DEFORESTATION ANALYSIS USING UNSUPERVISED CLUSTERING AND SATELLITE IMAGES

A PROJECT REPORT

Submitted by

RUSHAN MUKHERJEE [RA1611003011168]
MAYANK DEWANGAN [RA1611003010417]

Under the guidance of

Dr.K.PRADEEP MOHAN KUMAR, Ph.D
(Assistant Professor, Department of Computer Science & Engineering)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

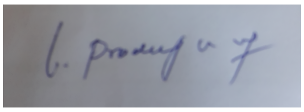
S.R.M. Nagar, Kattankulathur, Kancheepuram District

MAY 2020

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that this B.Tech project report titled "**DEFORESTATION ANALYSIS USING UNSUPERVISED CLUSTERING AND SATELLITE IMAGES"** is the bonafide work of **RUSHAN MUKHERJEE [Reg No: RA1611003011168] and MAYANK DEWANGAN [Reg No: RA1611003010417]**,who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.



Dr. K. PRADEEP MOHAN KUMAR
SUPERVISOR
Associate Professor
Department of Computer Science and Engineering

Dr. B. AMUTHA
HEAD OF THE DEPARTMENT
Department of Computer Science and Engineering

Signature of the Internal Examiner

Signature of the External Examiner

Own Work Declaration

Department of Computer Science and Engineering

SRM Institute of Science & Technology

Own Work* Declaration Form

This sheet must be filled in (each box ticked to show that the condition has been met). It must be signed and dated along with your student registration number and included with all assignments you submit – work will not be marked unless this is done.

To be completed by the student for all assessments

Degree/ Course : **Bachelor of Technology / Computer Science and Engineering**

Student Name : **Rushan Mukherjee , Mayank Dewangan**

Registration Number : **RA1611003011168 , RA1611003010417**

Title of Work : **Deforestation Analysis Using Unsupervised Clustering and Satellite Images**

I / We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism**, as listed in the University Website, Regulations, and the Education Committee guidelines.

I / We confirm that all the work contained in this assessment is my / our own except where indicated, and that I / We have met the following conditions:

- Clearly references / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook / University website

I understand that any false claim for this work will be penalised in accordance with the University policies and regulations.

DECLARATION:
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above.
If you are working in a group, please write your registration numbers and sign with the date for every student in your group.

Rushan Mukherjee
RA1611003011168
Rushan Mukherjee
25 May, 2020

Mayank Dewangan
RA1611003010417
Mayank
25/05/20

ACKNOWLEDGEMENT

We express our humble gratitude to **Dr. Sandeep Sancheti**, Vice Chancellor, SRM Institute of Science and Technology,for the facilities extended for the project work and his continued support.

We extend our sincere thanks to **Dr. C. Muthamizhchelvan**, Director, Faculty of Engineering and Technology, SRM Institute of Science and Technology, for his invaluable support.

We wish to thank **Dr.B.Amutha**, Professor & Head, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for her valuable suggestions and encouragement throughout the period of theproject work.

We are extremely grateful to our Academic Advisor **Dr.A.JeyaSekar,** Associate Professor, and **Dr.R.Annie Uthra**, Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for their great support at all the stages of project work.

We would like to convey our thanks to our Panel Head, **Dr.M.Prakash,** Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for his / her inputs during the project reviews.

We register our immeasurable thanks to our Faculty Advisor**, Mrs.R.Lavanya,** Assistant Professor, **Dr.G.Niranjana,** Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to my guide, **Dr.K.Pradeep Mohan Kumar,** Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for providing me an opportunity to pursue my project under his/her mentorship. He / She provided me the freedom and support to explore the research topics of my interest. Her / His passion for solving the real problems and making a difference in the world has always been inspiring.

We sincerely thank staff and students of the Computer Science and Engineering Department, SRM Institute of Science and Technology, for their help during my research. Finally, we would like to thank my parents, our family members and our friends for their unconditional love, constant support and encouragement.

Rushan Mukherjee

Mayank Dewangan

ABSTRACT

The expansion of farmland and unplanned land encroachments has increased as the earth's population boomed, which has led to uncontrolled deforestation across the world.Deforestation and industrialization have given rise to Global warming, which has caused mayhem in the current ecosystem. The weather patterns are disrupted, and natural calamities are occurring more frequently. The after-effects of these events have to lead to dramatic losses in flora and fauna. Even though a large part of India's land is urbanized, there are many protected areas in specific parts of the country that represent significant vegetation that has been affected if we observe from the scale of a subcontinent. In this paper, we aim to trace the deforestation which has happened in the Sundarbans over the period of 1988 to 2019. We have selected the Sundarbans because mangroves are a natural defense to cyclones and also provide shelter to a plethora of living organisms. We will make use of the K-means clustering which is an unsupervised learning algorithm on satellite imagery from Landsat-5, Landsat-7 and Landsat-8 which have been orthorectified.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

GIS	Geographical Information Systems
GEE	Google Earth Engine
NDVI	Normal Difference Vegetation Index
NDWI	Normal Difference Water Index
NDBI	Normal Difference Built-in Index
NIR	Near Wave Infrared
USGS	United States Geological Survey
PX	Pixels

LIST OF SYMBOLS

Σ	Summation
μM	Micro metre
ϵ	Belongs to
$\%$	Percentage
N^o	North
E^o	East
S^o	South
W^o	West

CHAPTER 1

INTRODUCTION

1.1 General

Geographical Information Systems (GIS) frameworks are used for accumulation , regulating, and understanding data. It combine many data despite being based on the geographics. It analyzes locations and organize them into different slices of data. These are then transformed into visualizations using maps and 3D scenes. We plan to integrate GIS with Machine Learning classification techniques to conduct our project. Unsupervised classification is a sub branch of machine learning classification algorithms which tries to find patterns from unlabeled data. This project will be a testament of interoperability between the above mentioned domains to analyze real world scenarios like deforestation. Deforestation is the result of rapid industrialization of rural land of India. To combat deforestation is to combat global warming and reduce pollution levels of the whole world.The earth just lost a huge chunk of its forest land in the Amazon because of illegal mining and human negligence which caused wildfires. The results of our study will not only help government reduce crucial time in identifying “stressed” areas prone to deforestation but will also provide a consolidated map of the locations where the forests are blooming again .Our information can also be used by NGOs who are willing to do grass-root level work of educating people not to destroy forests illegally.In this project, we aim to make a common architecture for deforestation analysis of India and also making it open source so that anyone can access the information and learn from it.

1.2 Purpose

The Sundarbans is the biggest mangrove forest region in the country, covers 2.47 million acres in Bengal and Bangladesh combined. In tropics and subtropics, along the coast most rich in carbon. Mangrove habitats, consist of salt-tolerant shrubs and trees, play a crucial role. Supplying natural coastal defence against storm surges by minimizing wave and wind strength. Such environment monitoring depends on satellite data to assess the present condition. From 2001 to 2012 a gross decline since 2000 of 1.38 per cent (or 0.13 per cent annually). Tropical deforestation stands at a record of 4.9 per cent loss (or 0.41 per cent annually). Advances in vegetation cover satellite imagery give fresh ways to track tree transition with much greater accuracy and precision than ever before. Annual revisions to the global tree cover depletion data provide an incentive for reliable tracking of forest transition in mangrove habitats. And satellite imagery mining records provide the ability to chart recent shifts, providing insight to the mangrove depletion situation prior to 2000. Due to variations in observational methodology we chose not to equate our estimates with those results. Deforestation has been a global issue for decades, largely because of our growing rising population and raising resource demand. Mangrove forests are seriously endangered by anthropogenic exploitation and deforestation because they depend on very limited environments for survival. The largest mangrove region of about 198,000 km2 worldwide was in 1980. The overall region was less than 150,000 km2 in 2000, a decline of 25 per cent from 1980. This indicates that the average pace of deforestation over these twenty years ranged from 1-per cent. The key causes for mangrove forest depletion are urbanization, aquaculture and over-exploitation of wood, shrimp, and shellfish.

1.3 Scope

The latest developments with remote sensoring technologies have increased the quality of low or even no cost satellite pictures, with high quality of data. This, in addition, extended the limits of our capabilities, so the creation of fairly medium to high-global land cover mappings to facilitate cost-forest surveillance. Likewise, modern cloud-such as Google Earth Engine (GEE) also created the capacity to store bytes and bytes of imagery of world level, as well as with free study access.The developments enable data collection and processing on a much larger scale and also in fairly inaccessible areas with inadequate data collection capability at higher precision. Based on these global strategies we utilize related methods in a regional environment to develop a land cover and measure forest change through Landsat image time series analysis.

