Land Cover Classification of Bay Area

Sri Sumanth - IMT2016081
Durga Yasasvi - IMT2016060
Manideep Nizam - IMT2016026

Background

In the context of Land cover classification, Using different classification techniques there are many papers that were proposed to classify regions for different parts of the world.

In this project we focus to classify the Bay Area's land cover, which is a part of the Californian family that's close to the Bay. Among the 10 counties of Bay Area's we considered 5 counties for classification. They are San Francisco, San Mateo, Santa Clara, Contra Costa and Alameda.

Using the very well known tool GRASS GIS we classified the land cover.

Objective

Our main objective is to do a pixel wise classification of the Bay area into four different categories: Vegetation, Urban, Water bodies, other(ex: Barren Land)

Contributions

All contributed equally to the project.

Dataset

From the open source website usgs.gov, we collected the Bay Area data from the Landsat-8 satellite. The bay area data contains all the 10 counties data.

Landsat 8 measures different frequencies along the electromagnetic spectrum. It contains 11 multispectral bands, each having 30 m resolution. The multiple spectrums ranges from

- Violet (senses the deep blues, air molecules, dust and moisture),
- Visible Spectrum (Blue, Green, Red),
- NIR (to capture vegetation, chlorophyll content),
- SWIR (to tell whether it is wet or dry earth),
- Panchromatic (sharpest band),
- a band for clouds,
- TIR (to see heat).

Methodology

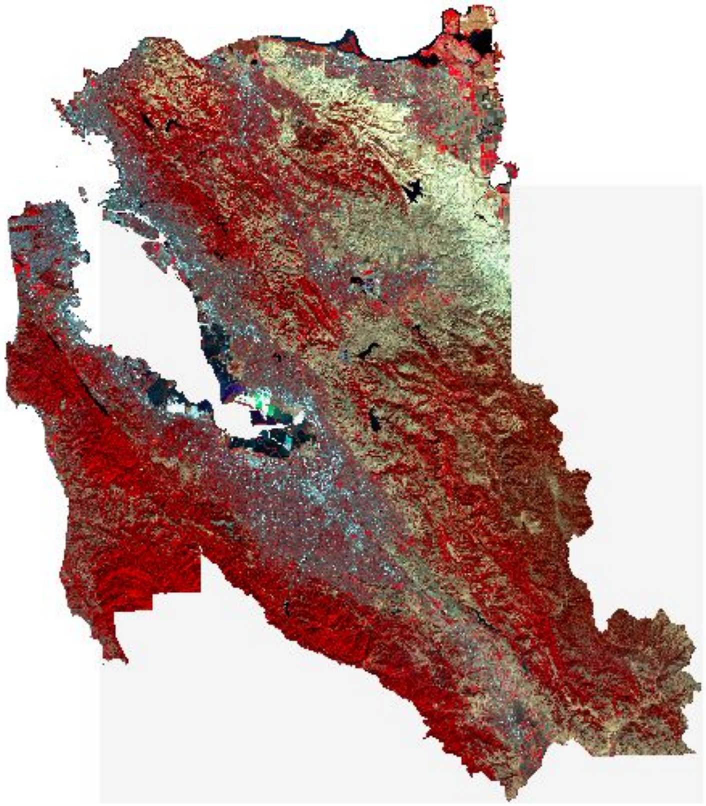
- Majority of the project, we used GRASS GIS application.
- Using the visible spectrum (blue, green, red) we constructed a True Color Composite map.
- As the vegetation/chlorophyll content is captured in the NIR band, using the NIR spectrum along with Visible spectrum we generated a False Colour Composite.
- As said earlier our focus is on the 5 counties of the Bay Area and the data obtained contains 10 counties, we need to cut the required counties from the data.
- Using the 5 counties shapefiles, we converted them into the raster data and overlaid onto the FCC layer using '*' operator on it.

Training and Inference

- Having the required counties FCC, we need to prepare the training samples to feed into the classification model.
- Using the GRASS GIS tool, we made training samples. Each of the 4 classes contained around 15-18 training samples. A total of around 7000 pixels were made.
- Now using the well known ML models like SVM, Random forests we fed our training samples into them and classified.
- In the same way, we generated testing samples and measured the accuracy.
- Apart from classification, without any training samples we performed clustering.

