Geomatics & Geoinformation
Exercise 6
Machine Learning with Google Earth Engine

Mattia Manna, 1886817
Enrolled in Data Science

Academic year 2023/2024, Second semester

Contents

1 Analysis of the United Arab Emirates landscape conducted through K-means clustering

1.1 United Arab Emirates

The United Arab Emirates (UAE) are a confederation of hereditary monarchies (emirates), in each of which the emir is absolute ruler in his state.
Regarded as a single state the United Arab Emirate stretches across the southeastern part of the Arabian Peninsula, and shares borders with Oman and Saudi Arabia; as well as maritime borders in the Persian Gulf with Qatar and Iran [9].

Political map of the United Arab Emirates

Figure 1: Image source *nationsonline.org* [1]

The terrain is flat covered almost everywhere with a sandy blanket that makes the landscape even more monotonous.
From the coastal belt, that is fronted by numerous islands, the gradual transition to the desert environment of inland Arabia begins; the climate, therefore, is arid, with large temperature ranges, especially during the day.
Because of the extreme scarcity of rainfall, surface hydrography is absent; groundwater feeds the vegetation of the oases.

The population is heavily concentrated in the most important cities, like Abū Dhābi and Dubai.
In the rest of the country, permanent settlement is limited to small towns, in oases or along the coast, while nomadism is still widespread.

To be more specific the United Arab Emirates covers an area of 83600 km^2, most of the surface is mountainous and barren desert covered with loose sand and gravel, the cities are few and the biggest are:

Nº	**City**	**Metro Area [km^2]**	**Emirate**
1	Dubai	1610	Dubai
2	Abu Dhabi	972	Abu Dhabi
3	Sharjah	235.5	Sharjah
Total area:		2772.5	

The extensions of the others cities weren't available.
To summarize the United Arab Emirate are covered only by desert, mountains and cities, and the extensions of these biomes are just partially know so **the purpose of the analysis will be to understand how the area is distributed among those three biomes**.
Since the landscape surface has very pronounced differences the *K-means clustering* will be the perfect methodology to carry on the analysis.

Of course since the methodogy used will not garantee measurement precise to the meter the results will be given in percentages, just to get an idea of how the territory is distributed.

1.2 Introduction to the K-means clustering

The *K-means clustering* is a machine learning procedure, more specifically an unsupervised methodology.
Unsupervised machine learning models are given unlabeled data and allowed to discover patterns and insights without any explicit guidance or instruction [4].
They should be thought as black boxes and the process to use those models is simple:

1. The user feed the model with data;
2. The model give back an answer, in this case a classification;
3. The user analyze the answer and draw is conclusions.

A very important concept to keep in mind is that those models do not always provides understandable and usable answers, sometimes they just gave the best output in mathematical terms but the results are useless for practical purposes, so the user must always be skeptical and rigorous when dealing with them.
Of course to understand better the behaviour of those algorithms some metrics are used, however those will not be covered in this report.

Going back to the *K-means clustering* is a non hierarchical cluster analysis methodology, the peculiarities of this type of cluster techniques is that they don't possess tree-like structures and so the number of clusters must be decided a priori, they are also quicker respect to hierarchical cluster analysis methodologies.

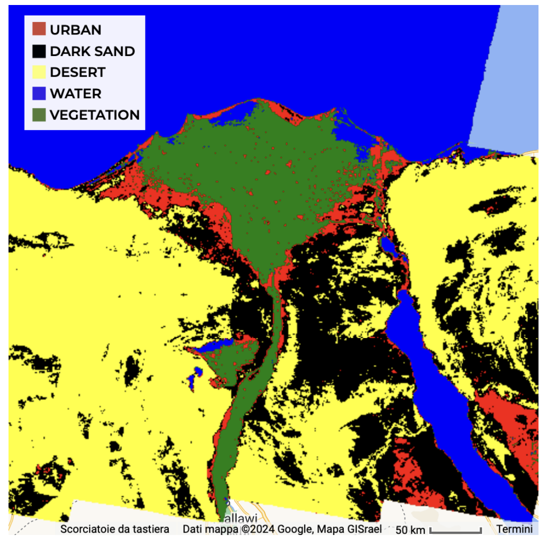
More in details the K-means algorithm aims to partition **n** observations into **k** parts (clusters).
To distinguish the **n** units and assign them to the clusters it look at all the features for each unit, after that it assign the unit to the cluster that have more similar features (less distance).

In the context of Google Earth Engine: the units to classify are the pixels and the features used to classify are the bands.
In this report the following bands will be used :

- SR_B1: Band 1 (blue) surface reflectance
- SR_B2: Band 2 (green) surface reflectance
- SR_B3: Band 3 (red) surface reflectance
- SR_B4: Band 4 (near infrared) surface reflectance
- SR_B5: Band 5 (shortwave infrared 1) surface reflectance
- SR_B7: Band 7 (shortwave infrared 2) surface reflectance
- ST_B6: Band 6 surface temperature.

So basically each pixel of the median image will be assigned to the nearest cluster by looking to all the bands.
Practically this procedure is a quick way to make a classification of the image, in fact with a correct use of the kmeans algorithm it is possible to distinguish among different types of areas, for instance: water, sand , green areas and urban areas.

An example of what has been said can be taken directly from the lectures of the course, let's have a look on the clustering process taken in the area of Nile delta.



It can be noted how the different areas have been identified. However it must be underline that this technique has several limitations, like satellite resolution that prevent to have more accurate clusters, and also the conformations of the object. Infact sometimes areas that are differents but results very similars by looking at the bands are misclassified.
For instance in this report sometimes buildings are identified as mountains and viceversa, but despite this the procedure perform well, and has proven to be useful.

1.3 Defining the pipeline

As has been said before the tool to find out in which percentage the three biomes are distributed will be the K-means.
The K-means algorithm will be applied using Google Earth Engine, therefore a pipeline must be defined.
To obtain the results sought the steps are multiple:

1. Define the area of interest (ROI);
2. Define a traning area;
3. Take the LANDSAT Collection, filter it as necessary;
4. Apply radiometric scaling and cloud masking to the collection;
5. Select the median image, cut it out in the roi;
6. Sample from the latter image in the training area and define a traning set;
7. Define the cluster algorithm (kmeans) and train it on the traning set;
8. Apply trained algorithm on the roi;
9. Visualize the result and choose the correct number of clusters;
10. Draw the conclusions by counting the pixels for each type land and so obtain the land type extension in km^2.

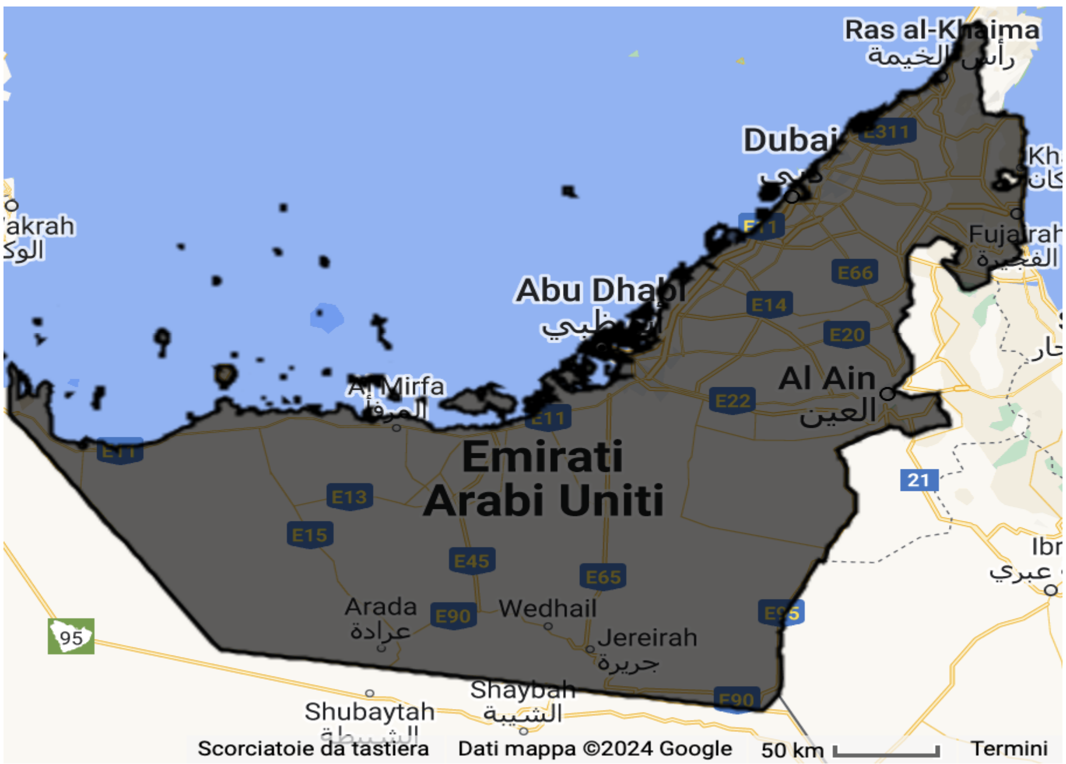
The Google Earth Engine script that implements the described pipeline will be available at the following link: `https://code.earthengine.google.com/b6235350f45f592e55705b81f8c601a3`.

