. Given a free text query as input it shows relevant Indian tourist spots
2. Given a destination picture as input it shows similar Indian tourist spots

Motivation & Challenges

- India is a diverse country country with enormous tourism potential
- To help promoting tourism in India and facilitate better travel planning
- Tourism based websites can handle only a limited variety of queries
- Existing websites fall short on providing personalized travel recommendation and general information service
- Lack of a comprehensive dataset that include information and pictures of numerous tourist destinations tourist destinations

Existing Realtime Solutions…

Existing Research works…

- Models for personalized travel recommendation based on user preferences and previous travel history :
  - Collaborative Filtering based model
  - Deep neural network based model
  - Ensemble recommender based on collaborative filtering & content-based filtering
- Tourism information service depending on freely accessible data and pictures over the web
- CNN based image classifier to provide similar destinations given a photo query

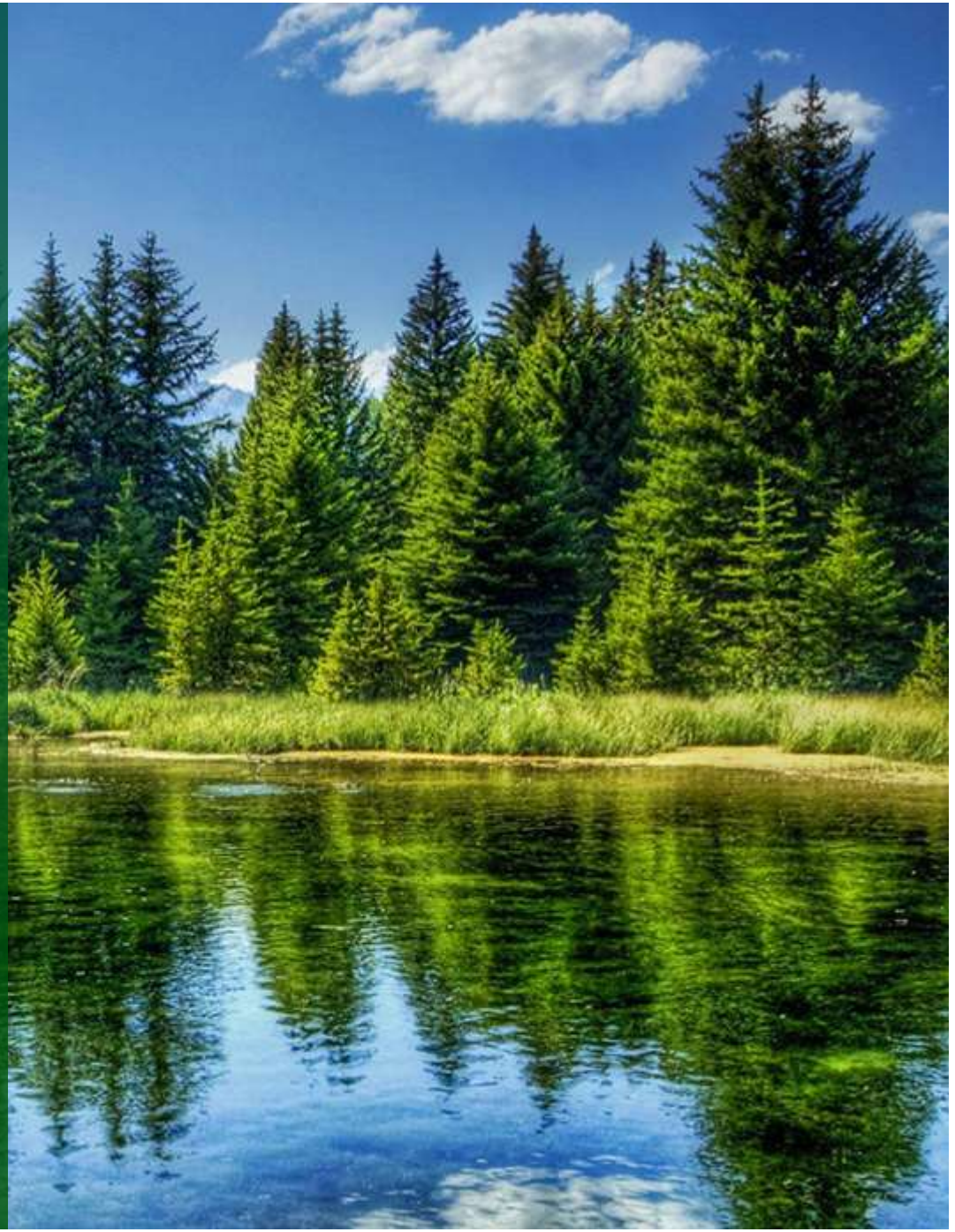
Data Collection & Pre-processing

- Scraped data from Google Maps using Apify
- Divided the tourist places into ten categories, including pilgrimage, beach, hill station, wildlife/forest, adventure, historical place, museums, trekking, desert, and smart city.
- Data cleaning :
  - Removed entries having null / non-alphabetic entries
  - Removed entries with same place title
- Sorted the entire data in non-increasing order based on a popularity-score,
  where popularity-score = Rating * Total count of ratings
- Extracted keywords and adjectives from the place description and user reviews column using KeyBERT and NLTK POS_Tagging respectively, and then concatenated them with all the reviewTags to create another column called feature_set