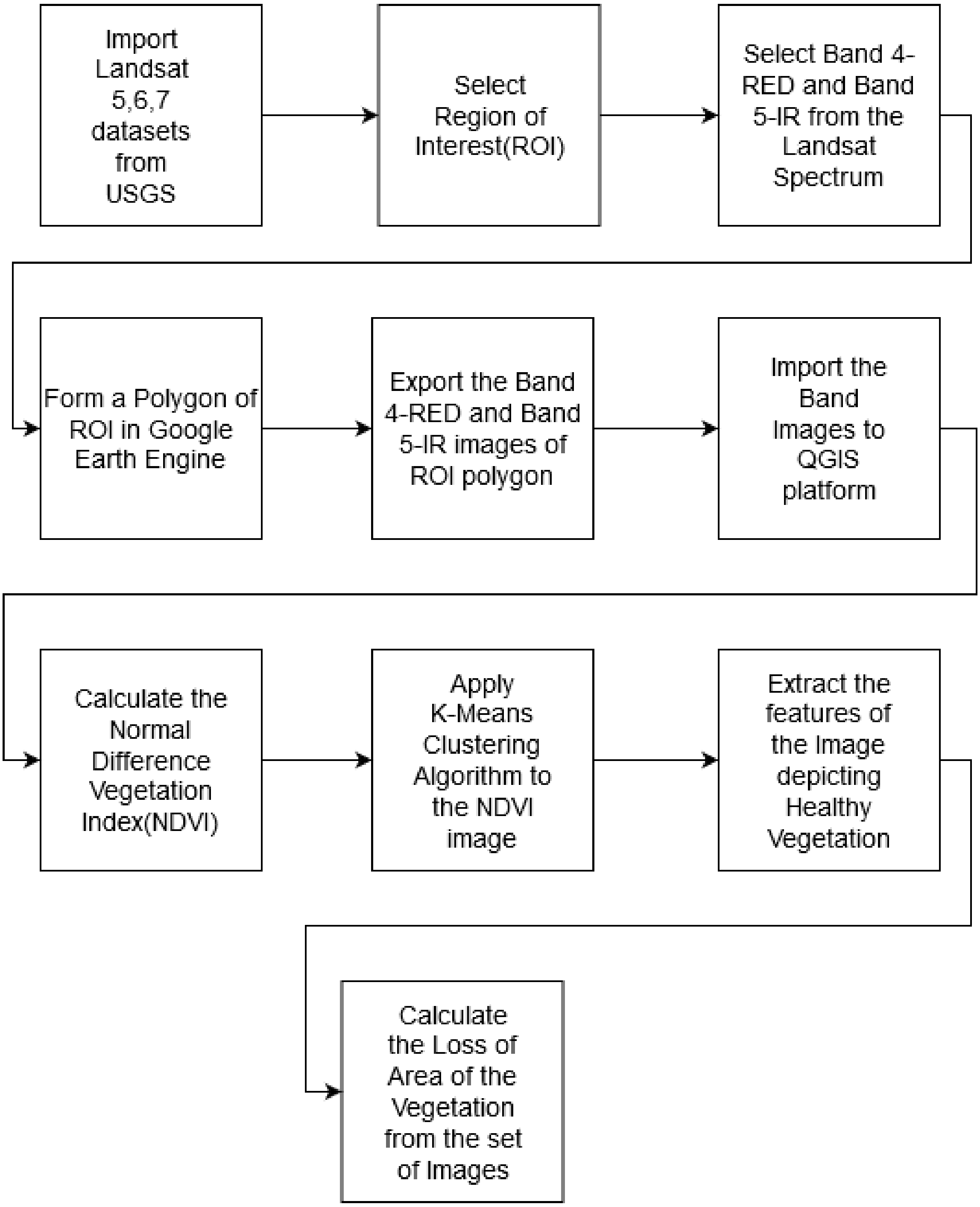


Figure 1.1: Architecture Diagram

CHAPTER 2

LITERATURE SURVEY

2.1 Review

[1] The registered correctness of the arrangement procedure were sensible, clarified way all-out accurately ordered large in the presence of Pixels (PX).The related basic property externalities required at nearby, local and worldwide scales, require the observing of land use elements crosswise over forested scenes in creating future techniques and approaches concerning rural enhancement, common woodland protection and monoculture tree estates. The managed arrangement (Maximum probability) on satellite information done to evaluate the land usage from 1998 to 2010.The utilization of remote detecting information for observing versatile part like crosswise over assorted scenes. [2]The study surveyed deforestation in woods environment of Rudraprayag area situated in Himalay Mountain Regions. Investigation built up a model for loss of flora-fauna vulnerability utilizing recurrence proportion in Geological Information systems condition. The outcomes showed variables (incline with height) with anthropogenic components (good ways from street, closeness of backwoods to the settlement and agrarian vicinity to timberland) have emphatically quickened deforestation. The investigation region encountered lost about 112.5 km square woods region in 90s. [3]The accessibility of the recorded information of Landsat images as the result of the innovation of GEE point towards a innovations in the mechanism, checking with analysing area use combined with over enormous geographic areas. Investigation effectively builds up a territorial scale examination and decides the classes and the circulation of land spread in Savannah River. It also distinguished the spatial and the worldly difference in the land spread that happened as a result of the revisions in land usage in the course of recent years in the Savannah River bowl. Change of land use was seen prevailing in the backwoods zones during all interims of the investigation timeframe.

[4]LULC pictures were acquired from LandSat8 and Sentinel2 informational indexes utilizing a MLC technique which is px based, outcomes was assessed utilizing exactness evaluation by 400 arbitrary focuses. Overall exactness also Kappa coefficient at LandSat8 implied LULCC & Sentinel2 determined LULCC were 83 % , 0.7 and 88% , 0.8 because of precision evaluation.

In spite of the fact that it appears that Sentinel-2 speaks to LULC superior to Landsat-8 by and large, this circumstance can change if distinctive grouping techniques and insights are utilized. Albeit generally speaking precision of Sentinel2 inferred is superior to Landsat-8 determined results. [5] Divided legitimate timberland limits of state-claimed woods and community/private woodland were utilized for the appraisal of the backwoods spread in 2005 and 2011.The limits were mapped and confirmed on VHR picture.The depicted limits were utilized to survey the LC and woodland spread change from 2005 to 2011 exclusively for the state-possessed woodland and community/private timberland. [6]The related basic property externalities required at nearby, territorial and worldwide scales, require the observing of land use elements crosswise over forested scenes in creating future systems and strategies concerning rural enhancement, regular woodland protection and monoculture tree estates. The figured correctness of the arrangement procedure were sensible can clarified way that the all-out num of effectively grouped px were high.

2.2 Inference from survey

Accurate analysis of deforestation with satellite imagery is difficult because there are various factors involved like clouds and agricultural land which classify as noise in our datasets. There are major takeaways from the survey.There is a need for a comprehensive dataset that has very low noise. Adequate area of interest mapping required to scope out the initial field of view. Proper labelling of images has to be done so as to reduce training time. Testing the model on a variety of forest types for better accuracy.

CHAPTER 3

DATA PREPROCESSING

3.1 Introduction

The Landsat system is Earth's longest operating effort to collect satellite imagery. Formerly known as Earth Resources Technology Satellite. Landsat satellite photos are stored in the U.S. and are a valuable tool for climate change studies and implementation in agriculture, geography, education, forestry, development work and aerial observation at numerous Landsat receiving stations across the world, and can be accessed via the website United States Geological Survey (USGS).The 8 spectral bands of Landsat 7 have their spatial resolution 15m-60m.Whereas resolution of Landsat 7 is 2 weeks. The data taken at Landsat are easy to download as they are split into scenes. A Landsat scene is 185km x 185km. Landsat 7 data has eight spatial resolution spectral bands 15m x 60m with a resolution of 2 weeks. These data are typically fragmented for convenience in transferring images. A Landsat coverage is about 185 by 185 kilometers.

<table>
<tr><th>Features</th><th>Landsat-5</th><th>Landsat-7</th><th>Landsat-8</th></tr>
<tr><td>Acquisition Date</td><td>19 February,1988</td><td>28 February, 2000</td><td>23 January, 2019</td></tr>
<tr><td>Number of Pixels</td><td colspan="3">50176px</td></tr>
<tr><td>ROI Polygon(Longitude, Latitude)</td><td colspan="3">88.23ºE, 22.16ºN, 88.23ºE, 21.47ºN, 89.12ºE, 21.47ºN, 89.12ºE, 22.16ºN</td></tr>
<tr><td>Band Combination</td><td colspan="2">Near Infrared = Band 4 ,
Red = Band 3</td><td>Near Infrared = Band 5 ,
Red = Band 4</td></tr>
<tr><td>Format</td><td>TIFF</td><td>TIFF</td><td>TIFF</td></tr>
</table>

Figure 3.1: Dataset Classification

3.2 Technique

Every Landsat picture comprises of 9 spectral bands.Combining these factors result in each pixel possessing a spectral designation. RGB bands generate contrasts in colors between different forms of land cover. These can be differentiated by unaided vision and these bands form a significant part of the spectral designation of any pixel. Including only the RGB will also narrow the spectrum of wavelengths.The Near Infrared (NIR) band with wavelength 0.7 m–0.9 m contains details on the green coefficient of the reflected surface that enables us classify live and stable plants, while the shortwave infrared (SWIR) bands with wavelengths of 0.9–2.5 mcontribute details on surface water quality that is helpful in separating the mangrove from the delta where vegetation is more likely to be present.Atmospheric disturbances don't affect higher wavelengths.These higher wavelengths built indicators that deteriorating land cover. Greater wavelengths are seldom influenced by ambient changes. Thus they are used to create indicators used to distinguish vegetated and barren land. The Surface-data provide important details for atmospheric corrections. When we were picking photographs with low cloud cover percentage. It may influence the sunlight absorption and, thus, the reflectance values of the wavelengths of the bands. As part of the prediction data, the information from these bands is used to provide evidence from pseudo-patterns which may influence the other bands ' reflectance values.The Landsat picture bands can be used to measure correlations capable of capturing different categories of land cover, which are instrumental in detecting vegetation and soil changes. Plain images give poor results in comparison to when vegetation indicators are used in tandem with machine learning algorithms.

CHAPTER 4

FEATURE SELECTION TECHNIQUES

4.1 K Means Clustering

Clustering is a process by which a collection of data is grouped into a particular category. K-means clustering algorithm divides a set of data into a set of k data groups. It classifies a given data set into dislocated clusters k. K-means algorithm consists of two steps. In the first phase it calculates the centroid k. The second step deals with the algorithm taking each point from the corresponding data point to the cluster which has the closest centroid. There are various techniques for determining the distance from the nearest centroid and Euclidean distance is one of the more widely known methods. When the grouping is done, it recalculates each cluster's current centroid. Depending on that centroid, a current Euclidean distance is determined between everycentre and every data point. Then the minimum Euclidean distance points in the cluster is determined. Each cluster in the partition is defined in terms of its member elements and centroid. The sum of distances from all the points in the cluster is minimized for each cluster at a point known as the centroid. Therefore k-means clustering is an iterative algorithm.In this, it minimizes the number of distances between each point as well as its cluster centroid for all clusters. Let's consider a picture with x×y resolution and also the picture must be clustered into k clusters.

Let coordinates-x,y be defined as input px for the clustor and CK be defined as centre of clusters. The algo is following below:

1. Initialize the number of clusters k, with centre Ck.
2. Calculate the Euclidean distance d, between the centre and each pixel of an image.

Euclidean Distance:$\boldsymbol{d = \left|\left|p(x,y) - Ck\right|\right|}$

3. Every pixel is assigned to the nearest centre based on the distance d.
4. New position of the centre is recalculated.

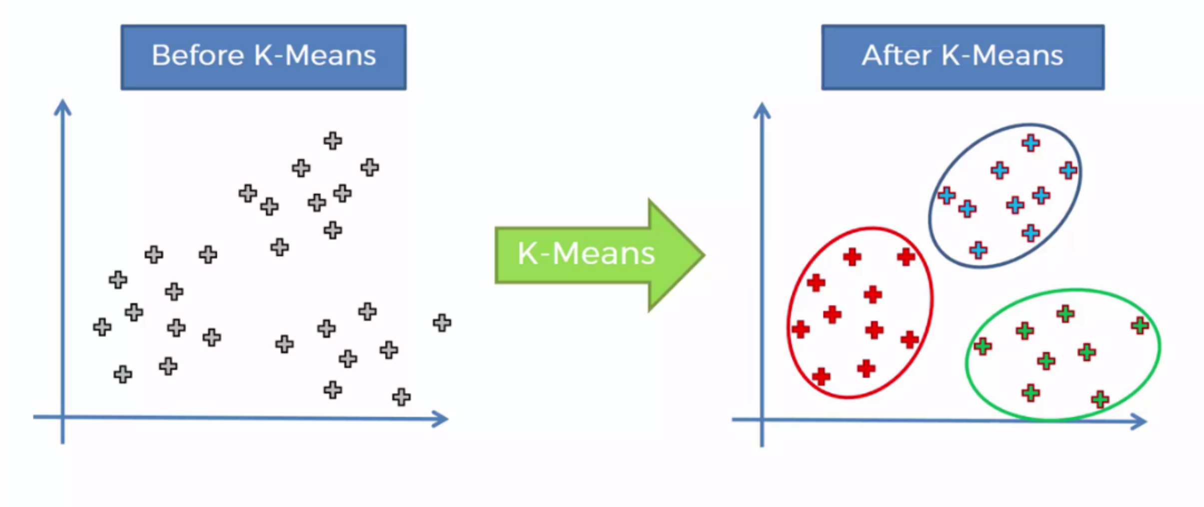


Figure 4.1: Kmeans Clustering

$$\textbf{New Centre Position:} Ck = \frac{1}{k} \sum_{y \in Ck} \sum_{x \in Ck} p(x, y)$$

5. Repeat until tolerance satisfied or error value.

6. Image formed by reshaping the cluster pixels.

Since the first centroid is randomly chosen, different initial centres can get different outcomes. Therefore the main centre should be consciously chosen such that we get segmentation of our desires. Computational complexity of the algorithm depends on the number of data units, the number of determined clusters and the size of the iterations.