1.4 Define the parameters

1.4.1 Define the area of interest (ROI)

As ROI has been selected the entire country of United Arab Emirates, this can be done using the dataset from the FAO called *FAO GAUL: Global Administrative Unit Layers 2015, Country Boundaries*.
This dataset divide the entire Map using the countries boundaries, and it is also possible using the proper code to select only a specified country, UAE[1] has code 255[2].

Then using this geometry together with the function clip has been possible to select only the area of the United Arab Emirates.



The total extension of this area is of 78233 km^2, not to much different from the 83600 km^2 indicated by the literature.

[1]United Arab Emirates
[2]The GAUL code of each country can be found at the following link: `https://www.fao.org/nocs/en`

1.4.2 Define a traning area

An area within the roi, specifically the most heterogeneous area, was selected as the training area. Thus, this area contains urban areas, sandy areas and mountainous areas.

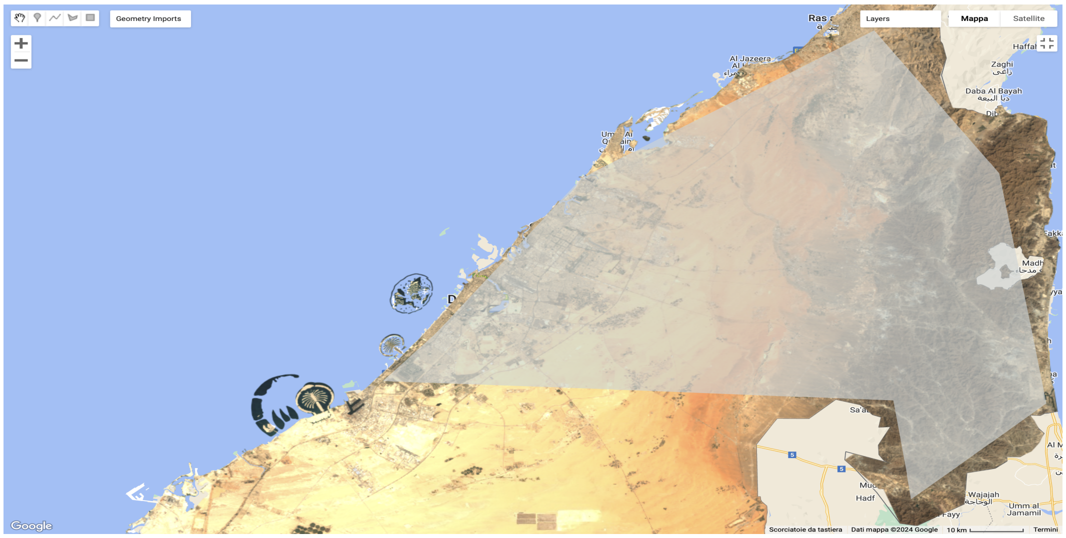


Figure 2: Scale 10 km

The total extension of this training area is of 7506 km^2, that is the the 9.6% of the region of interest.

1.4.3 Filter the collection

The LANDSAT Collection used is the *USGS Landsat 7 Level 2, Collection 2, Tier 1*, this collection is made up of images of all the world, in a time span that goes from **1999-05-28** to **2024-01-19** [3], dates are in yyyy/mm/dd format.

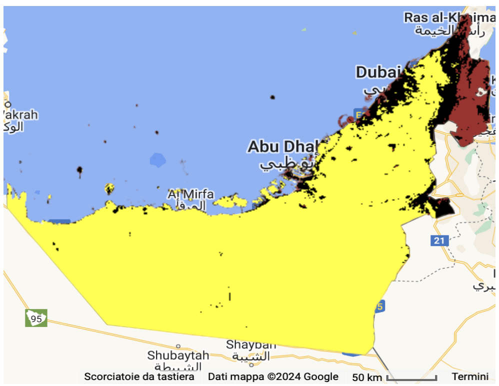
It goes without saying that work with all those images is computationally impossible, and also useless for the purpose of the analysis.
So what is usually done is filter the collection by region of interest and the time span desired.

Of course as **region of interest** the entire area described in 1.4.2 has been selected, and as **time interval** all the 2023 has been selected.
The entire year has been selected to minimize seasonals effects and to have enough images to apply succesfully cloud masking[4].
The collection is also f**iltered by clouds percentage**, in particular only images with less than 50 % of cloudy pixels have been selected.

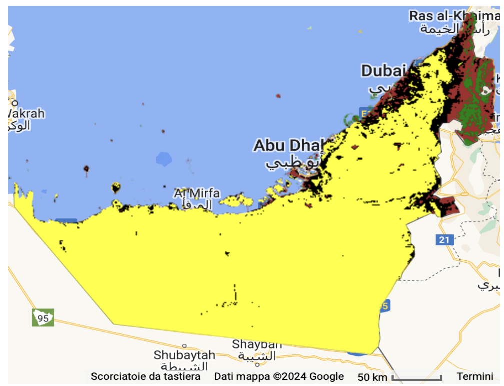
[3]Interval available at the day in which thee report has been written
[4]Process that consist in removing pixels that contains clouds, cloud shadow, cirrus, aerosols.
It is a fondumental step when working with certain type of band and images.

1.4.4 Number of clusters

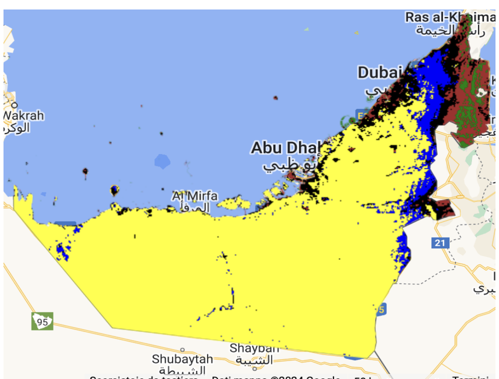
Since the *K-means clustering* is non hierarchical cluster analysis methodology a number of clusters must be choosen apriori.
Of course the choice cannot be random, an analysis must be carried out to choose the optimal number of clusters.
When dealing with images, the best way to understand how many cluster should be used is to apply many times the algorithm (choosing a different number of cluster) , visualize the results and then make a choice.
In this scenario since is required to distinct 3 different biomes the clusters number will start from 3.
By applying 4 times the algorithm specifing: 3,4,5,6 clusters the results are the following.



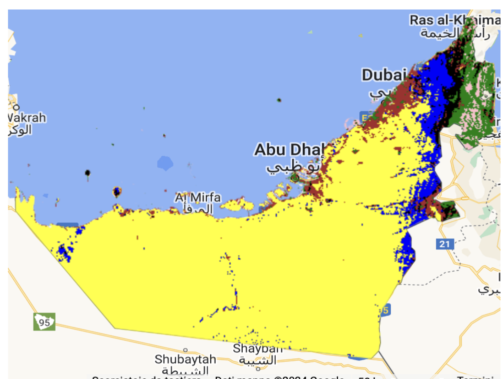
(a) 3 clusters



(b) 4 clusters



(c) 5 clusters



(d) 6 clusters

By looking at the clustered images below the best number of clusters is 3.
Despite occasional problems to distinguish cities and mountains area, errors commons also the others clusters, the *K-means clustering* perfomed with 3 clusters seems to well distinct the three biomes, given so it will be used to carry on the analysis.

Having a closer look to the *K-means clustering* (3 clusters) classification what has been said can be confirmed.

