**FOREST COVER TYPE (KERNELS ONLY)**

In this competition asked to predict the forest cover type (the predominant kind of tree cover) from cartographic variables. The actual forest cover type for a given 30 x 30 meter cell was determined from US Forest Service (USFS) Region 2 Resource Information System data. Independent variables were then derived from data obtained from the US Geological Survey and USFS. The data is in raw form and contains binary columns of data for qualitative independent variables such as wilderness areas and soil type.

This study area includes four wilderness areas located in the Roosevelt National Forest of northern Colorado. These areas represent forests with minimal human-caused disturbances, so that existing forest cover types are more a result of ecological processes rather than forest management practices.

## **DATASET**

The study area includes four wilderness areas located in the Roosevelt National Forest of northern Colorado. Each observation is a 30m x 30m patch. We are asked to predict an integer classification for the forest cover type. The seven types are:

1 - Spruce/Fir  
2 - Lodgepole Pine  
3 - Ponderosa Pine  
4 - Cottonwood/Willow  
5 - Aspen  
6 - Douglas-fir  
7 - Krummholz

The training set (15120 observations) contains both features and the Cover\_Type. The test set contains only the features. You must predict the Cover\_Type for every row in the test set (565892 observations).

## Data Fields

**Elevation** - Elevation in meters  
**Aspect** - Aspect in degrees azimuth  
**Slope** - Slope in degrees  
**Horizontal\_Distance\_To\_Hydrology** - Horz Dist to nearest surface water features  
**Vertical\_Distance\_To\_Hydrology** - Vert Dist to nearest surface water features  
**Horizontal\_Distance\_To\_Roadways** - Horz Dist to nearest roadway  
**Hillshade\_9am** (0 to 255 index) - Hillshade index at 9am, summer solstice  
**Hillshade\_Noon** (0 to 255 index) - Hillshade index at noon, summer solstice  
**Hillshade\_3pm** (0 to 255 index) - Hillshade index at 3pm, summer solstice  
**Horizontal\_Distance\_To\_Fire\_Points** - Horz Dist to nearest wildfire ignition points  
**Wilderness\_Area** (4 binary columns, 0 = absence or 1 = presence) - Wilderness area designation  
**Soil\_Type** (40 binary columns, 0 = absence or 1 = presence) - Soil Type designation  
**Cover\_Type** (7 types, integers 1 to 7) - Forest Cover Type designation

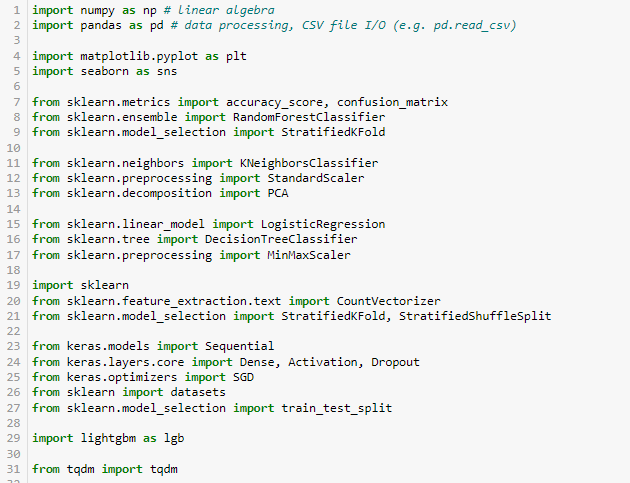
The wilderness areas are:

1 - Rawah Wilderness Area  
2 - Neota Wilderness Area  
3 - Comanche Peak Wilderness Area  
4 - Cache la Poudre Wilderness Area

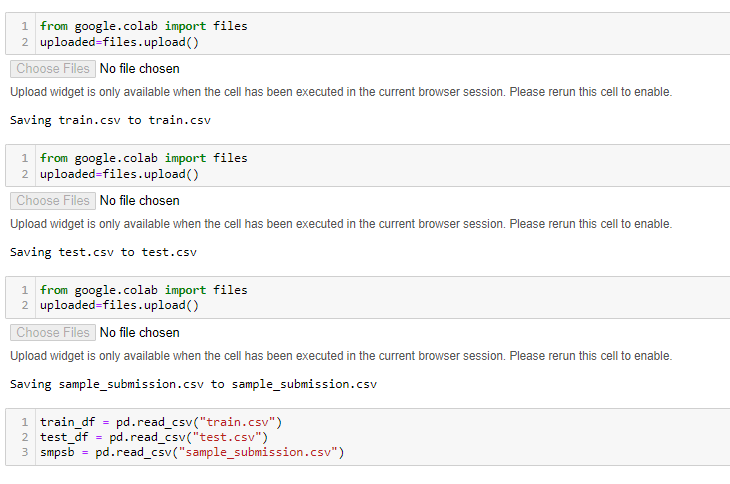
The soil types are:

