Water Bodies Classification

Machine Learning and Data Mining II

Team members:

Le Viet Anh
Do Thanh Dat
Vu Tuan Phuong Nam

Bui Quang Ha
Tran Thi Huong Giang
Le Duc Manh
Nguyen Huu Dung

Lecturer: Dr. Doan Nhat Quang

Content

1. Introduction
2. Project Overview
3. Process
4. Conclusion
5. Future work

1. Introduction

Shortly, this project aims to identify water bodies from maps, such as lakes, rivers, etc.

2. Project Overview

Detecting water bodies and classifying them based on colors without using CNN(Convolutional Neural Network).
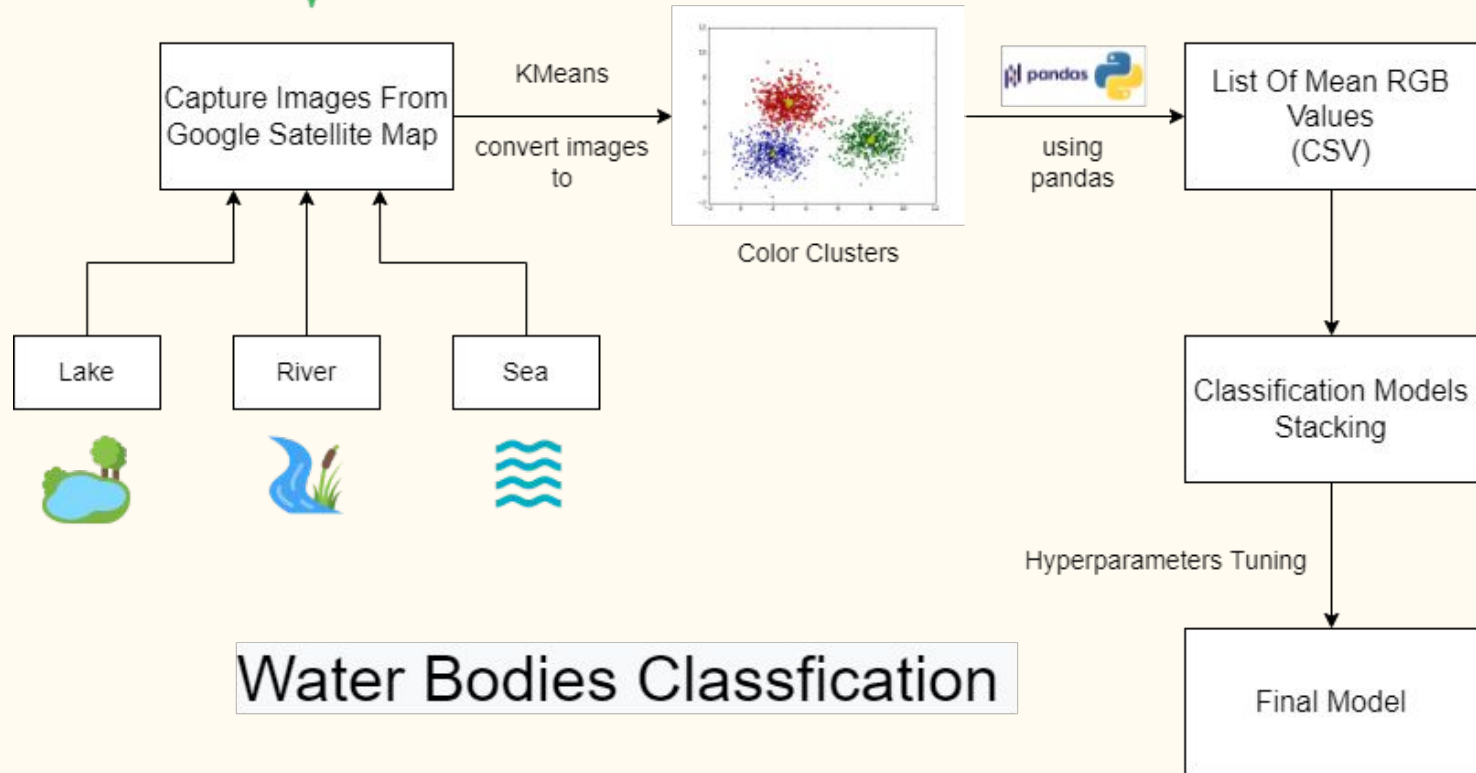
Programming Language: Python

Frameworks:

- Pandas
- Sklearn

3. Process

3.1 Procedure Diagram

3.2 Collecting photos from Google Satellite Map

3.3 1st images processing

- Using KMeans to group colors from an image
- Get a list of the images RGB values
- Scaling the dataset
- PCA - Principal Component Analysis

Evaluation:

After stacking SVC and Decision Tree, unfortunately the accuracy did not improve, going down from 0.89 to **0.86 or 86%**.