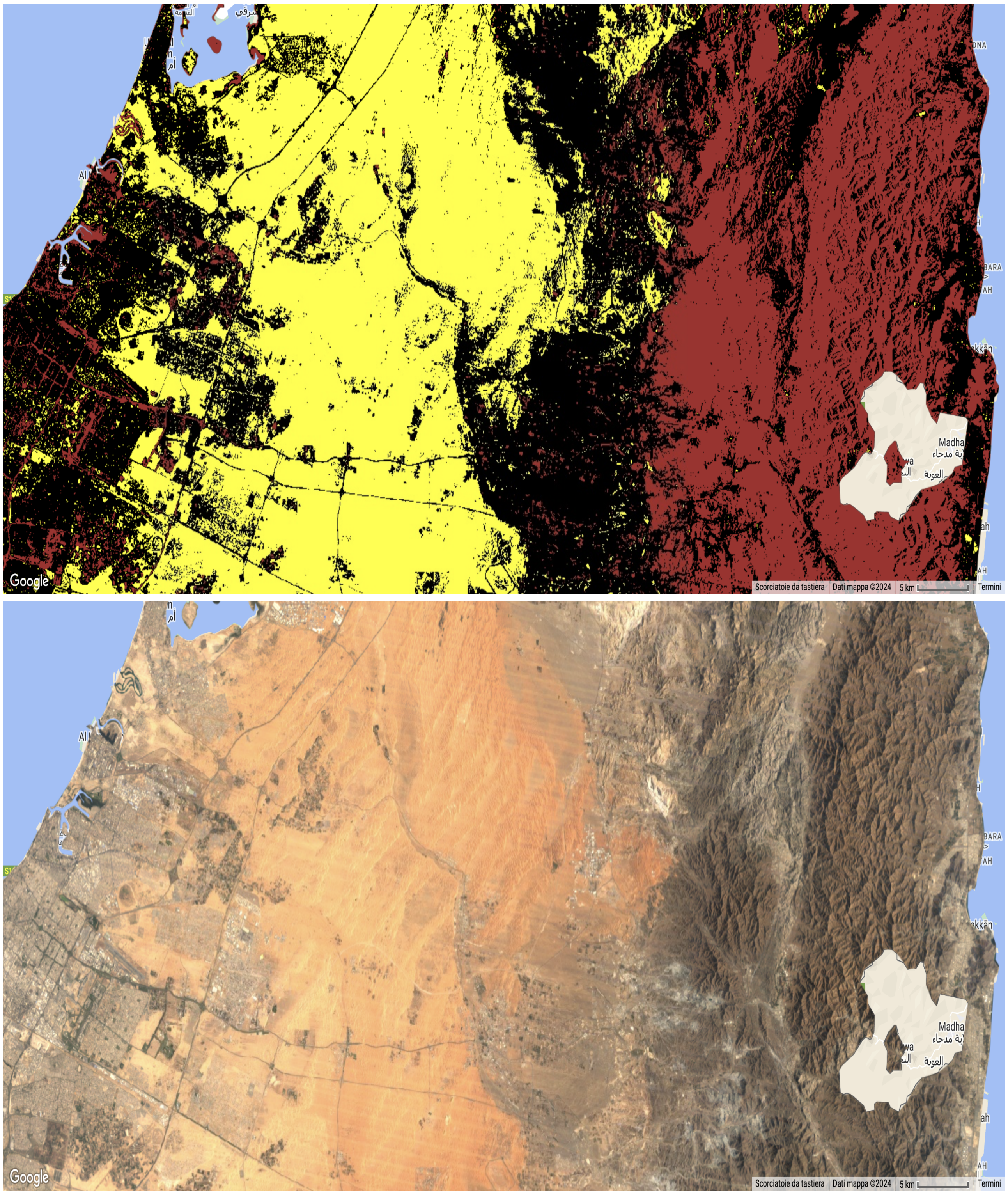
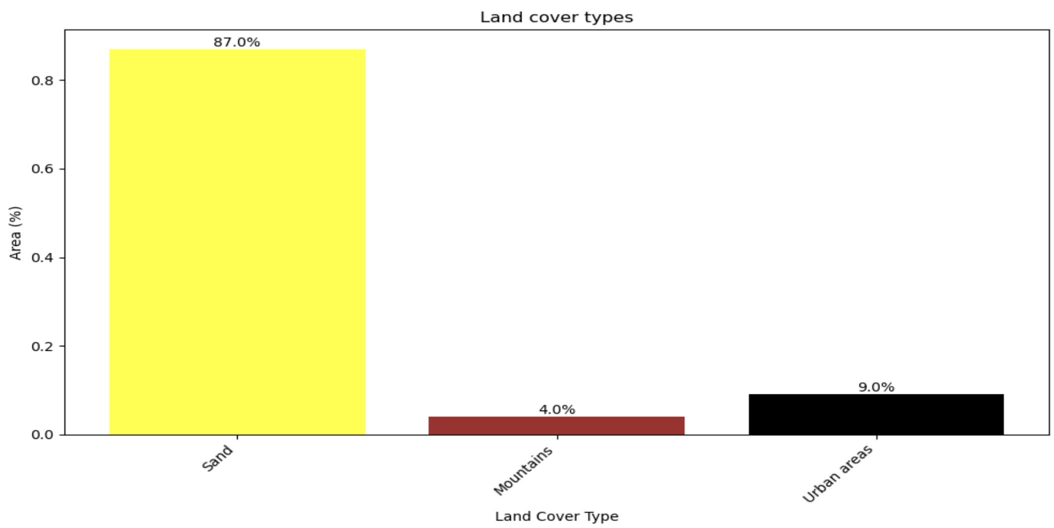


Figure 3: Scale is 5 km

1.5 Results

As expected there were several issue, in particular to distinct mountains and urban areas but the results seems credible.
It is clear that the dominant part of United Arab Emirates is desert (87 %), but there is also a not negligible part of mountains (4%).
For what concert the urban areas this value must be taken cautiously because has can been seen in 3 the urban areas values are inflated, infact a large portion of mountains has been identified as urban area.

For a better analysis maybe an image collection with better resolution can be considered as well as selecting differents bands.
In the end for what concern the attendibility of the analysis it seems credibile, infact the entire area of the analysis is 78233.42 km^2 and the one indicated by the literature is 80300 km^2.



To be more precise let's have a look to the table showing the surfaces for each land type.

Land type	**Extension [km^2]**	**Extension %**
Sand	68060.44	87
Mountains	3484.9	4
Urban areas	6688.08	9
Total:	78233.42	100

2 Analysis of green areas in Rome and Paris conducted through the random forest classifier

2.1 The importance of green areas

Urban green spaces provide environmental benefits through their effects on negating urban heat, offsetting greenhouse gas emissions, and attenuating storm water. They also have direct health benefits by providing urban residents spaces for physical activity and social interaction, and allowing psychological restoration to take place [6].
So it's clear that the presence of green areas in cities is very important, however how green are the capitals of Europe? A chart provided by *Statista* and using data from 2018 of the European Environment Agency provide this answer.

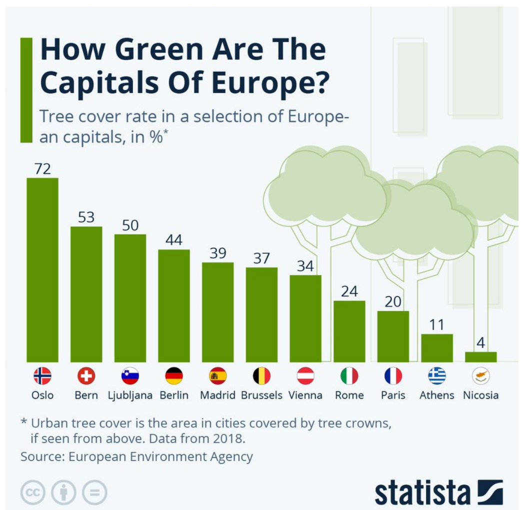


Figure 4: Source of the chart *Statista* [3], data used for the chart *European Environment Agency* [8]

It seems that Rome and Paris are not doing well in this on the latter chart, the percentage of green areas is far smaller respect to the others capitals of Europe.
However few considerations need to be made:

- The data are getting old, they are from 2018.
- The analysis is considering only the *tree cover* as green surface, but not considering also:
  - Shrubland [5]
  - Grassland [6]
  - Cropland [7]

Given those two considerations a new analysis will be carried out.
This new analiysis will be performed with the intent to use newer data, and also considering the other green surfaces like grassland, cropland, shrubland.
To perform this type of analysis a **random forest classifier** will be used, this classifier will be trained with data from 2021 and then applied to classify the most recent data available and obtain a new percentage of green areas in the cities.

[5] Arbusti
[6] Prato / Pascolo
[7] Terreni coltivati

Since the extensions of Rome and Paris are very different, the percentage will be considered the main focus on the analysis, infact making a comparison among only raw extensions is useless.
As evidence of this the extension of Rome is 1285 km^2 and the extension of Paris is 105.4 km^2.

2.2 Introduction to random forest classification

To understand what is a random forest classifier let's refer to an explanation provided by IBM [8].
Random forest is a commonly-used machine learning algorithm, trademarked by Leo Breiman and Adele Cutler, that combines the output of multiple decision trees to reach a single result.
Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems [9] .
The random forest model is made up of multiple decision trees, decision trees are common supervised learning algorithms, they can be prone to problems, such as bias and overfitting.
However, when multiple decision trees form an ensemble in the random forest algorithm, they predict more accurate results, particularly when the individual trees are uncorrelated with each other [10].