1 Cathedral family - Rock outcrop complex, extremely stony.  
2 Vanet - Ratake families complex, very stony.  
3 Haploborolis - Rock outcrop complex, rubbly.  
4 Ratake family - Rock outcrop complex, rubbly.  
5 Vanet family - Rock outcrop complex complex, rubbly.  
6 Vanet - Wetmore families - Rock outcrop complex, stony.  
7 Gothic family.  
8 Supervisor - Limber families complex.  
9 Troutville family, very stony.  
10 Bullwark - Catamount families - Rock outcrop complex, rubbly.  
11 Bullwark - Catamount families - Rock land complex, rubbly.  
12 Legault family - Rock land complex, stony.  
13 Catamount family - Rock land - Bullwark family complex, rubbly.  
14 Pachic Argiborolis - Aquolis complex.  
15 unspecified in the USFS Soil and ELU Survey.  
16 Cryaquolis - Cryoborolis complex.  
17 Gateview family - Cryaquolis complex.  
18 Rogert family, very stony.  
19 Typic Cryaquolis - Borohemists complex.  
20 Typic Cryaquepts - Typic Cryaquolls complex.  
21 Typic Cryaquolls - Leighcan family, till substratum complex.  
22 Leighcan family, till substratum, extremely bouldery.  
23 Leighcan family, till substratum - Typic Cryaquolls complex.  
24 Leighcan family, extremely stony.  
25 Leighcan family, warm, extremely stony.  
26 Granile - Catamount families complex, very stony.  
27 Leighcan family, warm - Rock outcrop complex, extremely stony.  
28 Leighcan family - Rock outcrop complex, extremely stony.  
29 Como - Legault families complex, extremely stony.  
30 Como family - Rock land - Legault family complex, extremely stony.  
31 Leighcan - Catamount families complex, extremely stony.  
32 Catamount family - Rock outcrop - Leighcan family complex, extremely stony.  
33 Leighcan - Catamount families - Rock outcrop complex, extremely stony.  
34 Cryorthents - Rock land complex, extremely stony.  
35 Cryumbrepts - Rock outcrop - Cryaquepts complex.  
36 Bruss family - Rock land - Crumbiest complex, extremely stony.  
37 Rock outcrop - Cryumbrepts - Cryorthents complex, extremely stony.  
38 Leighcan - Moran families - Cryaquolls complex, extremely stony.  
39 Moran family - Cryorthents - Leighcan family complex, extremely stony.  
40 Moran family - Cryorthents - Rock land complex, extremely stony.