4.2 Iterative Self Organizing Data Analysis Technique

This is also Unsupervised Clustering, uses either splits or merges of clusters. There are scenarios in merge and split, so it starts from a big cluster and then the split and merge operations are done according to the thresholds. The threshold depend upon their standard deviation and pixels. Kmeans uses priorii for assuming clusters.

The algorithm is following below:

1. With centroids placed at any random initial position with pixels in smallest distance with centre.

2.Cluster centres is calculated. Standard deviation(SD) for points in cluster calculated.

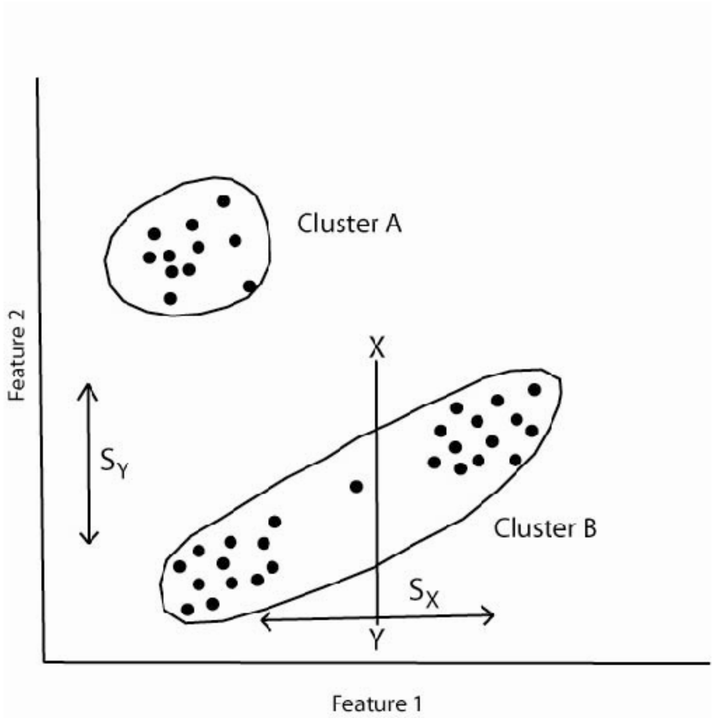


Figure 4.2: ISODATA Clustering

Clusters splitted if SD is greater than threshold(userdefined). Clusters merge when the in-between gap is less than the threshold(userdefined).

3.A second iteration is performed with the new cluster centers.

4.Further iterations are performed until:

4.1.The average inter-center distance falls below the user-defined threshold.

4.2.The average change in the inter-center distance between iterations is less than a threshold, or.

4.3.The max-iteration done.

CHAPTER 5

IMAGE SPECTRUM PROCESSING

5.1 NDVI

It represents the relation of red visible light (Pure, usually absorbed by chlorophyll in the plant) and Near Wave Infrared (NIR) wavelength (dispersed by the mesophyll structure of the leaf). The Normal Difference Vegetation Index (NDVI) is calculated as (NIR-RED)/ (NIR + RED). Using the following equation, the result from the Near-Infrared(NIR) band and RED band is drawn, the corresponding NDVI image includes values from-1 to 1 The value above 0.7 includes trees, 0.5-includes plants and 0.3-comprises shrubs and the negative value contains snow, mountains and other scattered structures.

5.2 NDWI

The water is a spectrum absorber but also reflects the visible spectrum. The Water presence in the geological site is depicted with infrared which is shortwave as the other infrared spectrum would either penetrate it or would get reflected. Short Infrared is about 1400 to 2400 nm. Normal Difference Water Index (NDWI) is calculated as (NIR-SWIR)/(NIR + SWIR).

5.3 NDBI

The Normal Difference Built-in Index (NDBI) is used to monitor urbanisation. The infrared wavelengths are used in this. Typically Shortwave and Near spectrum.The reflectance is high in such spectrum in the urbans and suburbans area. Non tree cover areas are basically barren land and the land which doesn't have any vegetation like open fields, roads and house rooftops. Sundarbans also shelter a big human populations consisting of fishermen and honeyhunters.It's formulae is (Shortwave - Near IR)/(Shortwave + Near IR).

CHAPTER 6

RESULTS & DISCUSSION

The major classes of palette were represented in correspondence to the Mangrove forest and the Delta. The Vegetation area from the Landsat satellite image is been analyzed by QGIS software. The palette of the NDVI images consists of five different colors representing the characteristics of the map components. The images of ROI from Near Infrared spectrum and Red spectrum are shown in image below.

6.1 NDVI RESULT

The analysis of vegetation is done by extracting the IR and R spectrum of the Region of Interest from the Landsat 5, Landsat 7 and Landsat 8. The NDVI is been implemented upon these images to calculate Normal Difference Vegetation Index of the respective area. The NDVI highlights the reflectance of the vegetation matter into the image. The resulting image has a set of different values from -1 to +1. The positive values reflect healthy vegetation and the negative values reflect the water bodies. The image shown below is of Sundarbans in the year of 1988. The image has a total of 50176 pixels. The total area of the region covered in the image is 7128.211 square Kilometers.

The Mangroves of Sundarbans in 1988 had a high density of healthy vegetation and areas that were undisturbed by any encroachments. The Amount of Mangroves is about 8562px out of the 50176px. The healthy vegetation is about 1216.3172 square kilometers. The Percentage of the vegetation cover is 17.0639%.The decade of 1980's had a considerably low population and demand when compared from today's standards.

Within 12 years there was a huge increase in demand and changes in the industrial policy in the country. The amount of healthy vegetation decreased to 7555px out of the 50176px. The Mangroves shrunk to 1073.2629 square kilometers. The percentage of the vegetation cover decreased about 2% to 15.0569%. The rate of urbanization has increased so much that in only a decade about 2% of Asia's Biggest Mangroves were diminished.