Schematic diagram of the random forest algorithm

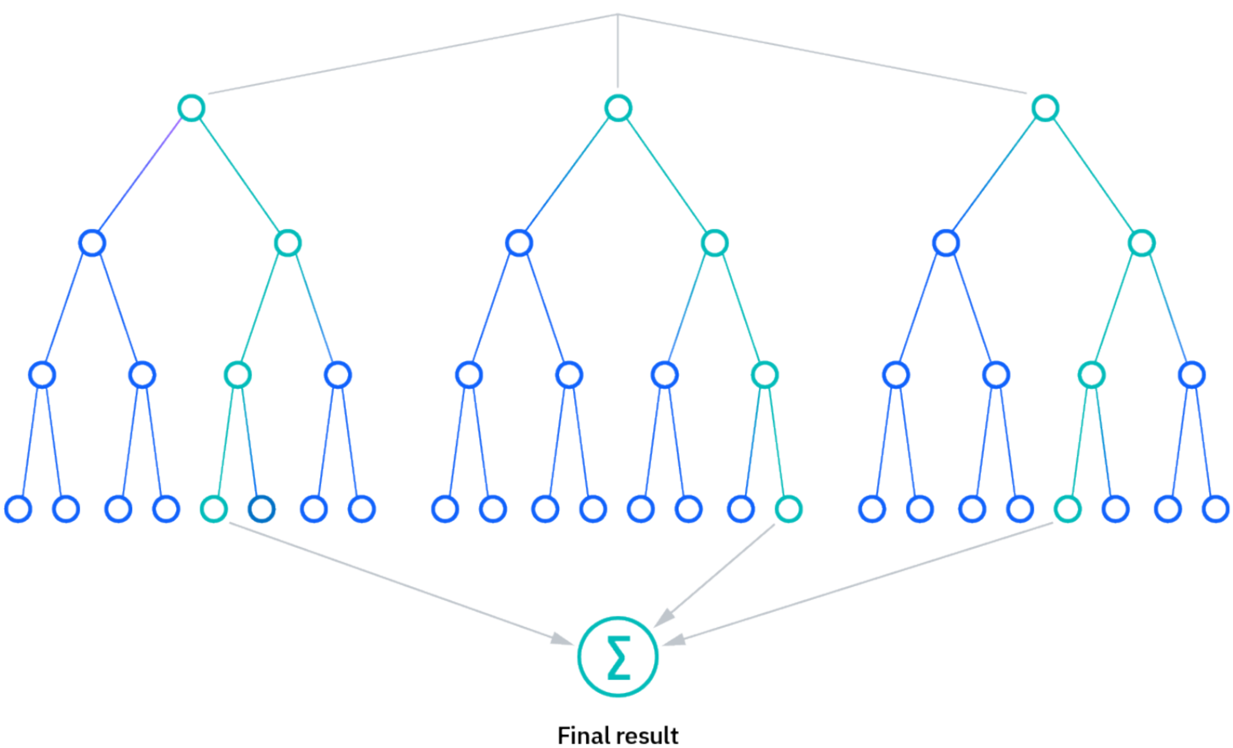


Figure 5: Source of image *IBM* [10]

[8]IBM (International Business Machines Corporation) is a multinational technology company, and is one of the world's largest companies in the sector.
[9]In this application *random forest* will be used to classify.

2.3 Defining the pipeline

As has been said before, the tool to find out in which percentage the green areas are present in the cities will be the random forest classification algorithm.
The *random forest* will be applied using Google Earth Engine, therefore a pipeline must be defined. To obtain the results sought the steps are multiple:

1. Define the area of interest (ROI);
2. Take the SENTINEL Collection, filter it as necessary;
3. Apply radiometric scaling and cloud masking to the collection;
4. Select the median image, cut it out in the roi;
5. Import the ESA collection *ESA WorldCover 10m v200*, extract its only image and define it as train image;
6. Cut the train image in the roi;
7. Sample from the latter image and define a traning set;
8. Define the random forest algorithm and train it on the traning set;
9. Test the model;
10. Apply trained algorithm on the median image defined in step 4;
11. Visualize the result;
12. Draw the conclusions by counting the pixels for each type land and so obtain the land type extension in km^2.

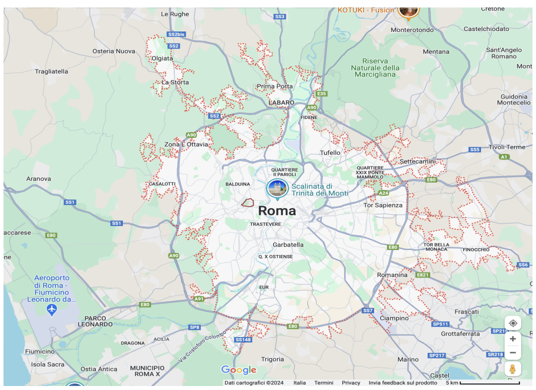
Those steps needs to be applied to both Paris and Rome, in order to make this application easier a function has been created. This function take as inputs only the ROI and the name of it to provide better outputs.

The code can be found at the following link:
`https://code.earthengine.google.com/fdc4c1e03211bc516f0847475a33e657`.

2.4 Define the parameters

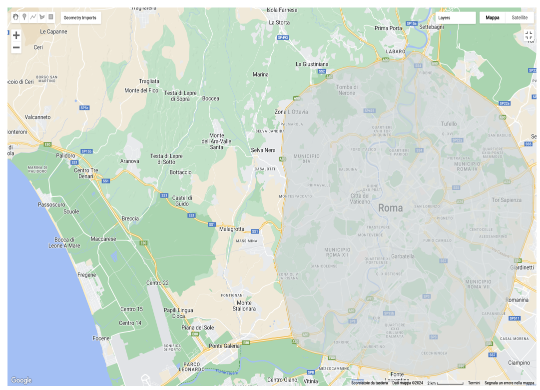
2.4.1 Define the area of interest (ROI)

Since the cities under analysis are Rome and Paris, there are two regions of interest to be defined. The areas will be selected using the polygon geometry from GEE.
In order to have a less biased analysis, given the large difference between Paris and Rome's extent, the regions of interest will be defined by the roads that surrounds the two cities.
For rome is the GRA has been selected [10], and for Paris has been selected the boulevard périphérique [11]. Of course the are areas inside the GRA that are not considered Rome, like the Vatican City, and also there are area outside the boulevard périphérique that are considered part of Paris, like the bois de Boulogne e del bois de Vincennes, but for this analysis we consider all and only the areas within those road rings.
Considering these new boundaries for cities, their extents change: from 1285 km^2 to 466.381 km^2 for Rome, and from 105.4 km^2 to 129.99 km^2 for Paris.
It is indeed a strange result for Paris, this another reason to consider only percentage results.



Boundaries of Rome by Google Maps

(scale 5km)



Boundaries of Rome given by the GRA

(scale 2km)

Boundaries of Paris by Google Maps

(scale 2km)

Boundaries of Rome given by the Périphérique

(scale 2km)

Note that the boundaries provides by Google Maps are defined by the red dotted lines.

[10] The Grande Raccordo Anulare (GRA), official name Autostrada A90, is the ring road that surrounds the city of Rome. it is characterized by a closed, discontinuous circular layout, entirely with three lanes per carriageway, with an average diameter of about 21 km and a length of 68 km

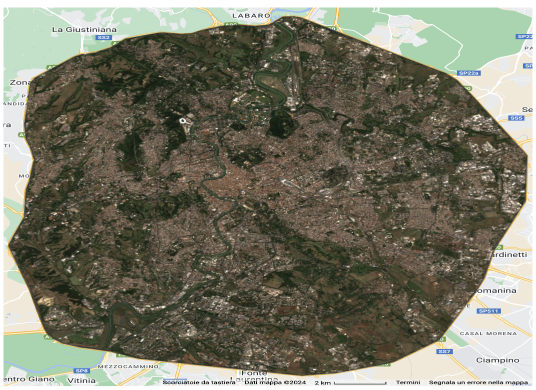
[11] The boulevard périphérique in Paris is a fast-flowing municipal road in the shape of a ring road with four or more lanes in each direction. 35.04 kilometers long, the périphérique encircles almost the entire municipal area of the French capital.

2.4.2 Filter the collection

The SENTINEL Collection used is the *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*, this collection is made up of images of all the world, in a time span that goes from **2017-03-28** to **2024-06-07**[12], dates are in yyyy/mm/dd format.

It goes without saying that work with all those images is computationally impossible, and also useless for the purpose of the analysis.
So what is usually done is filter the collection by region of interest and the time span desired.

As already has been said the pipeline will be applied for both Paris and Rome, so the **region of interest** will be the two described in the 2.4.1 section.
As **time interval** will be used a time span that goes from **2023-06-07,** to **2024-06-07**, this interval will allow to have the most recent data as possible and also selecting an entire year minimize seasonals effects, also the bigger time stamp will grant to have enough images to apply succesfully cloud masking.
The collection is also **filtered by clouds percentage**, in particular only images with less than 50 % of cloudy pixels have been selected.
Have a look on the median images obtained for the two cities.



Median image of Rome (RGB)

(scale 2km)

Median image of Paris (RGB)

(scale 2km)