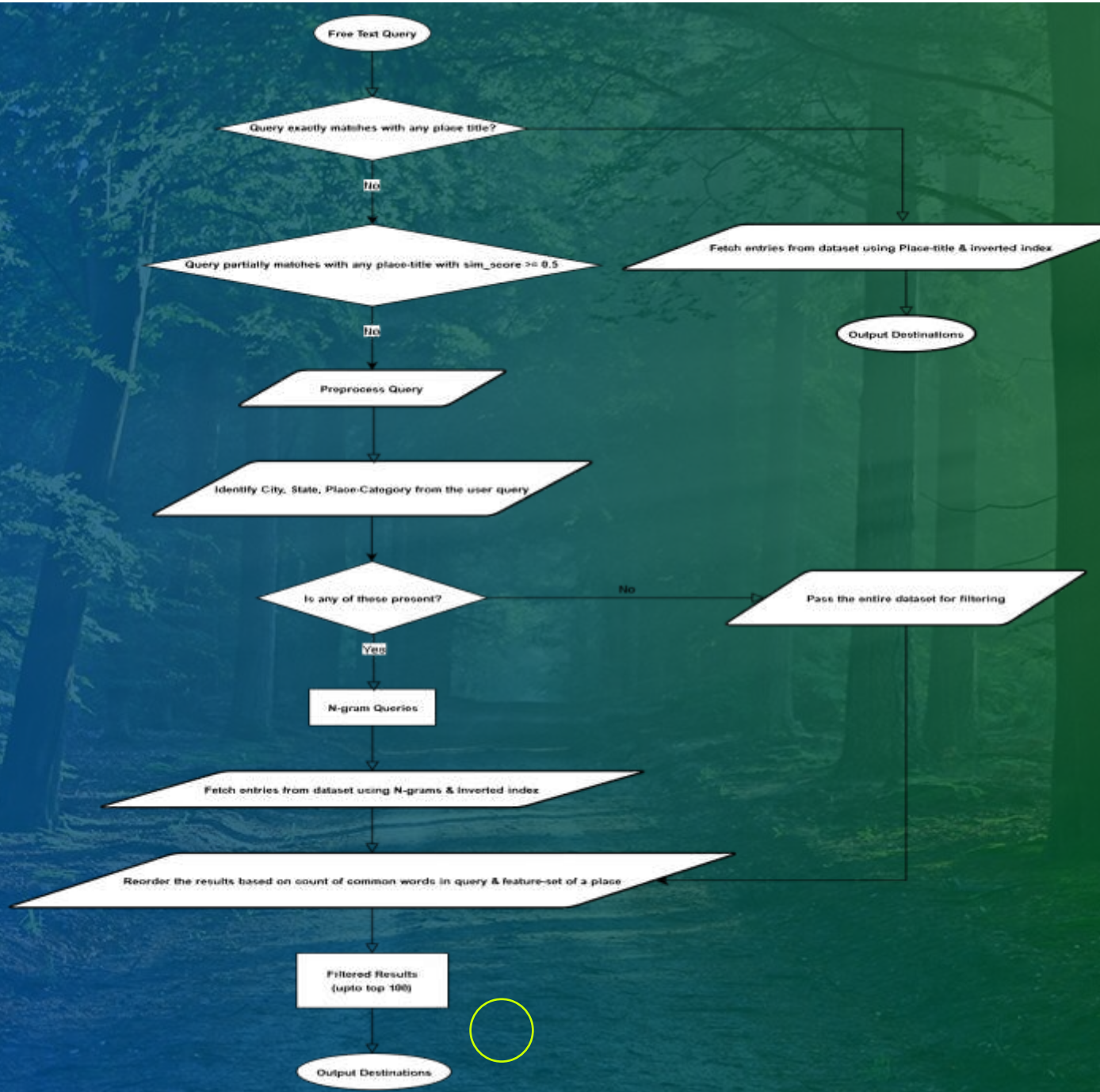
Proposed Solution

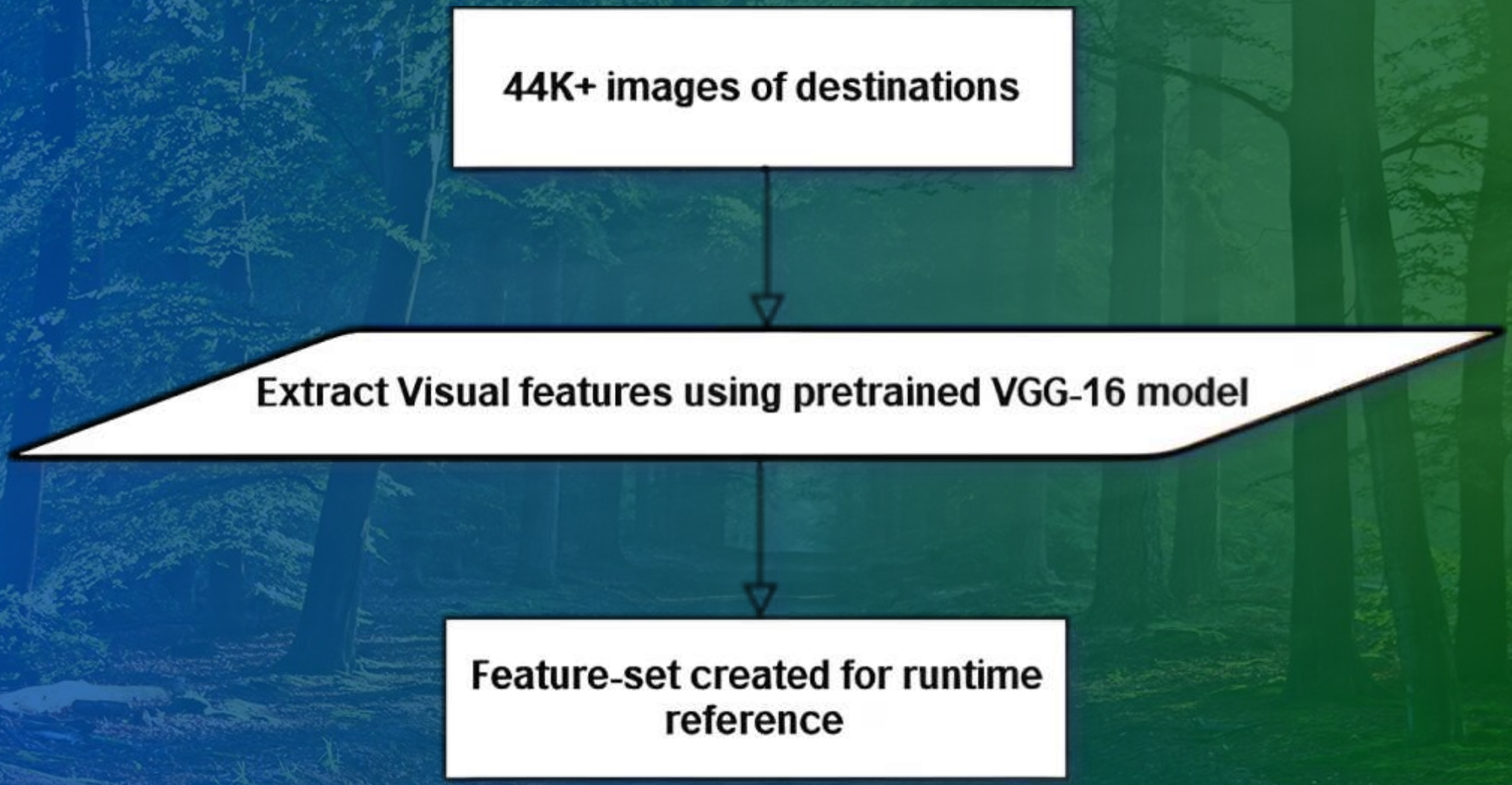
A multimodal search engine where a user can choose to either give a text query or a photo query to see a collection of relevant destinations

