

## Team 7 Project Phase 2 : Online Feature Selection Report

### ➤ Task 1: Feature Selection

Using SAOLA on the new data set and taking 4 days data, we found the following two important features:

1. t850 Day4 location 1209
2. t850 Day4 location 4179

t850 is the 850 hPa temperature

Location 1209 is India(Uttarakhand)

Location 4179 is Pacific Ocean somewhere closer to Panama

#This script is used to select features using SAOLA

```
import numpy as np
import math
import glob

Sr = 1/(math.sqrt(11300 - 3))
def zScore(np1, np2):
    pear_coef = np.corrcoef(np1.T, np2.T)[1,0]
    fisher_z = np.arctanh(pear_coef)
    z_score = (fisher_z / Sr)
    return z_score

target = np.load('target_1980_2010.npy')
targetSum = target[:, :1]

relevantFeatures = (np.random.randint(0,100, size =(11300,
1))).astype(float)
relevantFeatureNames = ["dummy"]

for day in range(0,4):
    for filename in
sorted(glob.glob('data_D_1980_2010_part*.npy')):
        data = np.load(filename).astype(float)
        data = data[day:11300+day, :]
        columnSize = data.shape[1]
        continueFeatureLoop = False
        for column in range(0, columnSize):
            feature = data[:, column: column+1]
            zScore_f_c = zScore(feature, targetSum)
            if (abs(zScore_f_c) < 1.96):
                continue
            if(relevantFeatures.shape[1]==1):
                relevantFeatures = np.hstack((relevantFeatures,
feature))
                relevantFeatureNames.append(str("Day" + str(day) +
filename[-9:]) + str(column))
                continue
```

## Team 7 Project Phase 2 : Online Feature Selection Report

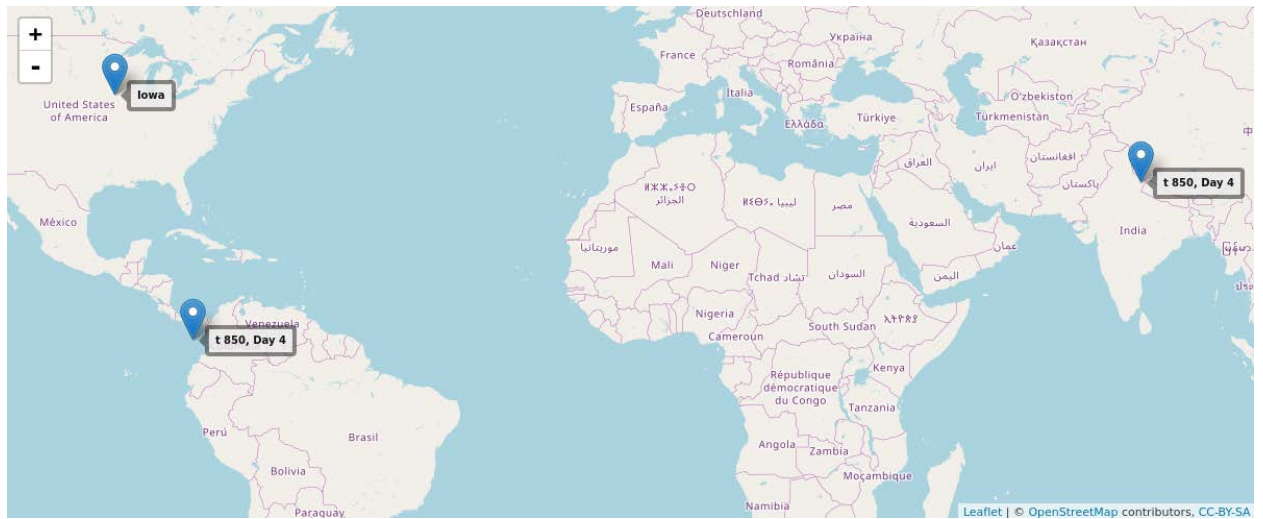
```
        relevantFeaturesIndex = 1
        while relevantFeaturesIndex <
relevantFeatures.shape[1]:
            relevantColumn = relevantFeatures[:,
relevantFeaturesIndex: relevantFeaturesIndex+1]
            zScore_y_c = zScore(relevantColumn, targetSum)
            zScore_f_y = zScore(feature, relevantColumn)
            if (abs(zScore_y_c) > abs(zScore_f_c) and
abs(zScore_f_y) > abs(zScore_f_c)):
                continueFeatureLoop = True
                break
            if(abs(zScore_f_c) > abs(zScore_y_c) and
abs(zScore_f_y) >abs(zScore_y_c)):
                relevantFeatures = np.delete(relevantFeatures,
relevantFeaturesIndex, axis=1)
                del relevantFeatureNames[relevantFeaturesIndex]
                relevantFeaturesIndex = relevantFeaturesIndex -
1
                relevantFeaturesIndex = relevantFeaturesIndex + 1
            if(continueFeatureLoop == True):
                continueFeatureLoop = False
                continue
            relevantFeatures = np.hstack((relevantFeatures,
feature))
            relevantFeatureNames.append("Day" + str(day) +
str(filename[-9:]) + str(column))

relevantFeatures = np.delete(relevantFeatures, 0, axis=1)
np.savetxt('SAOLAv2_4days_relevantFeatures.csv', relevantFeatures,
delimiter=',', fmt='%f')

del relevantFeatureNames[0]
with open("SAOLAv2_4days_relevantFeatureNames.txt", "w") as output:
    output.write(str(relevantFeatureNames))
```

## Team 7 Project Phase 2 : Online Feature Selection Report

### ➤ Task 3: Highlighting the relevant features we got on a map



Features effecting precipitation in Iowa

### Team member names and emails:

Name	Email
Sonica Kalmangi	sonica.kalmangi001@umb.edu
Jacob Robins	jacob.robins001@umb.edu
Sidhraj Solanki	Sidhraj.solanki001@umb.edu
Swapnil Patil	swapnil.patil001@umb.edu