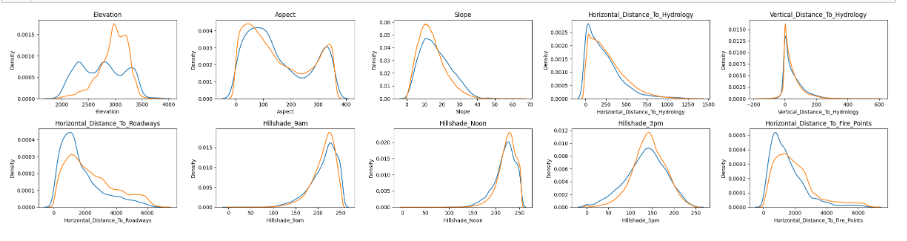
**STEP 1: IMPORTING LIBRARIES**

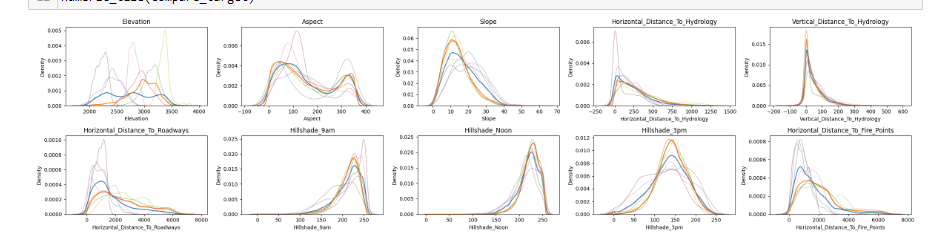
****

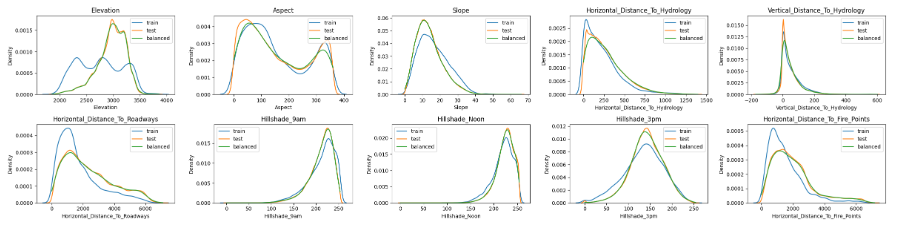
**STEP 2: DATA LOADING**

****

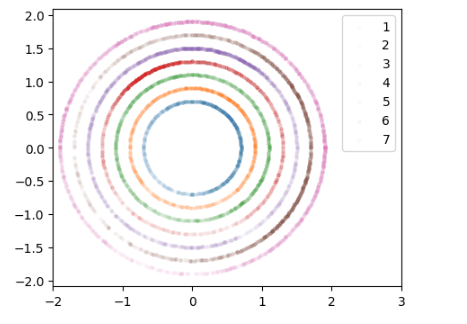
**STEP 3: PREPROCESSING**

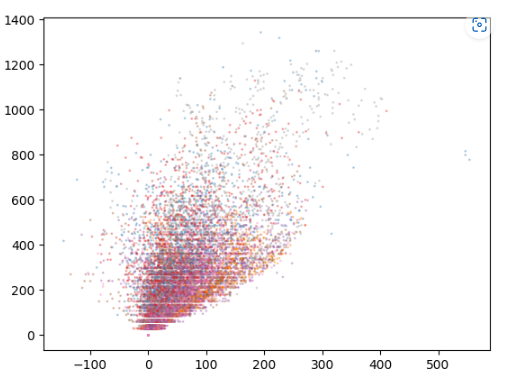
****

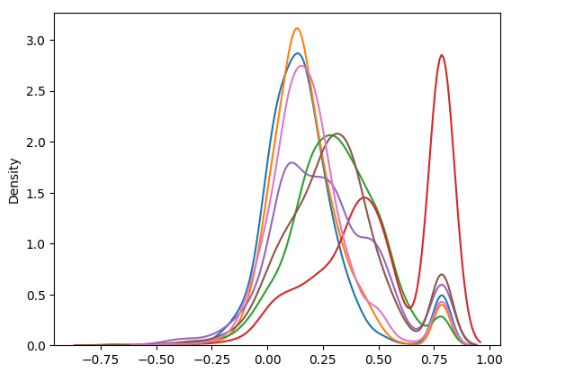
****

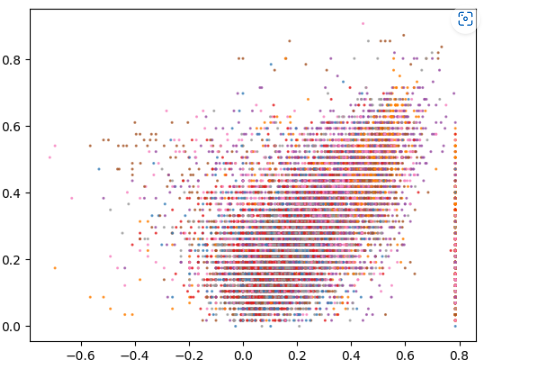
****

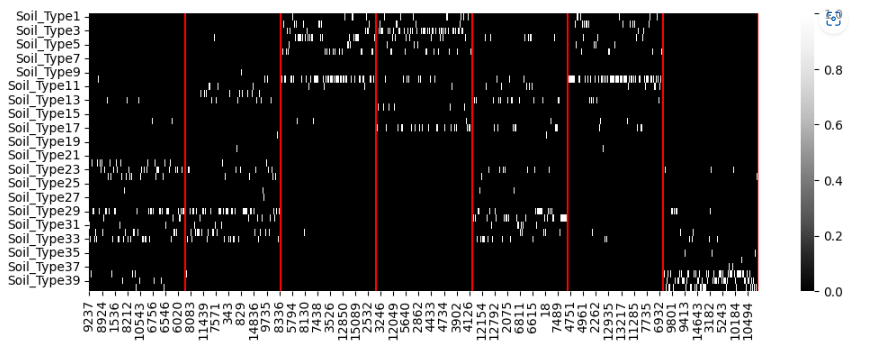