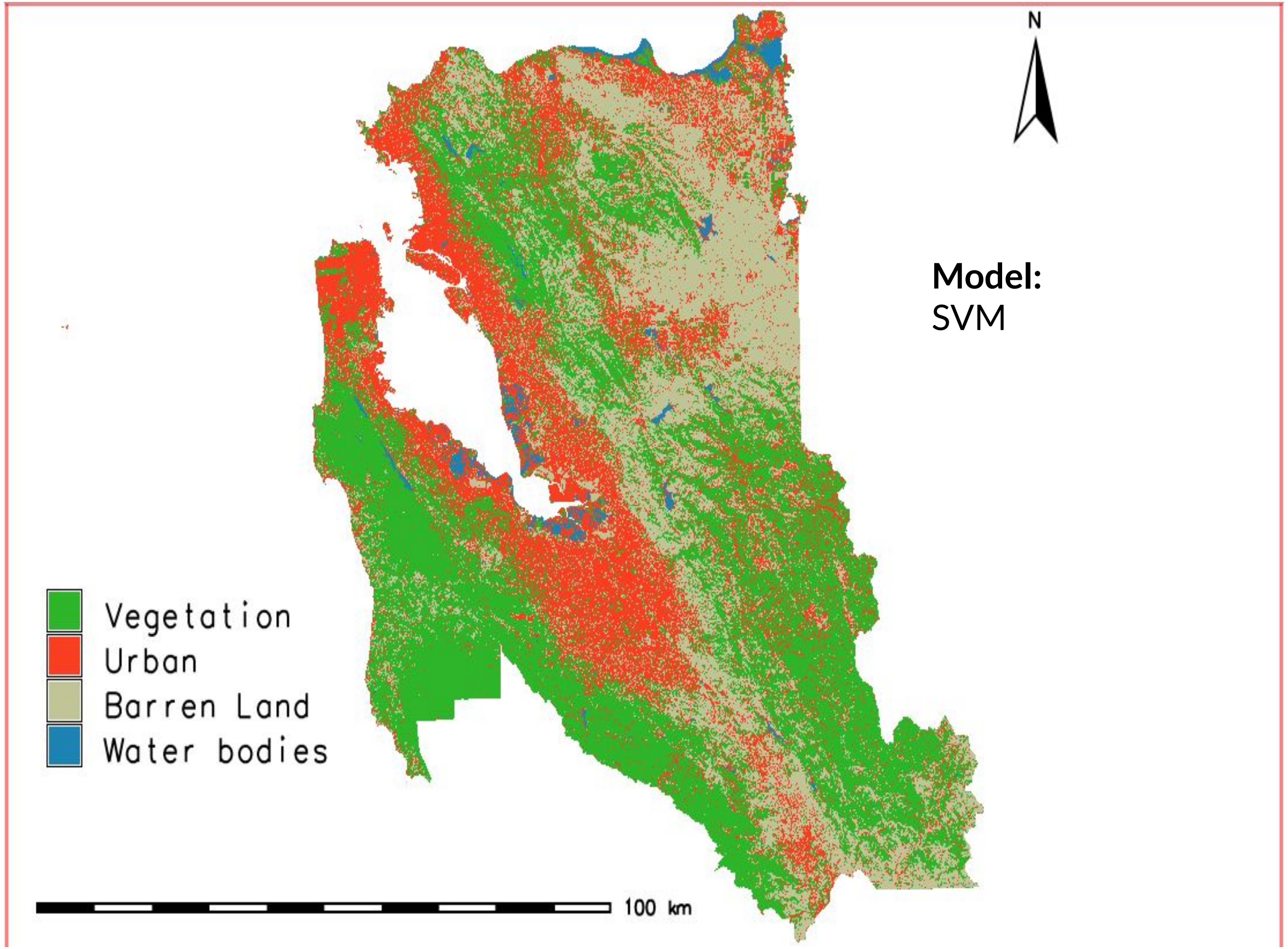
TCC

Cropping image using Shapefile and applying FCC



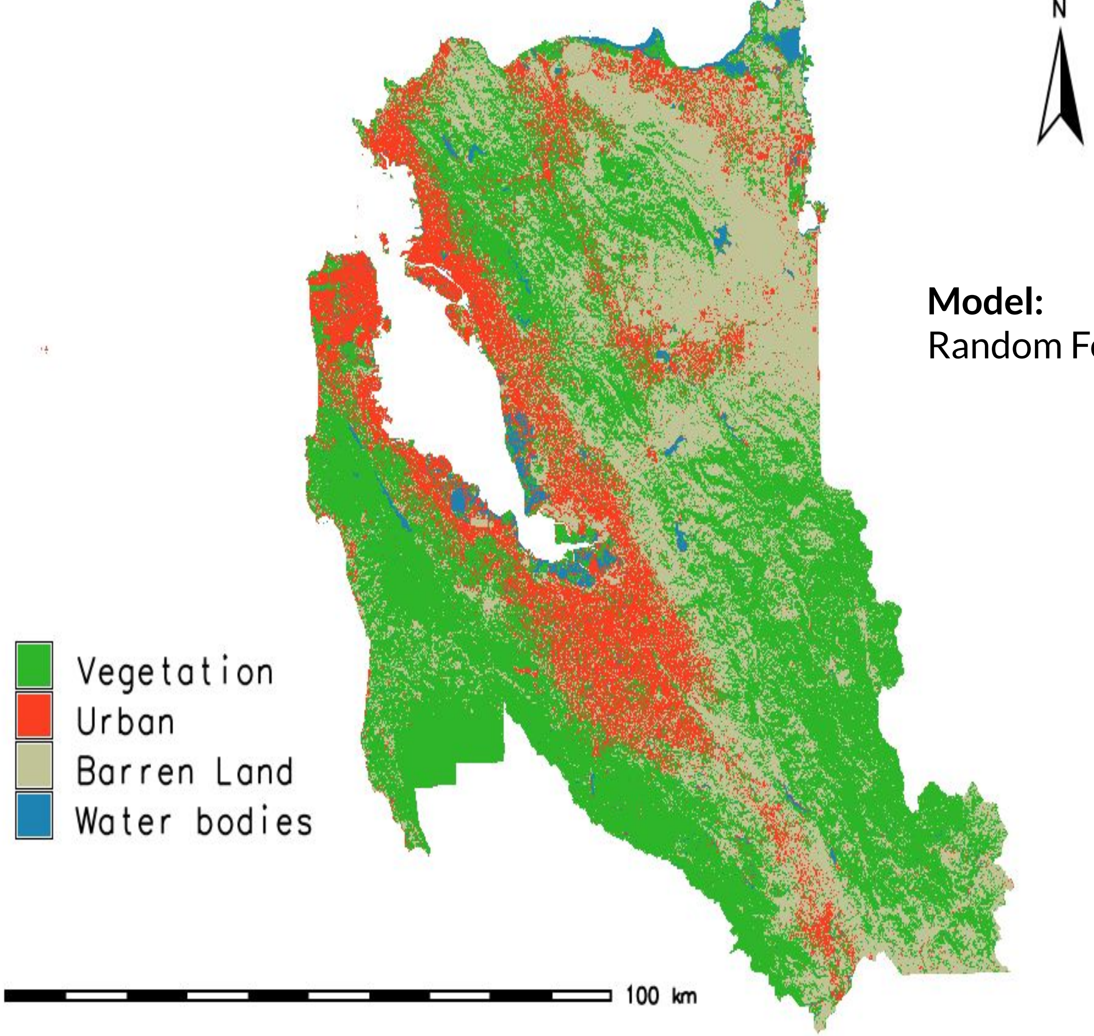
Training Sample

Results - SVM Classified



Model:
SVM

Results - Random Forest



Model:
Random Forest

Accuracy - SVM

```
                 MAP1
     cat#    1    2    3    4
 M      1    7420     56   0     0
 A      2    0    4275     728 0
 P      3    0    8    4539      0
 2      4    10   348  0    4782
Col Sum      7430     4687      5267      4782
cat#      Row Sum
S       1        7420              56              0              0         7476
V       2           0            4275            728              0        19955
M       3           0               8           4539              0        36981
_       4          10             348              0           4782        59147
      123559
Cats      % Commission     % Ommission Estimated Kappa
1    0.749064     0.134590     0.988733
2    14.551269    8.790271     0.815468
3    0.175940     13.821910    0.997692
4    6.964981     0.000000     0.911191
Kappa           Kappa Variance
0.929933        0.000004
Obs Correct Total Obs    % Observed Correct
21016           22166        94.811874
```

Accuracy - Random Forest

```
     cat#   1   2   3   4
 M     1    7471    1   0   4
 A     2    0   4271    731 1
 P     3    0   3   4544    0
 2     4    261 3   0   4876
Col Sum     7732    4278    5275    4881
cat#    Row Sum
R     1      7471           1           0           4       7476
a     2         0        4271         731           1      19955
n     3         0           3        4544           0      36981
d     4       261           3           0        4876      59147
   123559
Cats    % Commission    % Ommission Estimated Kappa
1   0.066881    3.375582    0.998973
2   14.631221   0.163628    0.818697
3   0.065978    13.857820   0.999134
4   5.136187    0.102438    0.934134
Kappa       Kappa Variance
0.938701    0.000004
Obs Correct Total Obs   % Observed Correct
21162       22166       95.470540
```

Future work

In this project we analyzed only for the year 2019. In future we would like collect the same area's data for different timestamps and analyze how the vegetation or the urban areas has been growing throughout the time.

Also we would like to use different Machine Learning models and Neural Networks to improve the results.