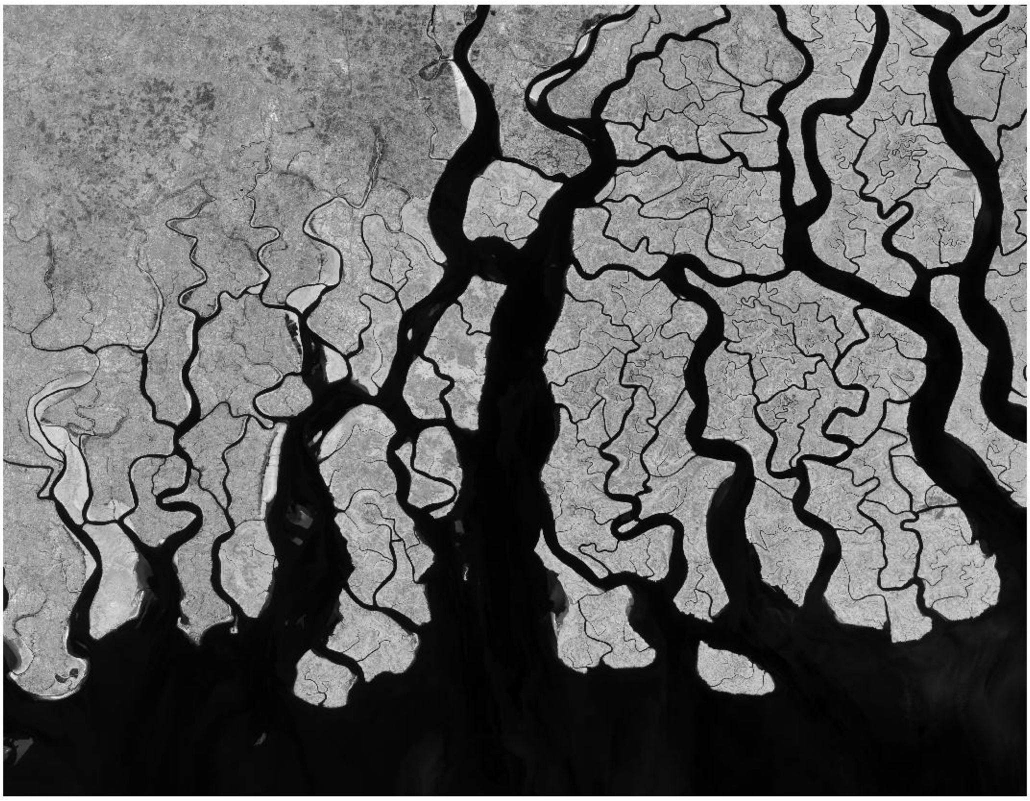


Figure 6.1: NIR Image : Sundarbans

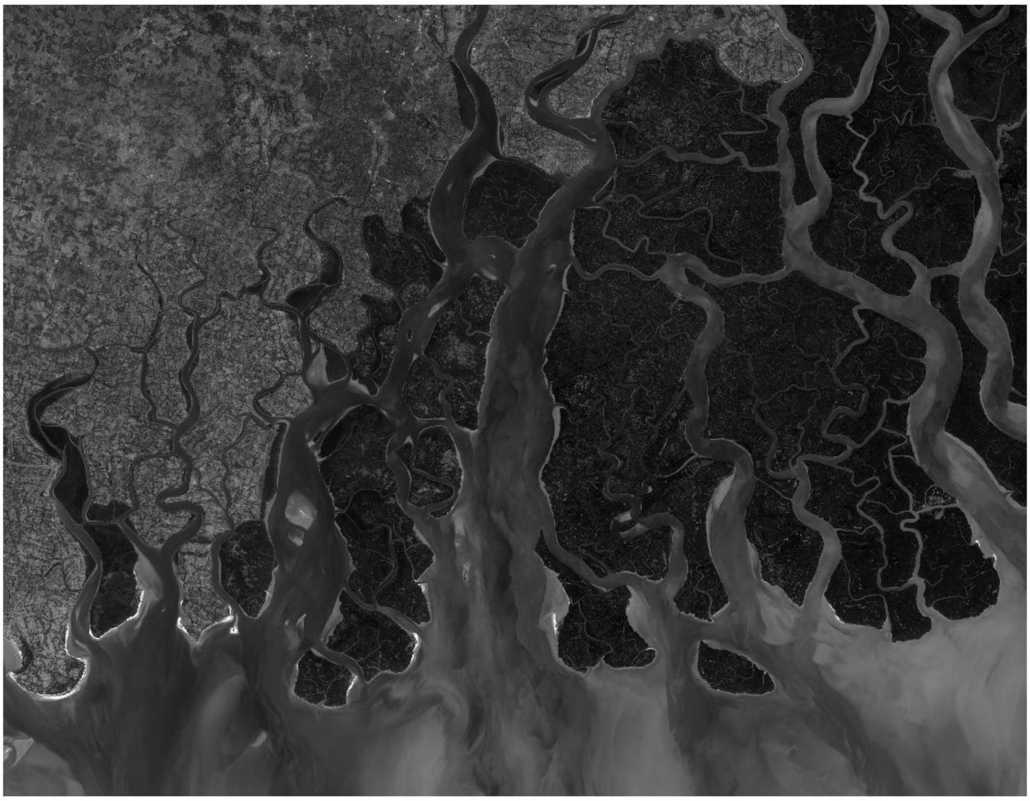


Figure 6.2: RED Image : Sundarbans

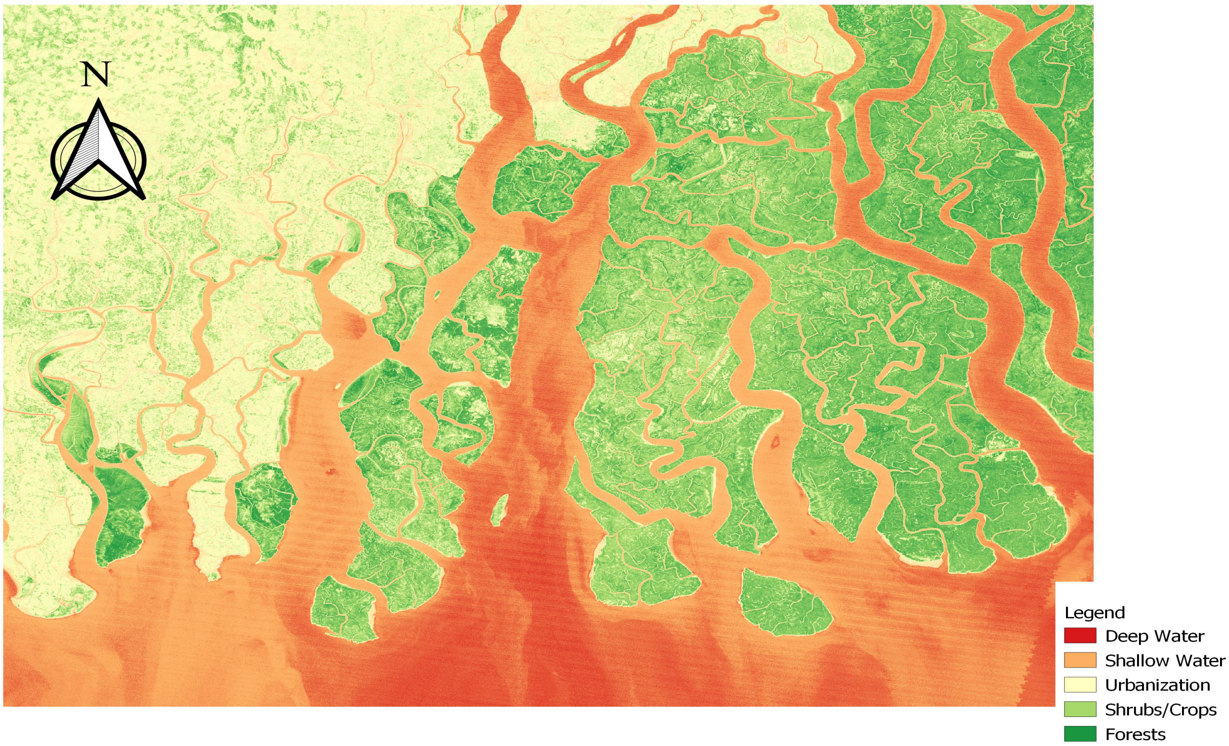


Figure 6.3: Sundarbans-1988

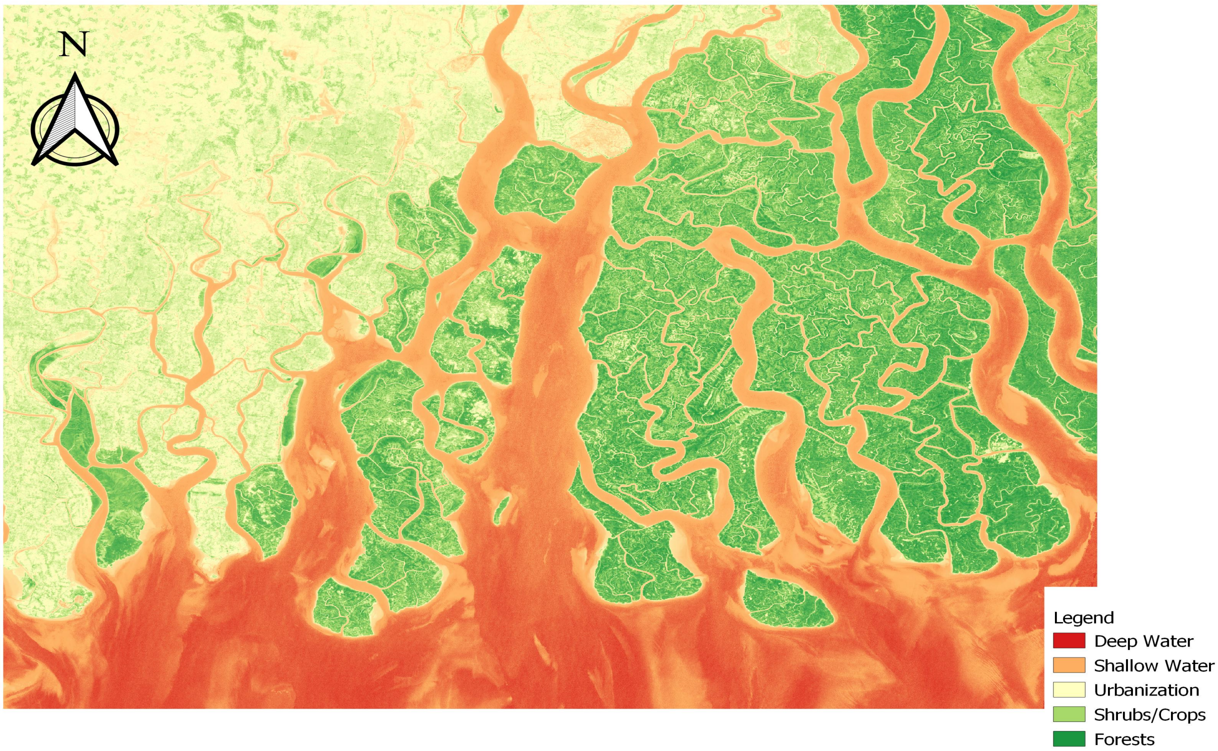


Figure 6.4: Sundarbans-2000

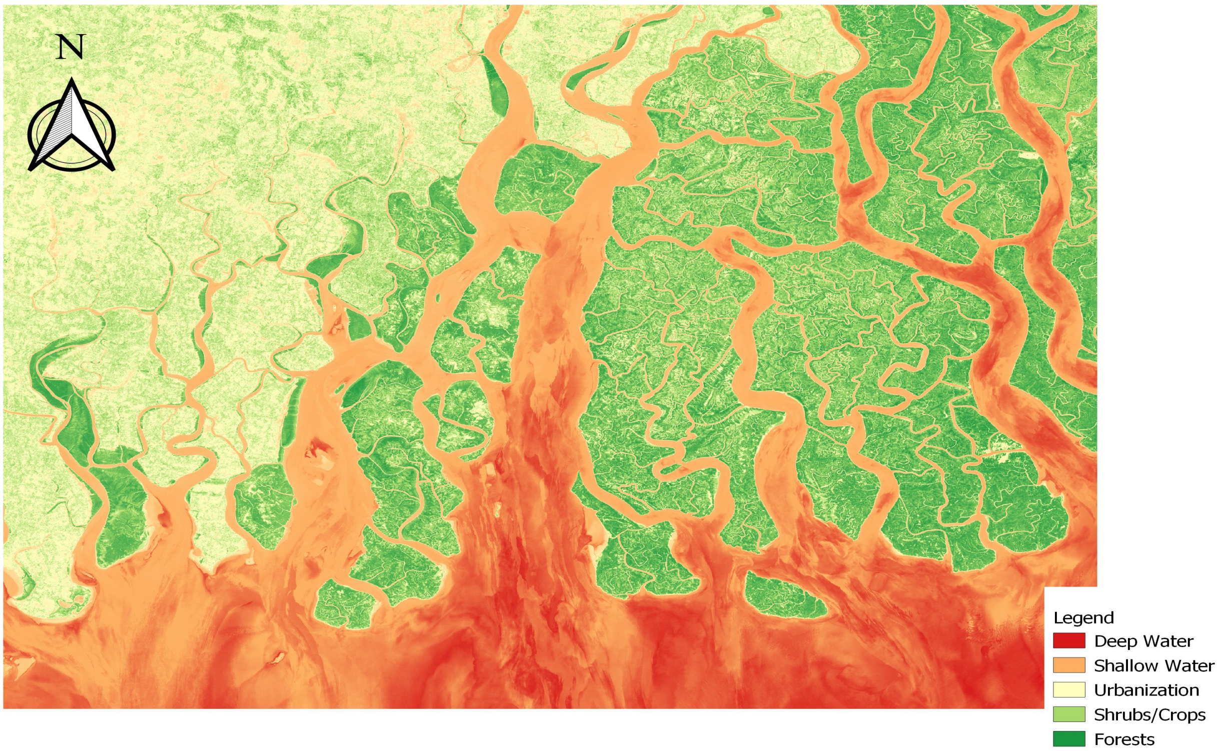


Figure 6.5: Sundarbans-2019

The Global trend has influenced the country policies regarding conservation of forests and wildlife. The amount of healthy vegetation increased to 7970px out of the 50176px. The mangroves have been under conserved and the area of the forests have increased to 1132.2177 square kilometers. The percentage of the vegetation cover increased about .8% to 15.840%. The upcoming decade has started to sense the consequences of the loss of these biodiversities and so the measures to facilitate the wildlife are taken and the drastic loss in only a small time gap would take a comparatively large time to recover.