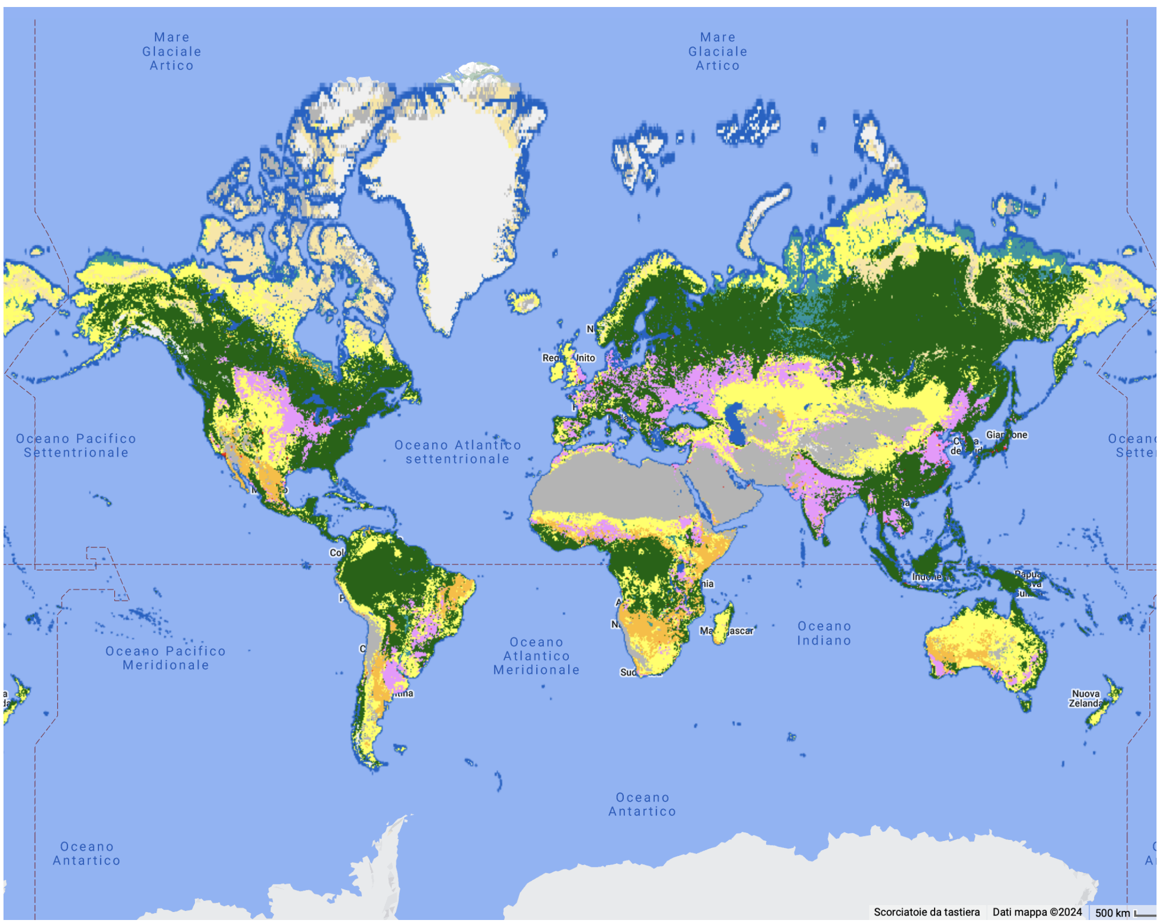
[12]Interval available at the day in which thee report has been written

2.4.3 ESA WorldCover 10m v200

The *European Space Agency (ESA) WorldCover 10 m 2021* product provides a global land cover map for 2021 at 10 m resolution based on Sentinel-1 and Sentinel-2 data.
The WorldCover product comes with 11 land cover classes and has been generated in the framework of the ESA WorldCover project, part of the 5th Earth Observation Envelope Programme (EOEP-5) of the European Space Agency [5].

This collection is made up by only one image, this image provide to classify the entire surface of the world in 11 classes.

European Space Agency (ESA) WorldCover 10 m 2021



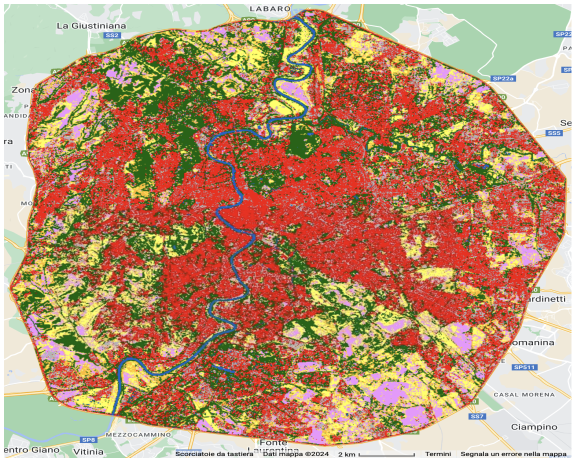
2.5 Test the models

City	Train accuracy %	Validation accuracy %
Rome	96	66
Paris	98	87

As far as Paris is concerned, the model it is very accurate. The model for Rome seems to be worse. However when considered that generally a model that has an accuracy that is better than a random choice, is considered usable and also that it is ranking among 12 classes the validation accuracy, although not the best, is acceptable.

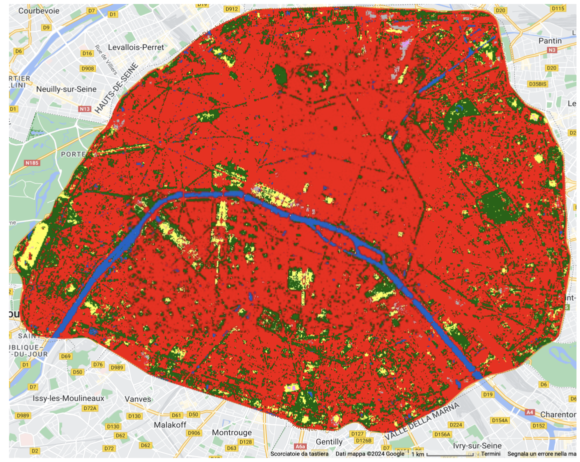
2.6 Results

After training the model and applying it, the results are as follows:



Classified image of Rome (RGB)

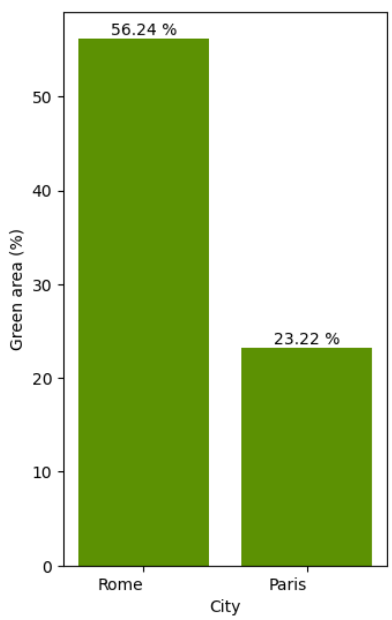
(scale 2km)



Classified image of Paris (RGB)

(scale 1km)

It is immediately noticeable that Paris have much more urban areas (red area) respect to Rome, and also less green areas (areas in green, orange, yellow and pink).

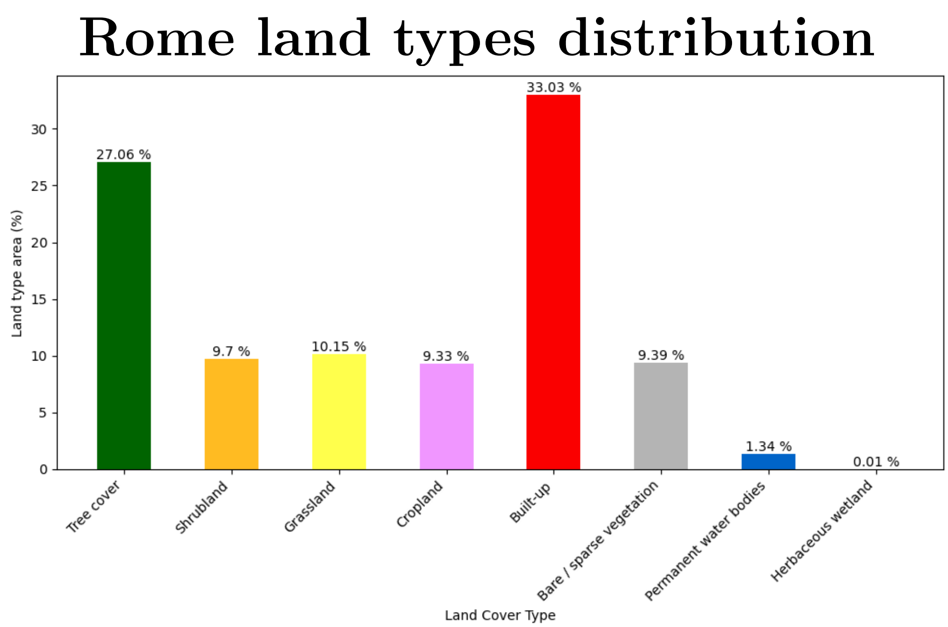


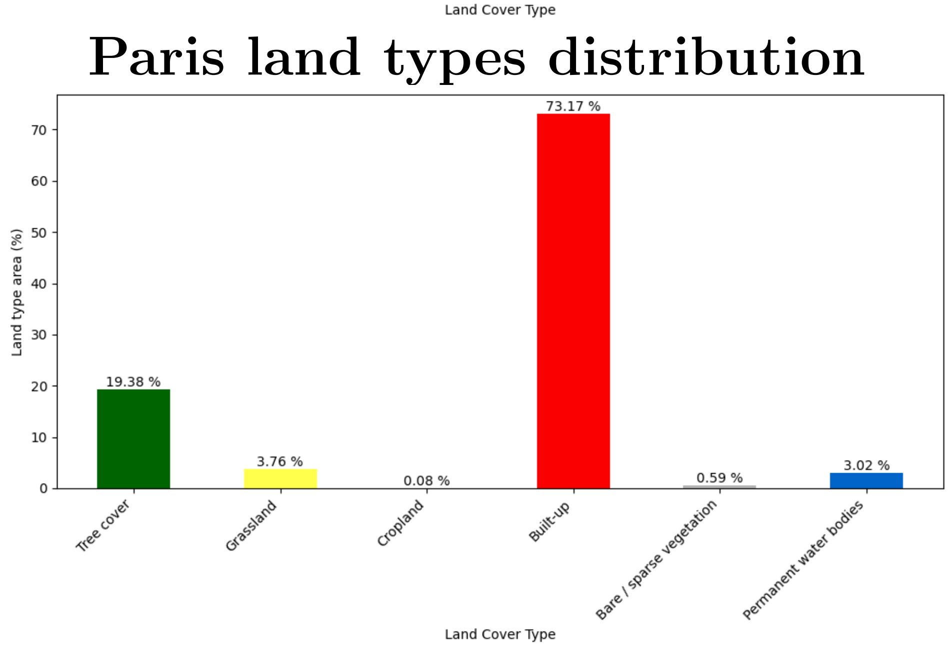
Considering as green areas:

- Tree cover
- Shrubland
- Grassland
- Cropland

it can be noted that Rome has more than twice of the green areas of Paris.
Infact Rome has the 56.24 % of the surface made up of green areas. Instead Paris has the 23.22 % of the surface made up of green areas.

