

# 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

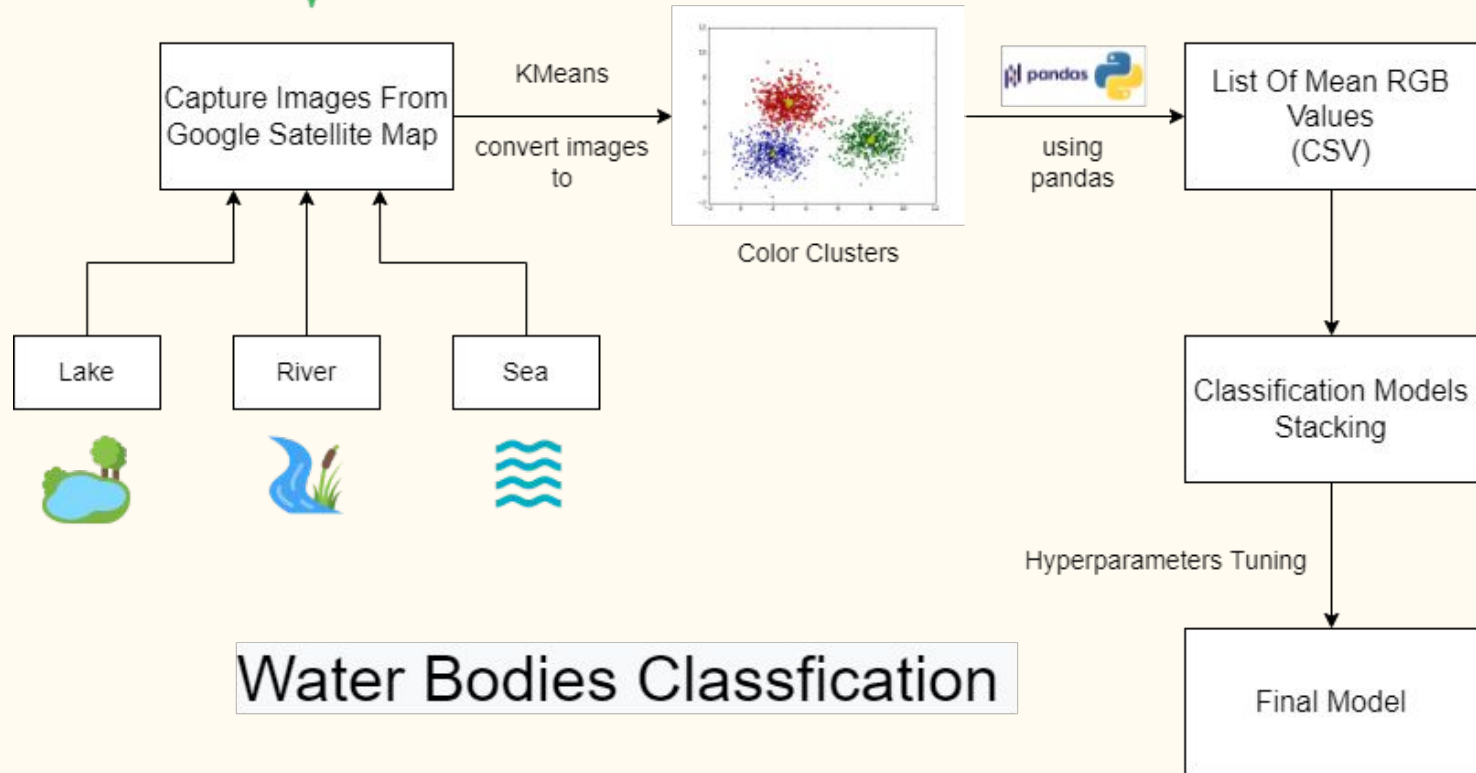
Frameworks:

- Pandas
- Sklearn

# 3. Process

—

## 3.1 Procedure Diagram



## 3.2 Collecting photos from Google Satellite Map



Ho_B52.png	Ho_Ba_Mau.PNG	Ho_Bay_Mau.PNG	Ho_Cau_4_mua.PNG	Ho_Co_Dung.PNG	Ho_cong_vien_a_m_nhac_Do_Ngh	Ho_Cong_vien_Cau_Giay.PNG	Ho_Cua_Khau.PNG	Ho_Dai_Dong.PNG
------------	---------------	----------------	------------------	----------------	------------------------------	---------------------------	-----------------	-----------------



Ho_Den_Lu.PNG	Ho_dieu_hoa_Nh	Ho_Dinh_Cong.P	Ho_Dong_Da.PN	Ho_Dong_Mo_1.	Ho_Dong_Mo_2.	Ho_Dong_Mo_3.	Ho_Dong_Mo_4.	Ho_Giam.PNG
---------------	----------------	----------------	---------------	---------------	---------------	---------------	---------------	-------------



Screenshot_2022 0320_211424.png	Screenshot_2022 0320_211453.png	Screenshot_2022 0320_211522.png	Screenshot_2022 0320_211544.png	Screenshot_2022 0320_211610.png	Screenshot_2022 0320_211629.png	Screenshot_2022 0320_211648.png	Screenshot_2022 0320_211723.png	Screenshot_2022 0320_211736.png
------------------------------------	------------------------------------	------------------------------------	------------------------------------	------------------------------------	------------------------------------	------------------------------------	------------------------------------	------------------------------------

[illegible]

1.png 2.png 3.png 4.png 5.png 6.png 7.png 8.png 9.png



10.png 11.png 12.png 13.png 14.png 15.png 16.png 17.png 18.png

## 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	<u>0.89</u>	<u>0.89</u>	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

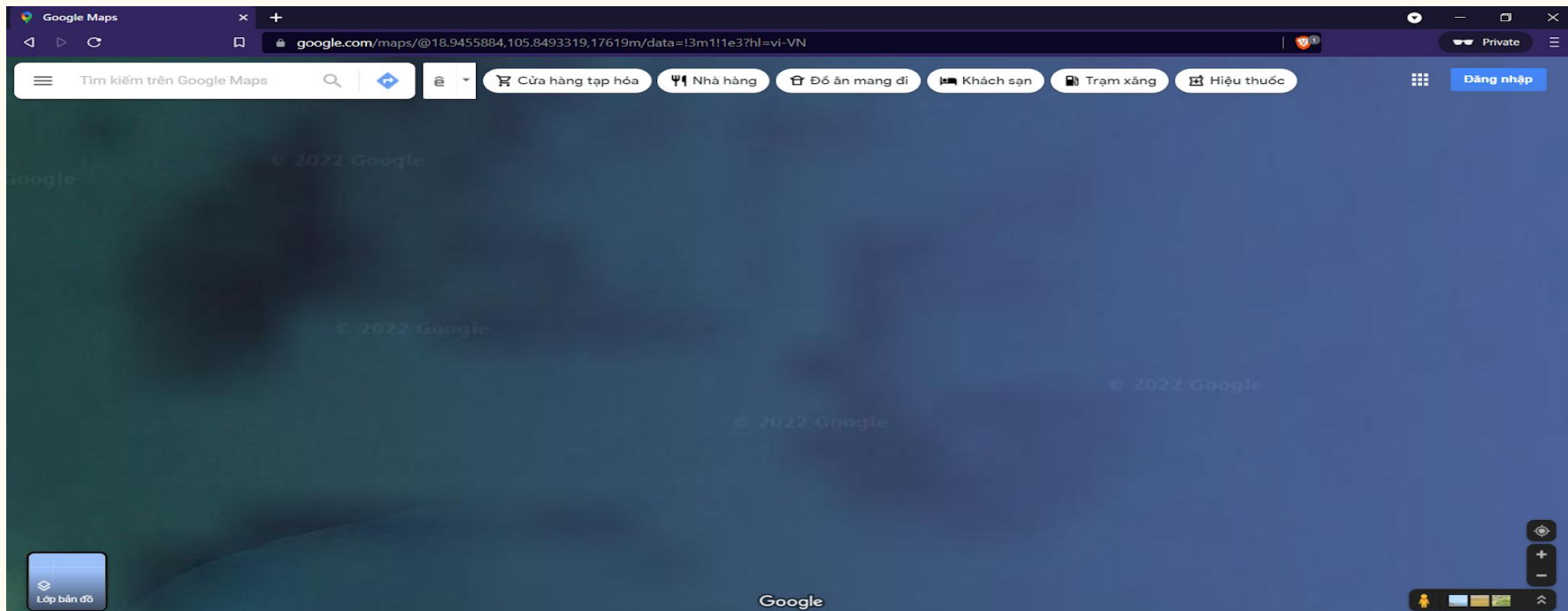
Promising field in the future

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!*

—