CHAPTER 7

CONCLUSION

In this paper, we have discussed about a proposed model which takes commercially available data gathered by Forest Survey of India and predicts the deforestation or primary forest cover loss.The extent of deforestation in the mangroves of Sundarbans has been calculated. K Means Clustering has been applied on the Landsat Satellite Imagery. As can be seen from the NDVI results and calculation, the mangroves and the freshwater swamp forests have dramatically decreased and the water cover's reach has also taken over the parts of the delta which were not flooded in the past years. To validate the future forest cover percentage we use satellite images and classification algorithm to compare the accuracy of the proposed model. Further improvements are possible, where we can improve the accuracy of the model with the help of an improved dataset.

The loss of vegetation has left the whole region prone to high tides and flash flooding. The native species of the area are very affected and are endangered due to the same. The inferences from this case study calls for the authorities to invest resources and elevate their efforts in devising an action plan to conserve the Sundarbans.

Year	Vegetation Area(Sq Kilometer)	Area Affected(Hectares)	Area Deforested%
1988	1216.317	121631.7	NA
2000	1073.262	107326.2	11.7 %
2019	1132.217	113221.7	-5.49%

Figure 7.1: Sundarbans Over the Decades

CHAPTER 8

FUTURE ENHANCEMENT

The methodology employed in this project can be further used to analyse flood patterns over the years to build sustainable housing. This can be accomplished by finding out the area flooded every time after a flood using NDWI (Normalized Difference Water Index) and clustering. These flood prone areas can then be demarcated, so that villagers don't build settlements in those parts. Due to global warming, the rising sea levels will pose an enormous problem for the villages. Therefore , we can use this technology to protect lives and build homes for the future.

APPENDIX A

SOURCE CODE

Listing A.1: Source Code to extract and export required Landsat Satellite Images of NIR and RED spectrum of the Sundarbans from the Google Earth Engine

```
var geometry =
    /* color: #d63000 */
    /* shown: false */
    /* displayProperties: [
      {
        "type": "rectangle"
      }
    ] */
    ee.Geometry.Polygon(
        [[[88.23215756023906, 22.164375317717916],
          [88.23215756023906, 21.47082196140555],
          [89.12754330242656, 21.47082196140555],
          [89.12754330242656, 22.164375317717916]]], null,
            false);
/// Geometry has been drawn

/// Define bands of interest here

var L8bands = ['B2','B3','B4']

var L7bands =['B3']

var L5bands = ['B3']
```

```
///////// Visualisation parameters
var L8vis = {
  bands: L8bands,
  min: 0,
  max: 0.5,

};



var L7vis = {
  bands: L7bands,
  min: 0,
  max: 0.5,
  //gamma: [0.95, 1.1, 1]
};



var L5vis = {
  bands: L5bands,
  min: 0,
  max: 0.5,
  //gamma: [0.95, 1.1, 1]
};

//mask clouds
function maskLandsatclouds(image) {
  var qa = image.select('BQA')
  var cloudBitMask = ee.Number(2).pow(4).int()
  var mask = qa.bitwiseAnd(cloudBitMask).eq(0)
  return image.updateMask(mask)
      .select("B.*")
      .copyProperties(image, ["system:time_start"])
```

```
}


//Landsat 5 TM, available from 1984-2012
//1985-1986
var L5 = ee.ImageCollection("LANDSAT/LT05/C01/T1_TOA")
  .filterDate('1985-01-01', '1989-12-30') // you can change
     date here
  .filter(ee.Filter.lt("CLOUD_COVER", 0.5))
  .filterBounds(geometry)
  .map(maskLandsatclouds)
  .select(L5bands)

var mosaic_L5 = L5.median().clip(geometry); // here we are
   taking the median at each pixel in the collection
Map.addLayer(mosaic_L5, L5vis, "mosaic_L5")




//Landsat 7, available from 1999.
//Nb striping is due to the scan line corrector
// on the satellite failing, median compositor smooths that
   out but be aware
var L7 = ee.ImageCollection("LANDSAT/LE07/C01/T1_TOA")
  .filterDate('2000-01-01', '2001-12-30') // you can change
     date here
  .filter(ee.Filter.lt("CLOUD_COVER", 0.5))
  .filterBounds(geometry)
  .map(maskLandsatclouds)
  .select(L7bands)
```

```
var mosaic_L7 = L7.median().clip(geometry); // here we are
   taking the median at each pixel in the collection
Map.addLayer(mosaic_L7, L7vis, "mosaic_L7")



//Landsat 8. Avalable from 2013
////2017-2018
var L8 = ee.ImageCollection("LANDSAT/LC08/C01/T1_TOA")
   .filterDate('2019-01-01', '2020-01-01') // you can change
      date here
   .filter(ee.Filter.lt("CLOUD_COVER", 0.5))
   .filterBounds(geometry)
   .map(maskLandsatclouds)
   .select(L8bands)


var best = ee.Image(L5.sort('CLOUD_COVER').first());

//print('Least Cloudy Image is ', best);



//get specific metadata from image

var date_taken = best.get('DATE_ACQUIRED');
print('Date taken is', date_taken);


var mosaic_L8 = L8.median().clip(geometry); // here we are
   taking the median at each pixel in the collection
Map.addLayer(mosaic_L8, L8vis, "mosaic_L8")
```

```
Map.centerObject(geometry)

//Print the area of the selected mosaic
// Print polygon area in square kilometers
print('Polygon area in square km : ', geometry.area(10).divide
   (1000 * 1000));

// Print polygon perimeter length in kilometers.
print('Polygon perimeter: ', geometry.perimeter(10).divide
   (1000));

//// For Exporting mosaic images to drive



Export.image.toDrive({
  image: mosaic_L5, /// To choose which imagery to export
     change the number (e.g. L5, L7 or L8)
  description: 'L_RED_band',  //give correct name
  scale: 30,
  maxPixels: 1e13,
  region: geometry
});

Export.image.toDrive({
  image: mosaic_L8, /// To choose which imagery to export
     change the number (e.g. L5, L7 or L8)
  description: 'L7_RED_band',  //give correct name
  scale: 30,
  maxPixels: 1e13,
  region: geometry
});
```

```
Export.image.toDrive({
  image: mosaic_L8, /// To choose which imagery to export
     change the number (e.g. L5, L7 or L8)
  description: 'L8_RGB_new',  //give correct name
  scale: 30,
  maxPixels: 1e13,
  region: geometry
});
```

Listing A.2: Python code to compare and recognise a suitable machine learning algorithm

```
import cv2
import matplotlib.pyplot as plt
from google.colab.patches import cv2_imshow
from sklearn.cluster import KMeans
from collections import Counter
from sklearn import preprocessing
img = cv2.imread("/content/NDVI Sundarbans 2019.jpg")
img = cv2.cvtColor(img,cv2.COLOR_BGR2RGB)
#cv2_imshow(img)
img = cv2.GaussianBlur(img,(5,5),0)
img= cv2.resize(img,(224,224))
mod_img = img.reshape(img.shape[0]*img.shape[1],3)
classifier =KMeans(n_clusters = 5)
labels = classifier.fit_predict(mod_img)
count = Counter(labels)
center_color = classifier.cluster_centers_
count.keys()
ordered_colors = [center_color[i] for i in count.keys()]
def RGB2HEX(color):
        return "#{:02x}{:02x}{:02x}".format(int(color[0]), int(
            color[1]), int(color[2]))
hex_colors = [RGB2HEX(ordered_colors[i]) for i in count.keys()
    ]
rgb_colors = [ordered_colors[i] for i in count.keys()]

if (1):
        plt.figure(figsize = (8, 6))
        plt.pie(count.values(), labels = hex_colors, colors =
            hex_colors)
maxi = 0
rgb_colors[1][1]
```

```
for i in range(5):
  if rgb_colors[i][1]>rgb_colors[i][0] and rgb_colors[i][1]>
     rgb_colors[i][2]:
    s = abs(rgb_colors[i][1]-rgb_colors[i][0])+abs(rgb_colors[
       i][1]-rgb_colors[i][2])
    if s>maxi:
      maxi = s
      c=i
ctr=0

for i in count.keys():
  if c==ctr:
    break
  ctr+=1
idx =i

su = 0  #sumation variable
for i in count:
  su = su + count[i]          #counting the sum of pixels

per = (count[idx]/su) * 100
```