Model	KNN	SVC	Decision Tree	GNB	MNB	CNB
Accuracy Score	0.80	0.89	0.89	0.67	0.78	0.56
n_neighbors	5	-	-	-	-	-
kernel	-	rbf	-	-	-	-
C	-	10000	-	-	-	-
gamma	-	0.1	-	-	-	-
max_depth	-	-	5	-	-	-
min_samples_leaf	-	-	5	-	-	-

3.4 2nd images processing

- Improve input images
- Choosing the best K number of clusters in KMeans

Evaluation:

After trying several different models, our new accuracy is much better, at **0.96 or 96%.**

Model	Decision Tree	SVC	KNN	GNB	CNB	MNB
Accuracy Score	0.92	<u>0.96</u>	0.94	0.91	0.74	0.87
C	-	10	-	-	-	-
gamma	-	1	-	-	-	-

4. Conclusion

Most accurate model ?

Bases on 2 attempts

SVC or Support Vector Classifier

Application

Promising field in the future

Detecting water bodies and classifying them based on the colors is a promising approach that can be applied in geographical analysis and geomarketing.

Interesting Findings

Google Map satellite images were unreliable in certain regions, due to geopolitical reasons. Belows the image were of strange colors and patterns.
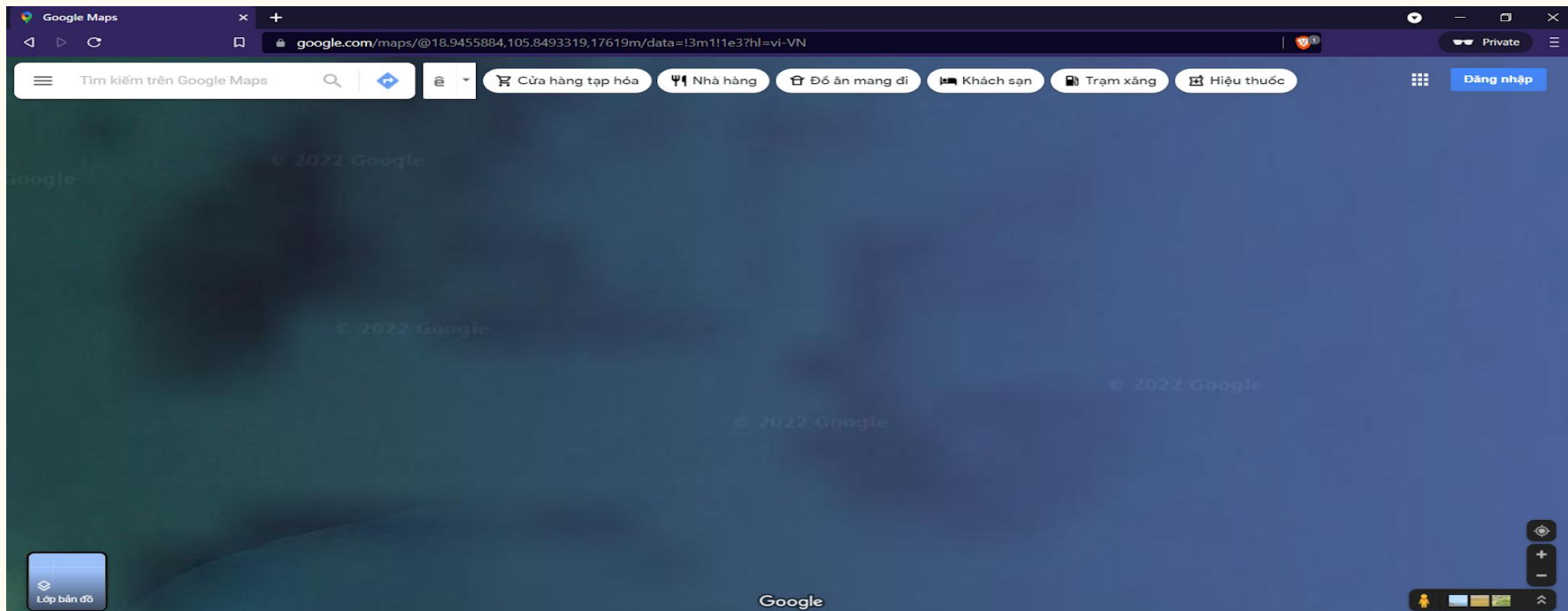
5. Future work

The only feature we are missing in the dataset is the location of the water source. Rivers in Northern Vietnam will have different color shades compared to the Ganga River in India since they have different sources.

Thank you for listening!