To go into more detail the distributions by land type can be observed.
It can been seen how much the Rome and Paris are different, or at least how much the land type insides the GRA and the périphérique are different.
Infact Paris has the 73.17 % of **built-up** land type, instead with 33.03 % Rome has less than half of it.
Also Rome has agricultural area, 9.33 % of **cropland**, Paris has none.
Instead for what concern **permanent water bodies** both of the cities have a considerable amount, this is because both cities have a major river, the Senna for Paris and the Tevere for Rome.
Lastly with regard to the **tree-covered** area, it can be noted that the extensions recorded are not so different from the ones specified in figure 4, Rome has tree cover extension of the 27.06 %, the one recorded in 2018 was 24 %, instead Paris has tree cover extension of the 19.38%, the one recorded in 2018 was 20 %.





An important disclaimer is that the analysis done so far is very subjective.

The boundaries selected are choosen manually and are not the real ones, and also the concept of green area have been redefined as the union of several land types.

However since the pipeline has been defined, if the desire is to replicate the analysis using different boundaries for the cities the only thing to do is change the regions of interest.

The following tables are intended to **summarize** what has been done and also provide a **legend** of the colors used in the classification.

Id	**Color**	**Surface**	**Extension [km^2]**	**Extension %**
0	Dark green	Tree cover	126.221	27.06
1	Ochre	Shrubland	45.240	9.70
2	Yellow	Grassland	47.329	10.15
3	Pink	Cropland	43.507	9.33
4	Red	Built-up	154.035	33.03
5	Grey	Bare / sparse vegetation	43.775	9.39
6	Light grey	Snow and ice	0	0
7	Blue	Permanent water bodies	6.231	1.34
8	Teal	Herbaceous wetland	0.043	0.01
9	Light green	Mangroves	0	0
10	Faded yellow	Moss and lichen	0	0
Total surface of Rome (GRA)			466.381	100.00

Id	**Color**	**Surface**	**Extension [km^2]**	**Extension %**
0	Dark green	Tree cover	25.195	19.38
1	Ochre	Shrubland	0	0
2	Yellow	Grassland	4.891	3.76
3	Pink	Cropland	0.100	0.08
4	Red	Built-up	95.107	73.17
5	Grey	Bare / sparse vegetation	0.767	0.59
6	Light grey	Snow and ice	0	0
7	Blue	Permanent water bodies	3.927	3.02
8	Teal	Herbaceous wetland	0	0
9	Light green	Mangroves	0	0
10	Faded yellow	Moss and lichen	0	0
Total surface of Paris (Le Periperique)			129.987	100.00

3 Supervised machine learning, classification with support vector machine

3.1 Vaia storm

This topic has been already discussed in the first part of homework 3, however that homework ends with the suggestions that using machine learning technique the results could have been better. By applying the support vector machine it will be found out whether the suggestion was true or not. Since the topic has already been covered, some repetitions will be found in the following paragraphs, but they must be included in order to have a self-concluding report.

The Vaia storm hits Italy on Oct. 29, 2018 causing casualties, 3 billion euros worth of damage and destroying approximately 42500 hectares[13] of forest.[7]
Particularly the storm hits 4 regions:

- Lombardia
- Veneto
- Trentino-Alto Adige
- Friuli-Venezia Giulia

In Alto Adige a total of 6000 hectares [14] of forest were felled and razed to the ground.
The spruce[15] forests of the Latemar massif[16], of the municipalities of Nova Ponente, Nova Levante Fontanefredde and San Vigilio di Marebbe, appear to be the forest areas most affected and severely damaged [2].
The analysis will be focused on this particular area, with the intent to verify if the results obtained, with the satellites data, are in line with those indicated in the literature.
To try to measure the extension of the damaged area a support vector machine classificator will be used.

3.2 Introduction support vector machine classification

A Support vector machine (SVM) classificator splits the data using a linear decision surface (hyperplane) and is ideal for single class classification, for instance water - no water.
In this case of study will be very useful to distinguish between forest-damaged and undamaged forest areas.

[13] 425 km^2
[14] 60 km^2.
[15] Abete rosso
[16] Massiccio del Latemar

3.3 Defining the pipeline

As has been said before the main tool to find to measure the damage that the Vaia Storm will be the support vector machine algorithm.
The algorithm will be applied using Google Earth Engine, therefore a pipeline must be defined. To obtain the results sought the steps are multiple:

1. Define the area of interest (ROI);
2. Take the Sentinel-2 Collection, filter it as necessary;
3. Apply radiometric scaling and cloud masking[17] to the collection;
4. Select the median image, cut it out in the roi;
5. Apply NDVI masking with threshold fixed to 0.3;
6. From the latter image defined a training set;
7. Define the classifier and train it on the traning set, specify the bands on which the classifier will train;
8. Apply trained algorithm on the roi;
9. Draw the conclusions by counting the pixels for each type land and so obtain the damage area extension in km^2.

The script containing the pipeline described and all the code used to measure the damage of the Vaia storm will be the following link: `https://code.earthengine.google.com/8c770b6ac1d1aada78a913e021ee2398`.

[17] Process that consist in removing pixels that contains clouds, cloud shadow, cirrus, aerosols. It is a fondumental step when working with certain type of band and images.

3.4 Define the parameters

3.4.1 Define the area of interest (ROI)

As a region of interest the forest area near: Nova Ponente, Nova Levante Fontanefredde and San Vigilio di Marebbe has been selected.
The area has been defined using the polygon geometry from GEE.

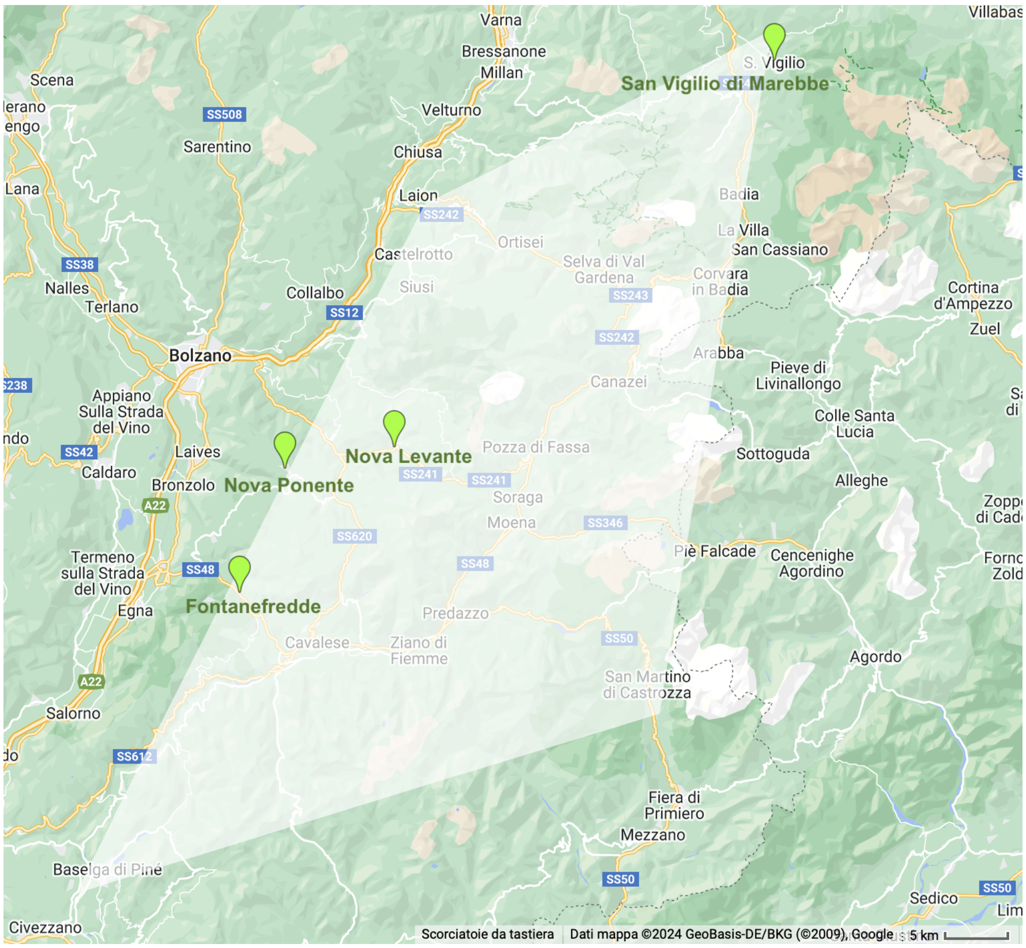


Figure 6: Region of interest

In total the area of interest is made up of 2523 km^2.