APPENDIX B

SCREENSHOTS

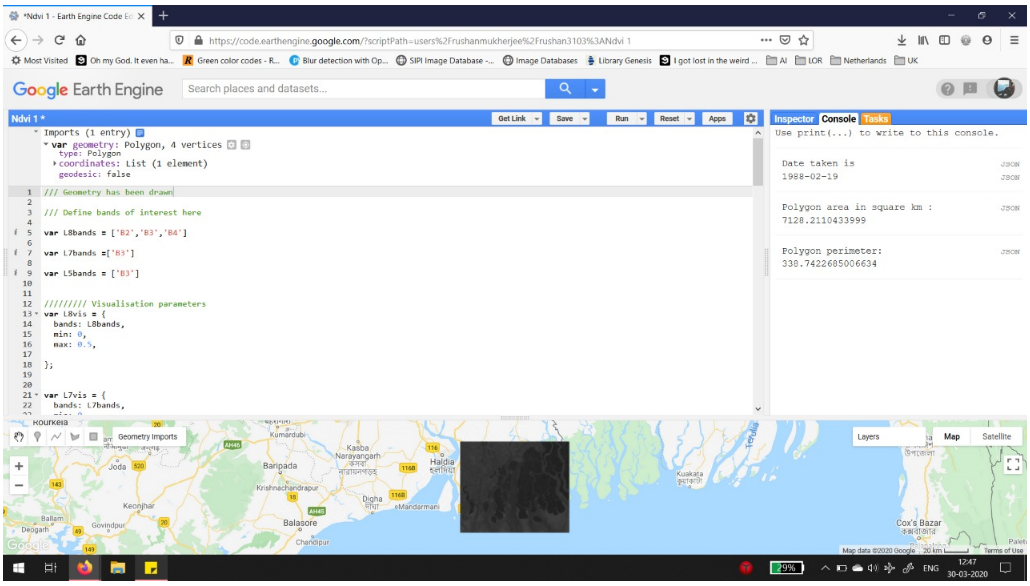


Figure B.1: Google Earth Engine : Importing Landsat Images

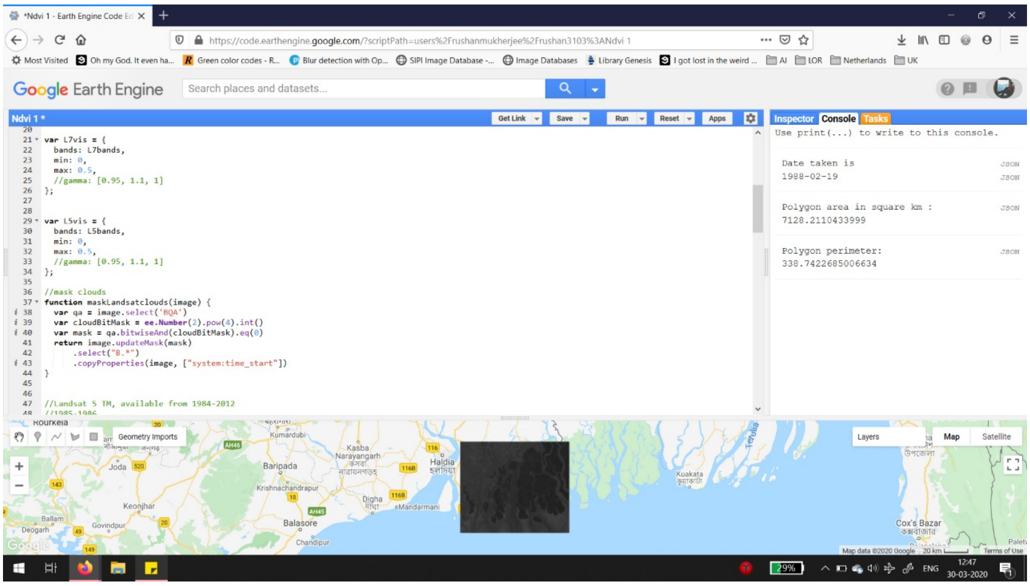


Figure B.2: Google Earth Engine : Searching Images with least Cloud Cover

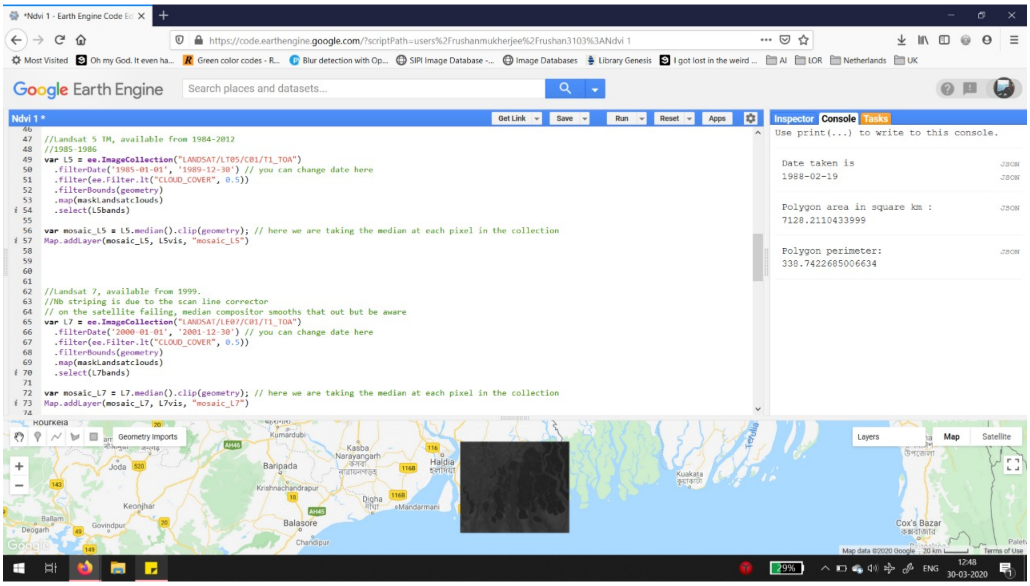


Figure B.3: Google Earth Engine : Noise(Cloud Cover) Filtering

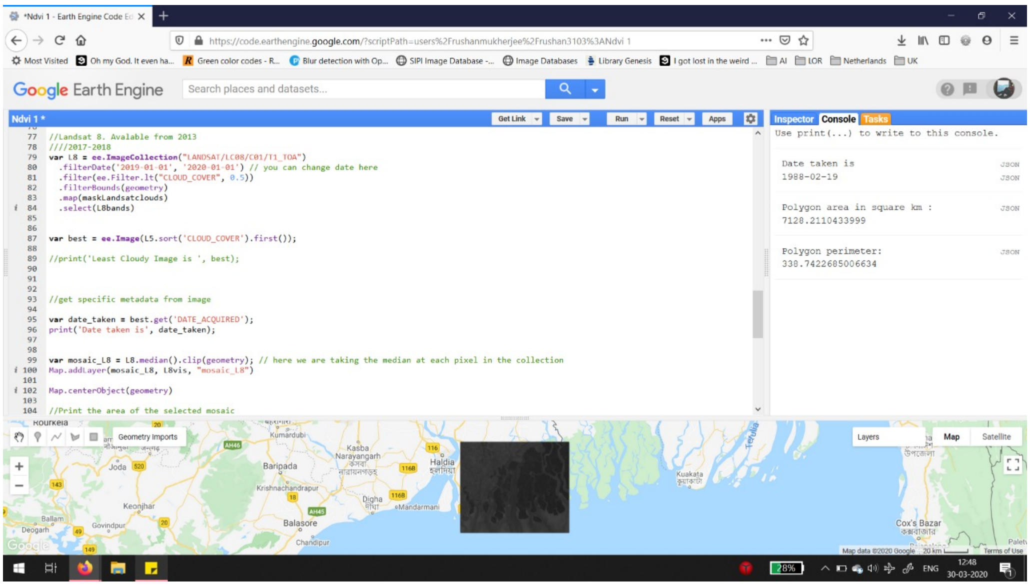


Figure B.4: Google Earth Engine : Creating Polygon over Region Of Interest

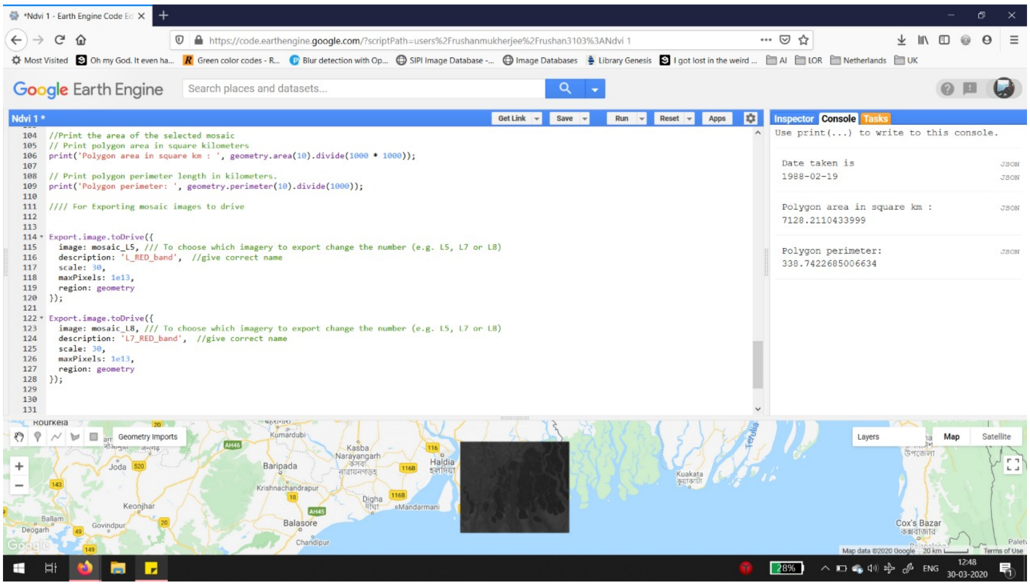


Figure B.5: Google Earth Engine : Extracting images of NIR and RED spectrum

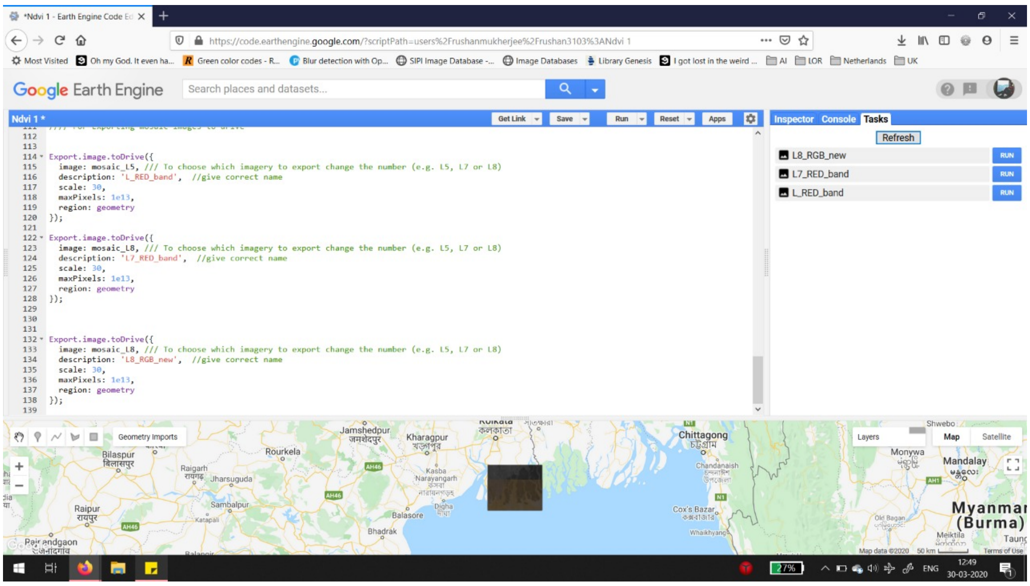


Figure B.6: Google Earth Engine : Exporting Satellite Images

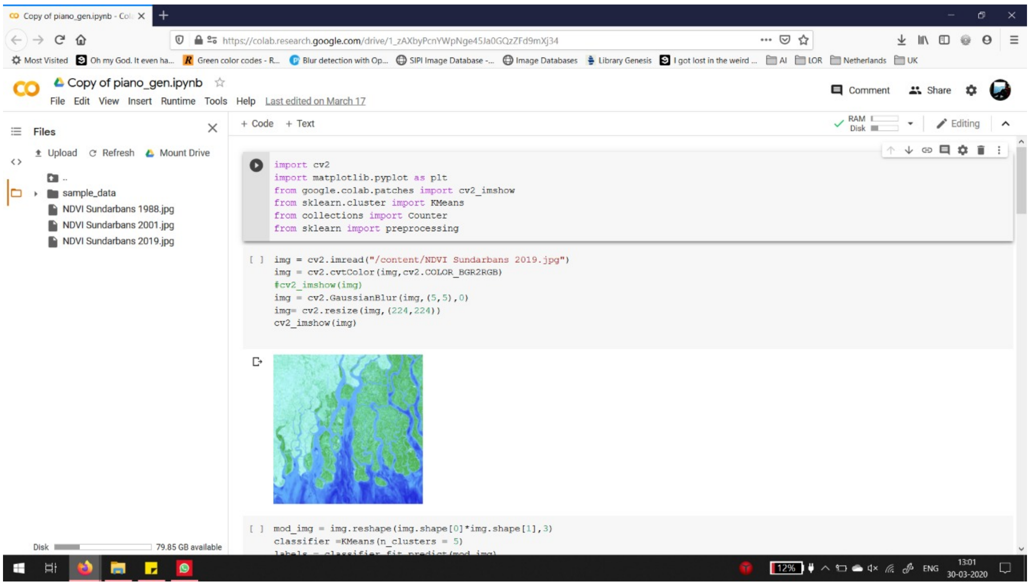


Figure B.7: Importing Libraries

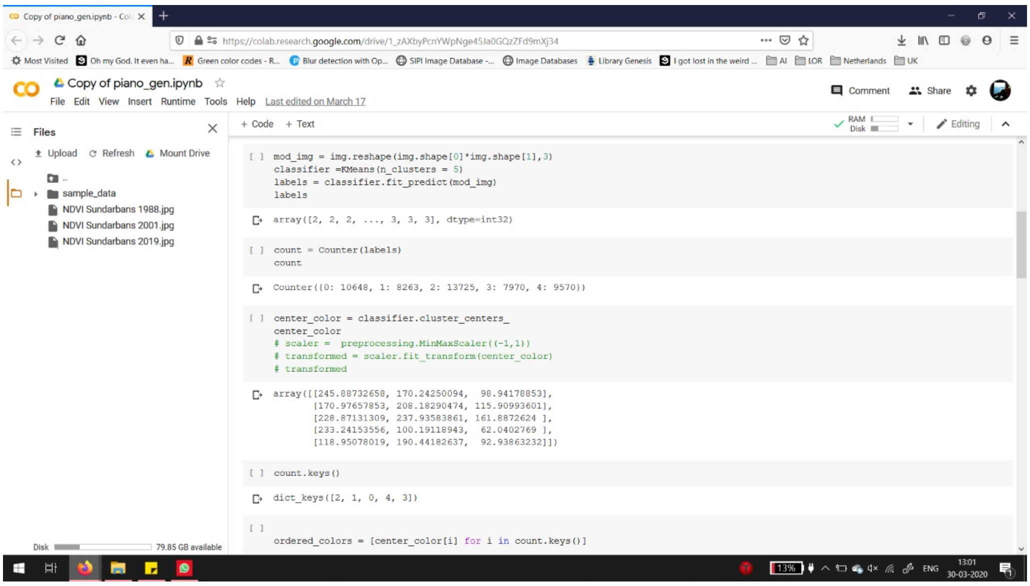


Figure B.8: K-Means Clustering

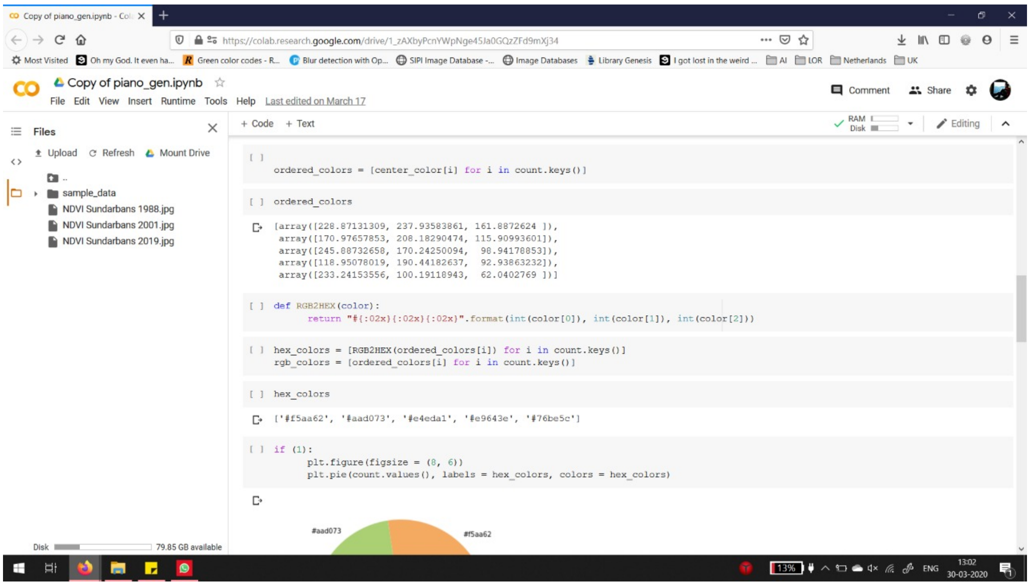


Figure B.9: Ordering Clusters

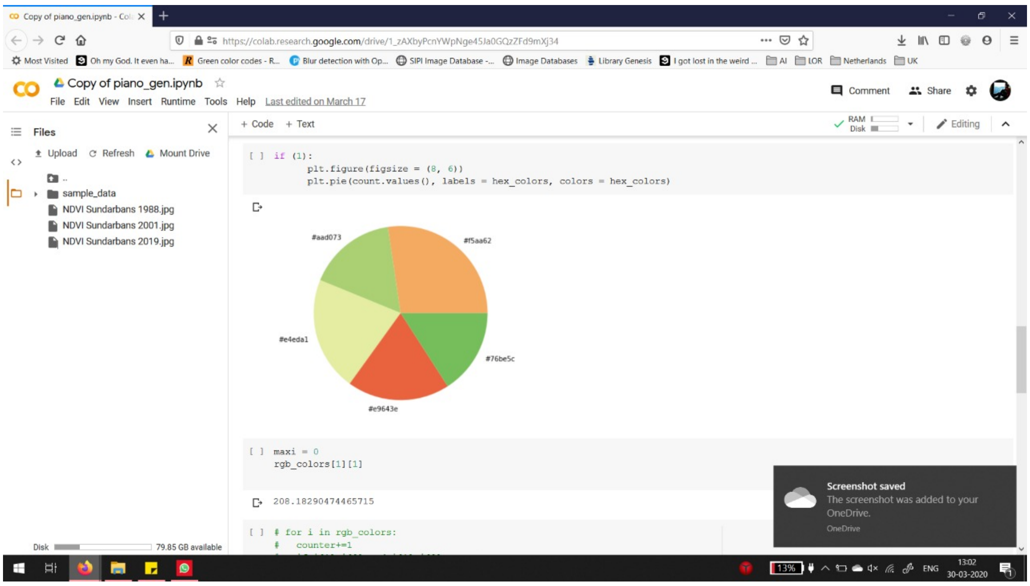


Figure B.10: Plotting Colors in Pie Chart

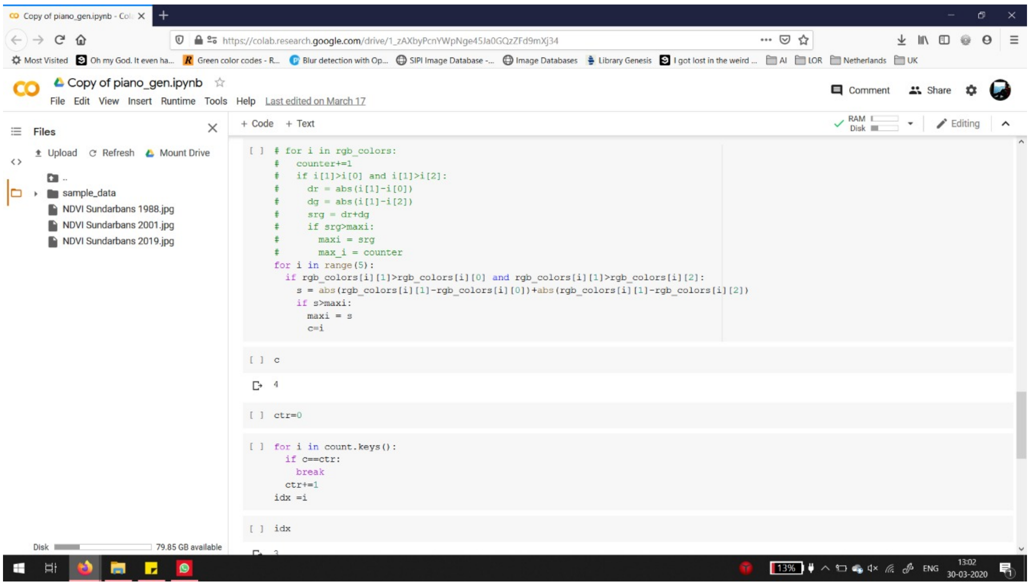


Figure B.11: Filtering Out Forest Vegetation

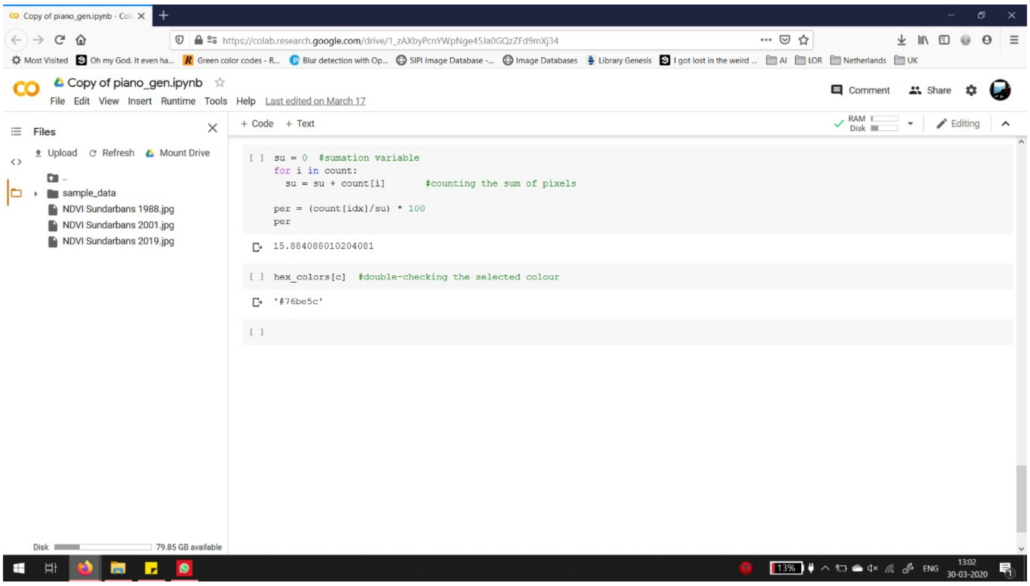


Figure B.12: Calculating Proximity of the Healthy Vegetation

APPENDIX C

GRAPHS

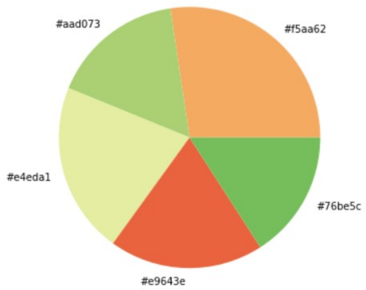


Figure C.1: Range of Vegetation

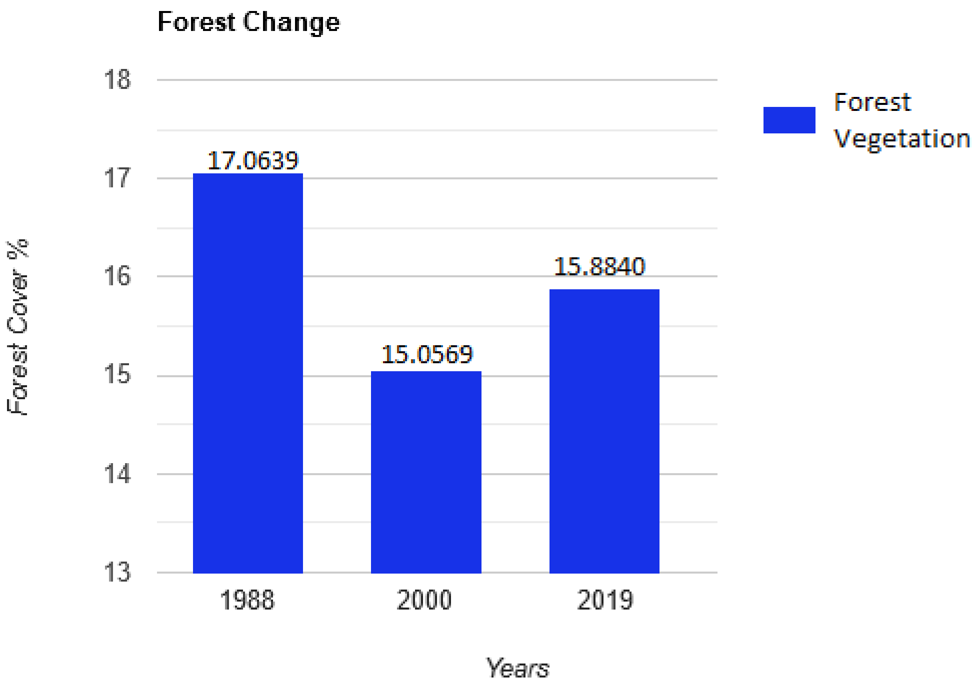


Figure C.2: Forest Change

REFERENCES

[1]Shah, S., Sharma, D. Land use change detection in Solan Forest Division, Himachal Pradesh, India.For.Ecosyst.2, 26 (2015) doi:10.1186/s40663-015-0050-7

[2]Sahana, Mehebub & Hong, Haoyuan & Sajjad, Haroon & Liu, Junzhi & Zhu, A-Xing. (2018). Assessing deforestation susceptibility to forest ecosystem in Rudraprayag district, India using fragmentation approach and frequency ratio model. Science of The Total Environment. 627. 1264-1275. 10.1016/j.scitotenv.2018.01.290.

[3]Zurqani, Hamdi & Post, Christopher & Mikhailova, Elena & Schlautman, Mark & Sharp, Julia. (2018). Geospatial analysis of land use change in the Savannah River Basin using Google Earth Engine. International Journal of Applied Earth Observation and Geoinformation. 69. 175-185. 10.1016/j.jag.2017.12.006.

[4]Sekertekin, A. , Marangoz, A. M., and Akcin, H. Pixel-based classification analysis of land use land cover using sentinel-2 and landsat-8 data, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-4/W6, 91-93, https://doi.org/10.5194/isprs-archives-XLII-4-W6-91-2017, 2017.

[5]Lv, Zhiyong, Tongfei Liu, Yiliang Wan, JónAtliBenediktsson and Xiaokang Zhang. "Post-Processing Approach for Refining Raw Land Cover Change Detection of Very High-Resolution Remote Sensing Images." Remote Sensing 10 (2018): 472.