- Given a text query multiple NLP based matching algorithms will work on the backend to filter out the desired destinations and then rank them based on relevance and popularity score.
- Given a photo query pretrained CNN model is used to extract its visual features and places having similar features are shown as recommendation ranked on cosine similarity score.
- Nearby places of each destination is obtained using its latitude and longitude information.



Pipelines of the Solutions


Free Text Query
Query exactly matches with any place title?
No
Query partially matches with any place-title with sim_score >= 0.5
No
Preprocess Query
Identify City, State, Place-Category from the user query
Is any of these present?
No
Yes
N-gram Queries
Fetch entries from dataset using N-grams & inverted index
Fetch entries from dataset using Place-title & inverted index
Output Destinations
Pass the entire dataset for filtering
Reorder the results based on count of common words in query & feature-set of a place
Filtered Results
(upto top 100)
Output Destinations


44K+ images of destinations
Extract Visual features using pretrained VGG-16 model
Feature-set created for runtime reference

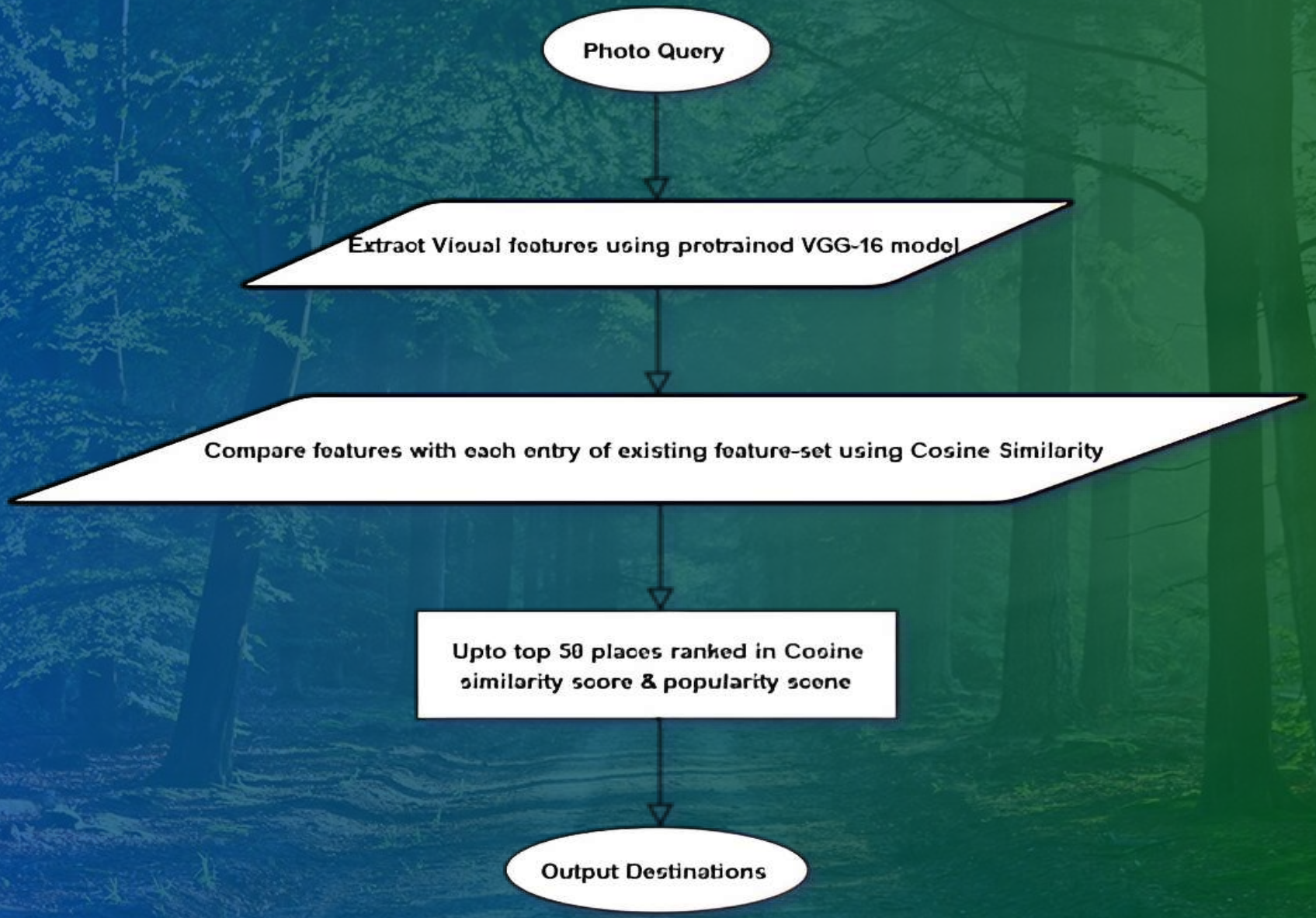
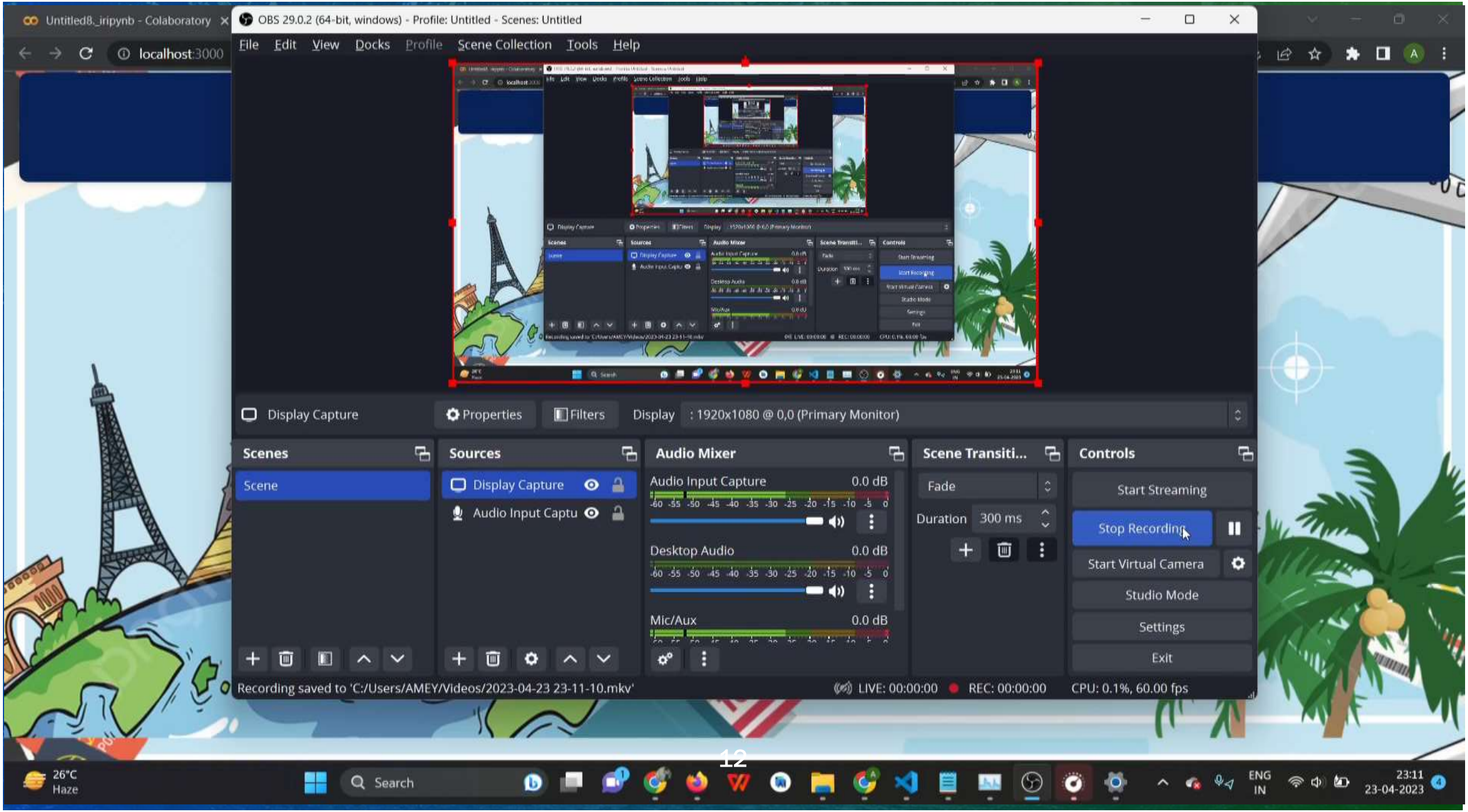