3.4.2 Filter the collection

The Sentinel Collection used is the *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*, this collection is made up of images of all the world, in a time span that goes from **2017-03-28** to **2024-06-02** [18], dates are in yyyy/mm/dd format.

It goes without saying that work with all those images is computationally impossible, and also useless for the purpose of the analysis.
So what is usually done is filter the collection by region of interest and the time span desired.

Of course as **region of interest** the entire area described in 3.4.1 has been selected.
The collection is also **filtered by clouds percentage**, in particular only images with less than 50 % of cloudy pixels have been selected.

The selected **interval** goes to the **2019-06-01** to the **2019-07-30**.
Why the focus is on an interval that goes from june to july when the storm happened on the 30 october?
Of course the ideal would have been to collect images right after the disaster but since the region of interest is often cover by clouds and snow in the winter, taking the images collection during summer can avoid those problems.
The great length of the intervals is to be implied to the research of the accuracy, in fact, when selecting only a month an excessive cloud masking was disturbing the analysis.

[18] Interval available at the day in which thee report has been written

3.4.3 NDVI threshold

A threshold fixed at 0.3 is choosen because it leave in the image only vegetation and damaged vegetation making the classification easier.
Of course this step could have been skipped, however it makes classification easier and more efficient.
Looking at the following images, it can be seen that the cities and snow in the mountains are removed from the image by the use of NDVI.

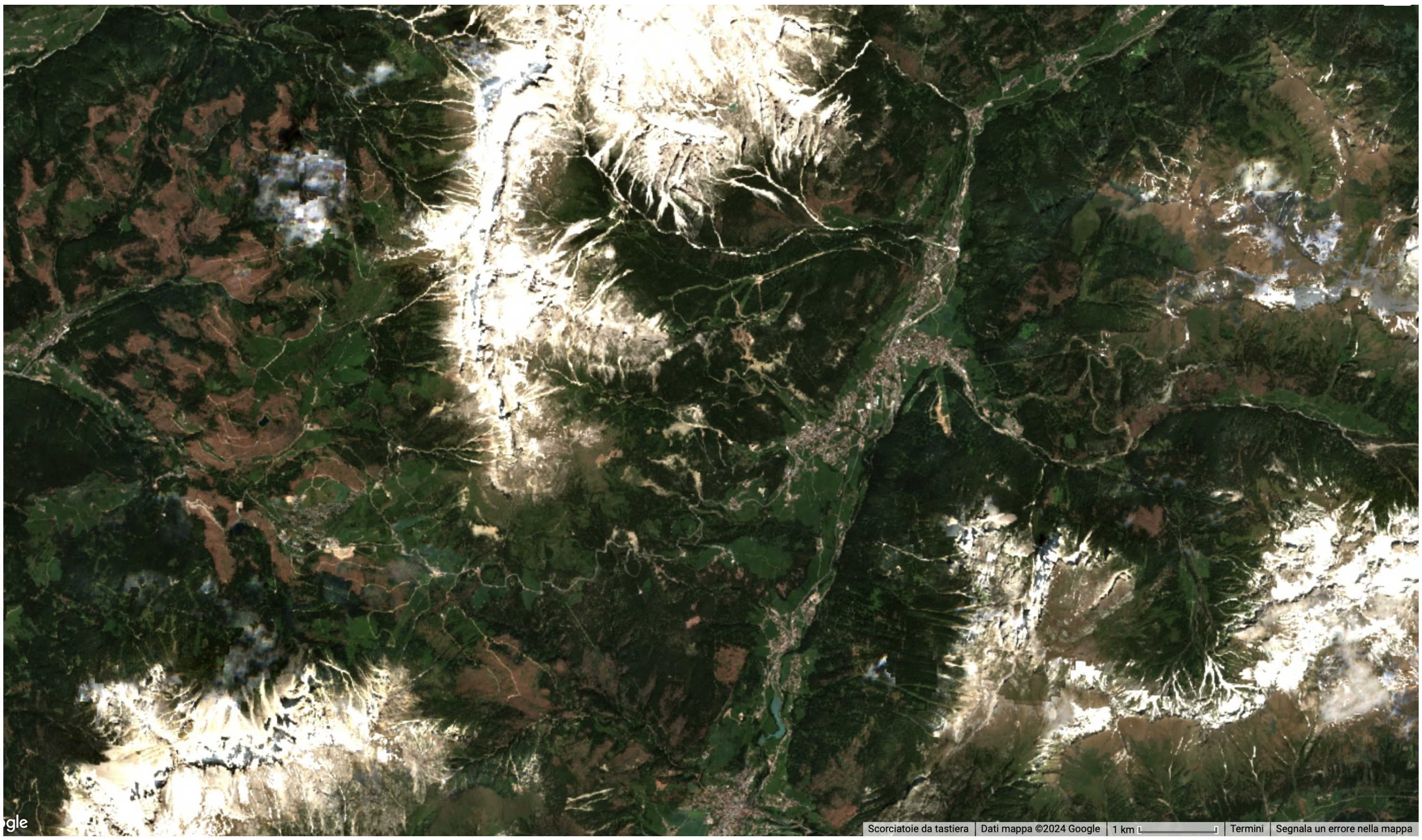


Figure 7: Region of interest RGB

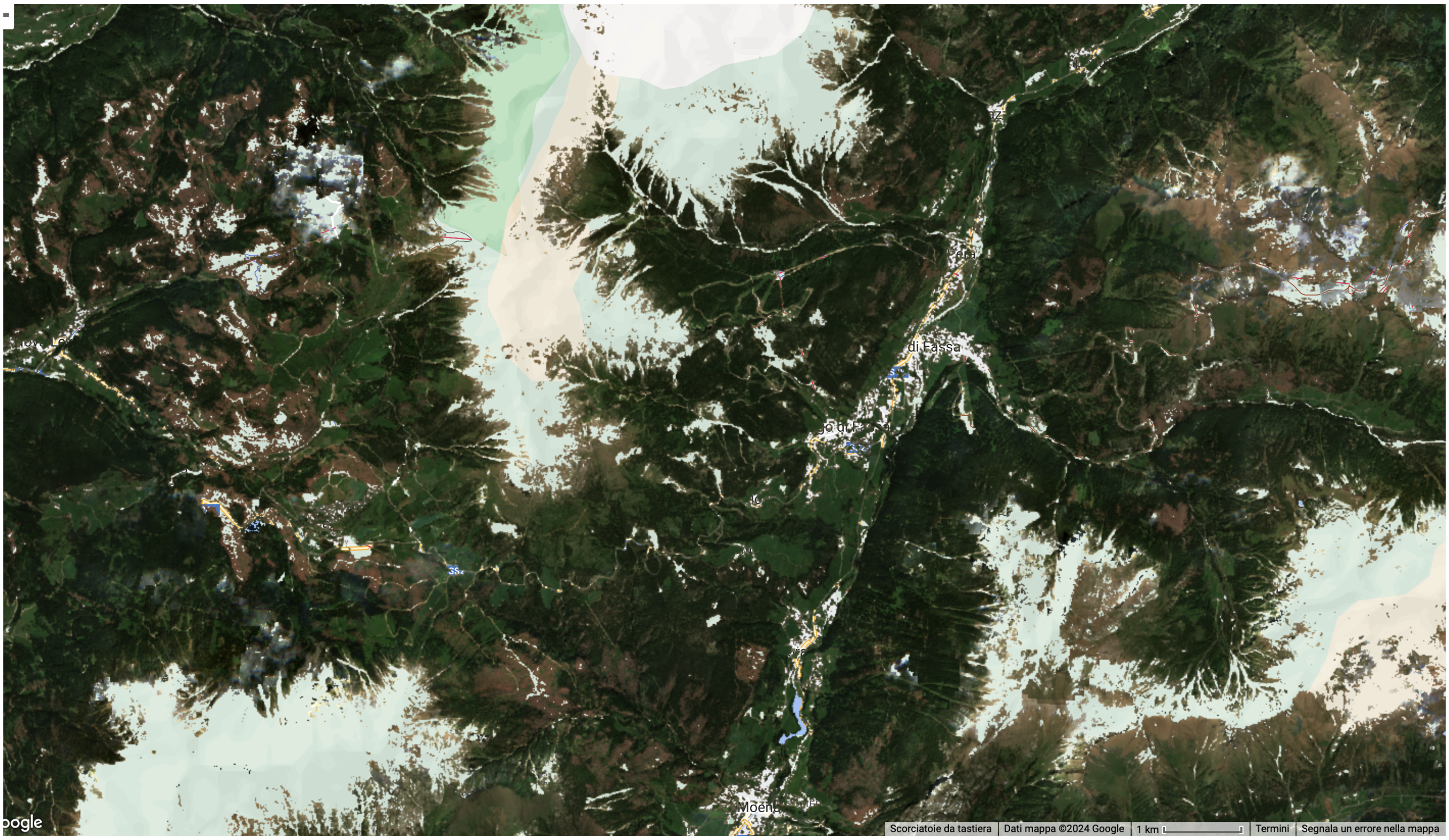


Figure 8: Region of interest masked with NDVI threshold at 0.3

3.4.4 Defining the training set

Since the support vector machine is a supervised learning procedure a training set with labels must be defined.
In Google Earth Engine to create this type of dataset is only necessary to cut out some areas and pass them to the classifier specifying their label.
In this application, 4 areas of damaged forest (areas in red) and 4 areas of non-damaged forest (areas in green) were selected. These non-damaged forest areas are divided between forest and mountain.
Doing it this way will result in a more correct classification than just specifying areas of only forest and damaged forest, it is more effective because it makes up for any errors derived from the application of NDVI.