[6]M.D. Behera, P. Tripathi, P. Das, S.K. Srivastava, P.S. Roy, C. Joshi, P.R. Behera, J. Deka, P. Kumar, M.L. Khan, O.P. Tripathi, T. Dash, Y.V.N. Krishnamurthy, Remote sensing based deforestation analysis in Mahanadi and Brahmaputra river basin in India since 1985, Journal of Environmental Management,Volume 206,2018,Pages 1192-1203, ISSN03014797, https://doi.org/10.1016/j.jenvman.2017.10.015

[7]Guigonan Serge Adjognon, Alexis Rivera-Ballesteros, Daan van Soest,Satellite-based tree cover mapping for forest conservation in the drylands of Sub Saharan Africa (SSA): Application to Burkina Faso gazette forests,DevelopmentEngineering, Volume 4, 2019, 100039, ISSN 2352-7285, (https://doi.org/10.1016/j.deveng.2018.100039.)

[8]Quy Van Khuc, BaoQuang Tran, Patrick Meyfroidt, Mark W. Paschke, Drivers of deforestation and forest degradation in Vietnam: An exploratory analysis at the national level,Forest Policy and Economics, Volume 90, 2018, Pages 128-141,ISSN 1389-9341, https://doi.org/10.1016/j.forpol.2018.02.004.

[9]Shahzad, N., Saeed, U., Gilani, H. et al. Evaluation of state and community/private forests in Punjab, Pakistan using geospatial data and related techniques. For. Ecosyst.2, 7 (2015) doi: 10.1186/s40663-015-0032-9

[10]Rodrigo Antônio de Souza, Paulo De Marco, Improved spatial model for Amazonian deforestation: An empirical assessment and spatial bias analysis,Ecological Modelling,Volume 387, 2018, Pages 1-9,ISSN 0304-3800,https://doi.org/10.1016/j.ecolmodel.2018.08.015.

Format - I

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University u/s 3 of UGC Act, 1956)

Office of Controller of Examinations

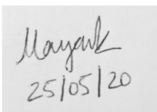
REPORT FOR PLAGIARISM CHECK ON THE DISSERTATION/PROJECT REPORTS FOR UG/PG PROGRAMMES

(To be attached in the dissertation/ project report)

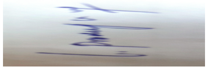
1	Name of the Candidate **(IN BLOCK LETTERS)**	Mayank Dewangan | RushanMukherjee
2	Address of the Candidate	H.no-22 ,Hospital Colony,Ganj Road,Gobra Nawapara Raipur - 493881 E-301, Smarana Apartments, Bakeri City , Vejalpur, Ahmedabad -380051 **Mobile Number :** 8602279136 / 87586 08200
3	Registration Number	RA1611003010417 / RA1611003011168
4	Date of Birth	30 April 1999 / 31 March1998
5	Department	Computer Science & Engineering
6	Faculty	Dr. G. Niranjana / Mrs. R. Lavanya
7	Title of the Dissertation/Project	Deforestation Analysis Using Unsupervised Clustering and Satellite Images
8	Whether the above project/dissertation is done by	~~Individual~~ or group : Group (Strike whichever is not applicable) a) If the project/ dissertation is done in group, then how many students together completed the project : 2 b) Mention the Name & Register number of other candidates : (Both have been Mentioned)
9	Name and address of the Supervisor / Guide	Dr. K. Pradeep Mohan Kumar pradeep.k@ktr.srmuniv.ac.in 9362447221 **Mail ID : Mobile Number :**
10	Name and address of the Co-Supervisor / Co- Guide (if any)	NOT APPLICABLE **Mail ID : Mobile Number :**

11	Software Used	TURNITIN		
12	Date of Verification	02-June-2020		
13	**Plagiarism Details: (to attach the final report from the software)**			
Chapter	**Title of the Chapter**	**Percentage of similarity index (including self citation)**	**Percentage of similarity index (Excluding self citation)**	**% of plagiarism after excluding Quotes, Bibliography, etc.,**
1	INTRODUCTION	0.1%	0%	0.1%
2	LITERATURE SURVEY	0%	0%	0%
3	DATA PREPROCESSING	0%	0%	0%
4	FEATURE SELECTION TECHNIQUES	3.2%	0%	3.2%
5	IMAGE SPECTUM PROCESSING	0.6%	0%	0.9%
6	RESULTS AND DISCUSSION	0%	0%	0%
7	CONCLUSION	0.5%	0%	0.9%
8	FUTURE ENHANCEMENT	0.9%	0%	0.9%
9	Nil	-	-	-
10	Nil	-	-	-
Appendices		0.7%	0%	0.7%

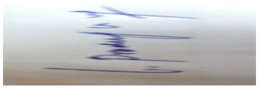
I / We declare that the above information have been verified and found true to the best of my / our knowledge.



Signature of the Candidate



Name & Signature of the Staff (Who uses the plagiarism check software)



Name & Signature of the Supervisor/Guide

(NOT APPLICABLE)

Name & Signature of the Co-Supervisor/Co-Guide

Name & Signature of the HOD

Project Report - Rushan Mukherjee Team

ORIGINALITY REPORT

6%	1%	1%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to SASTRA University Student Paper	2%
2	Submitted to University of Malaya Student Paper	1%
3	bookfrom.net Internet Source	1%
4	web.pdx.edu Internet Source	<1%
5	Submitted to CVC Nigeria Consortium Student Paper	<1%
6	Submitted to Heriot-Watt University Student Paper	<1%
7	serdp-estcp.org Internet Source	<1%
8	Submitted to University of Durham Student Paper	<1%
9	Submitted to Edith Cowan University Student Paper	<1%

International Journal of Innovative Technology and Exploring Engineering
ISSN: 2278-3075 (Online) | Exploring Innovation | A Key for Dedicated Services
Published by: Blue Eyes Intelligence Engineering and Sciences Publication
G18-19-20, Block-B, Tirupati Abhinav Homes, Damkheda, Bhopal (Madhya Pradesh)-462037, India
Website: www.ijitee.org Email: submit2@ijitee.org

CERTIFICATE

This certifies that the research paper entitled '**Deforestation Analysis using Unsupervised Clustering and Satellite Images**' authored by '**K. Pradeep Mohan Kumar, Rushan Mukherjee, Mayank Dewangan**' was reviewed by experts in this research area and accepted by the board of 'Blue Eyes Intelligence Engineering and Sciences Publication' which has published in '**International Journal of Innovative Technology and Exploring Engineering (IJITEE)**', ISSN: 2278–3075 (Online), Volume-9 Issue-6, April 2020, Page No. 1651-1656.

The Value of Citation (VoC): of IJITEE is 6.03 for year 2019. Your published paper and Souvenir are available at: https://www.ijitee.org/download/volume-9-issue-6/

Blue Eyes Intelligence Engineering and Sciences Publication BHOPAL

Jitendra Kumar Sen
(Manager)

Dr. Shiv Kumar
(CEO)

IJSCE IJEAT IJRTE IJITEE IJISME IJMH IJIES IJEANT IJESE IJBSAC