Photo Query
Extract Visual features using pretrained VGG-16 model
Compare features with each entry of existing feature-set using Cosine Similarity
Upto top 50 places ranked in Cosine similarity score & popularity score
Output Destinations


Untitled8_iripynb - Colaboratory
localhost:3000
OBS 29.0.2 (64-bit, windows) - Profile: Untitled - Scenes: Untitled
File Edit View Docks Profile Scene Collection Tools Help
Display Capture
Properties
Filters
Display : 1920x1080 @ 0,0 (Primary Monitor)
Scenes
Scene
Sources
Display Capture
Audio Input Captu
Audio Mixer
Audio Input Capture 0.0 dB
Desktop Audio 0.0 dB
Mic/Aux 0.0 dB
Scene Transiti...
Fade
Duration 300 ms
Controls
Start Streaming
Stop Recording
Start Virtual Camera
Studio Mode
Settings
Exit
Recording saved to 'C:/Users/AMEY/Videos/2023-04-23 23-11-10.mkv'
LIVE: 00:00:00
REC: 00:00:00
CPU: 0.1%, 60.00 fps
26°C
Haze
Search
ENG
IN
23:11
23-04-2023

Novelty

- None of the existing works provides the flexibility to give both text query and photo query.
- Unlike our solution majority of the websites fails to show appropriate results when there is no city or state name in the text query.
- Our solution also emphasizes on the descriptive words in the text query to make the output content more relevant.
- Curated a new comprehensive Indian tourism dataset

Future Work

- We can train a new CNN based model or fine tune a pretrained model on our image collections for better visual feature extraction. For this we also need to accurately annotate the category labels of the images.
- We can build a NER model based on Indian places to automatically identify city, state from text query.
- Few other techniques can be designed to optimize time.

Contributions

- ✓ **Backend Code –** Antara, Indraayudh
- ✓ **Data Collection & Preprocessing –** Taasir, Pranshu, Indraayudh, Antara
- ✓ **UI –** Taasir, Pranshu, Amey, Arun
  - **Search by text + Nearby places :** Taasir, Pranshu
  - **Search by Image + Place information :** Arun, Amey

THANK YOU!

Antara Das – MT22014
Amey Subhash Pawar – MT22010
Arun Sen – MT22018
Indraayudh Talukdar – MT22031
Mohammed Taasir – MT22029
Pranshu Patel – MT22117