**STEP 4: FEATURE ENGINEERING**

****

****

****

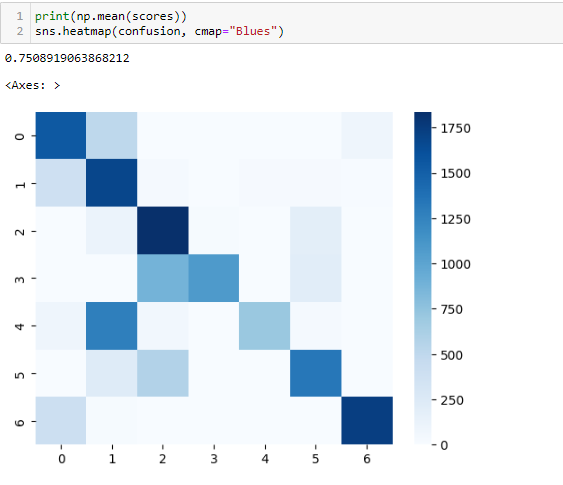
****

****

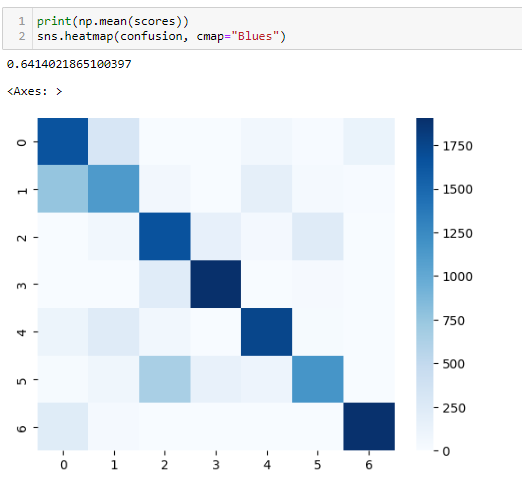
**STEP 5: MODELING**

without KNN&DT features

\* Random Forest Classifier

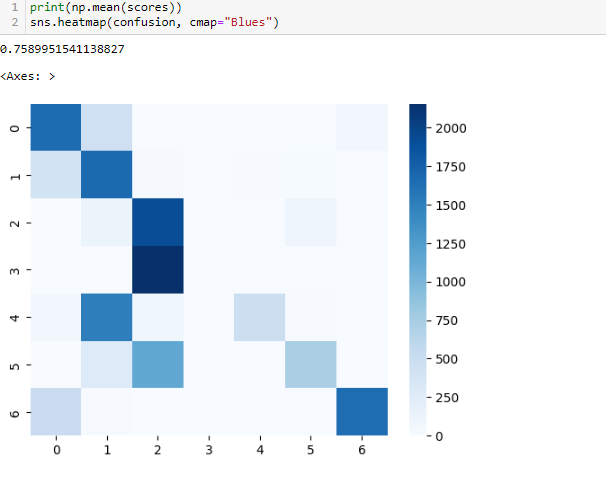


\* PCA & K-nearest Neighbors Classifier

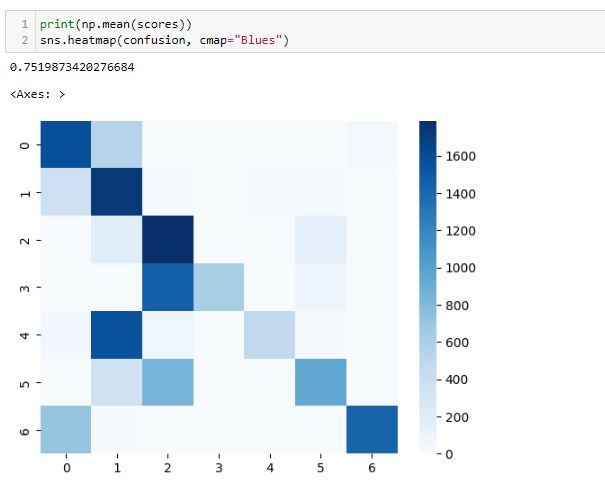


with KNN & DT features

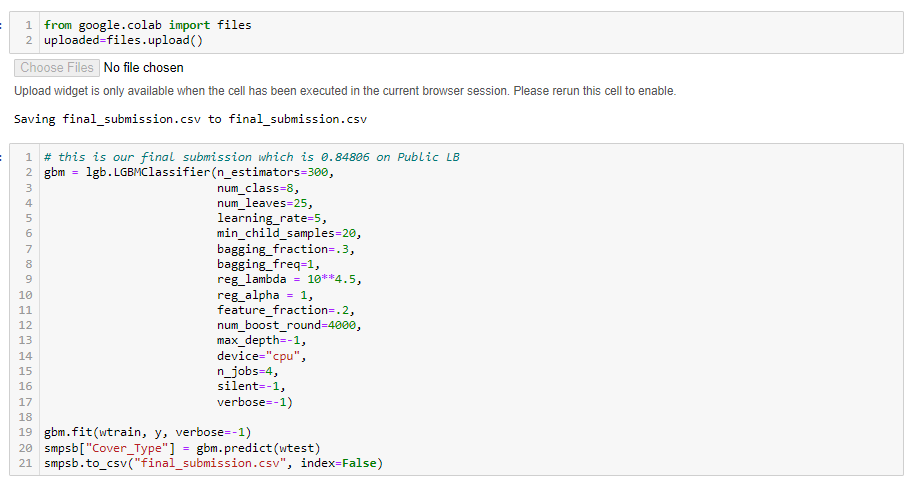
\* Random Forest Classifier



\* Logistic Regression



**STEP 6: STACKING WITH LIGHTGBM**

****