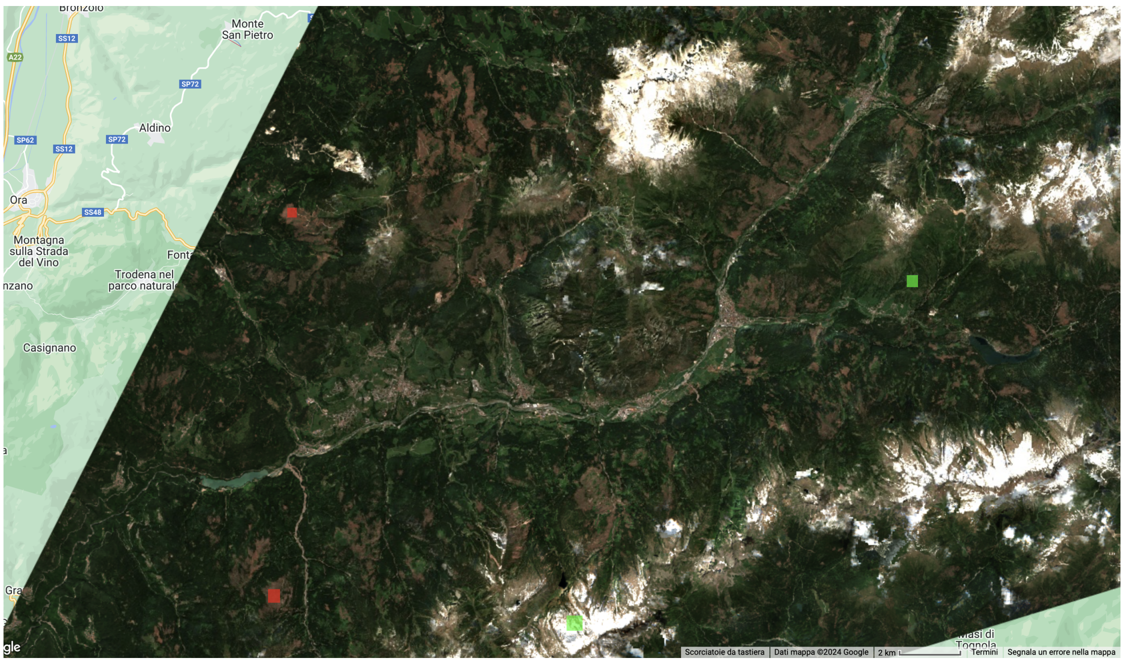


Figure 9: Define the training set

3.4.5 Specify the bands

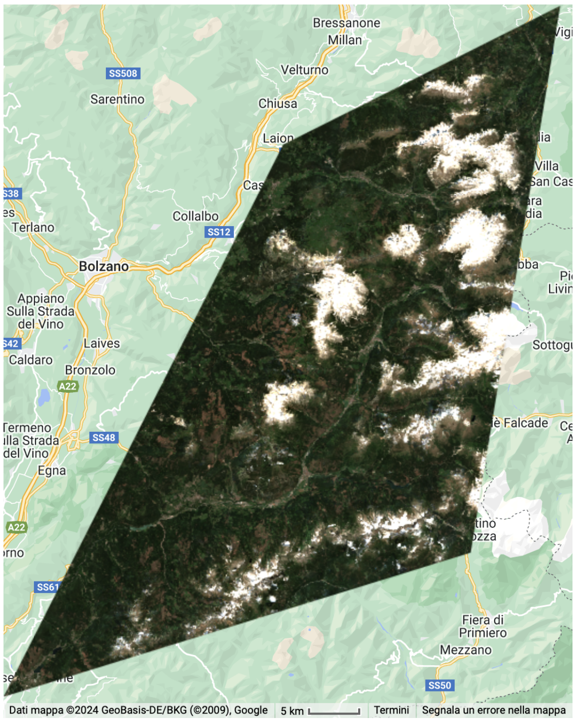
The selected bands are:

- B2: Blue
- B3: Green
- B4: Red
- B8: NIR

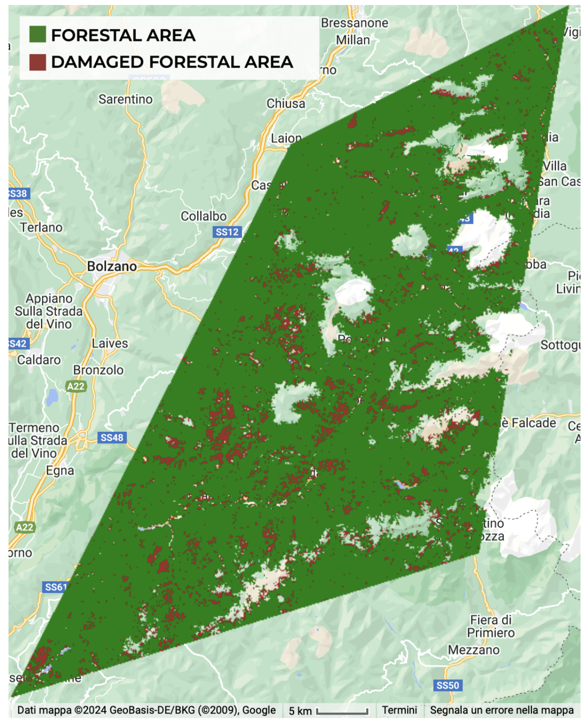
Te choice of bands B2,B3,B4 was made to have the bands define RGB and thus have a classifier that discriminates by color.
Since NIR, is a band that vegetation strongly reflects is used to try to discriminate healty vegetation and damaged vegetation.
A new NDVI band could also be defined and have the model use it , however since NDVI was widely used in homework 3 this option was discarded.

3.5 Result

After training the model and applying it, the results are as follows.



ROI after the storm RGB



ROI after the storm classified

The damaged area is 215.24 km^2, not to much different from 216.2 km^2 result given by the NDVI methodology used in homework 3.
This convergence in results is a strong indicator in favor of the correctness of the analysis.

By looking at the following tables what has been said before can be observed, but it can also be noted that the extent of the two areas of interest is different.
However this is not strange, the smallest extension of the SVM methodology is due to the fact that a masked image is used.

NDVI methodology results	
Area of interest	2523.6 km^2
Forest area	1890.9 km^2
Damaged forest area	216.2 km^2

SVM methodology results	
Area of interest	2175.70 km^2
Forest area	1960.46 km^2
Damaged forest area	215.24 km^2

References

[1] “”Map of United Arab Emirates, Middle East””. In: *nationsonline.org* (). URL: https://www.nationsonline.org/oneworld/map/united_arab_emirates_map.htm/.

[2] ”Provincia autonoma di Bolzano - Alto Adige”. “”Tempesta Vaia un anno dopo, sgomberati i due terzi degli alberi””. In: *News ufficio stampa* (Oct. 2019). URL: https://news.provincia.bz.it/it/news-archive/631895.

[3] Anna Fleck. “How Green Are The Capitals of Europe?” In: *Statista* (). URL: https://www.statista.com/chart/27950/the-greenest-cities-of-europe/.

[4] Google. “”What is unsupervised learning?”” In: *Google Cloud* (). URL: https://cloud.google.com/discover/what-is-unsupervised-learning.

[5] Google. “ESA WorldCover 10m v200”. In: *Earth Engine Data Catalog* (). URL: https://developers.google.com/earth-engine/datasets/catalog/ESA_WorldCover_v200.

[6] Andrew Chee Keng Lee, Hannah C Jordan, and Jason Horsley. “Value of urban green spaces in promoting healthy living and wellbeing: prospects for planning”. In: *Risk management and healthcare policy* (2015), pp. 131–137. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4556255/.

[7] ”Riccardo De Marco”. “”La tempesta Vaia del 2018 e i danni devastanti dell’evento meteorologico sulle foreste alpine””. In: *Geopop* (Oct. 2023). URL: https://www.geopop.it/la-devastante-tempesta-vaia-del-2018-i-danni-dellestremo-evento-meteorologico/.

[8] “Percentage of total green infrastructure, urban green space, and urban tree cover in the area of EEA-38 capital cities (excluding Liechtenstein)”. In: *European Environment Agency* (). URL: https://www.eea.europa.eu/data-and-maps/daviz/percentage-of-total-green-infrastructure#tab-googlechartid_chart_11.

[9] “United Arab Emirates”. In: *Istituto della Enciclopedia Italiana fondata da Giovanni Treccani* (). URL: https://www.treccani.it/enciclopedia/emirati-arabi-uniti/.

[10] “What is random forest?” In: *IBM* (). URL: https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems..