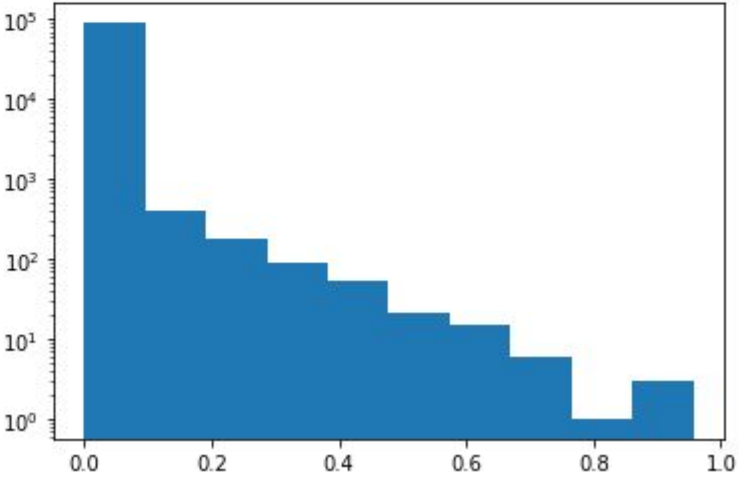
- Matrix Size: 250*361
- Total Observation: 90250
- Top 0.5% CutOff: 0.16

Very Sparse Matrix, after processing, there are 451 observations of 1, the rest are set to 0.

All Observation histogram

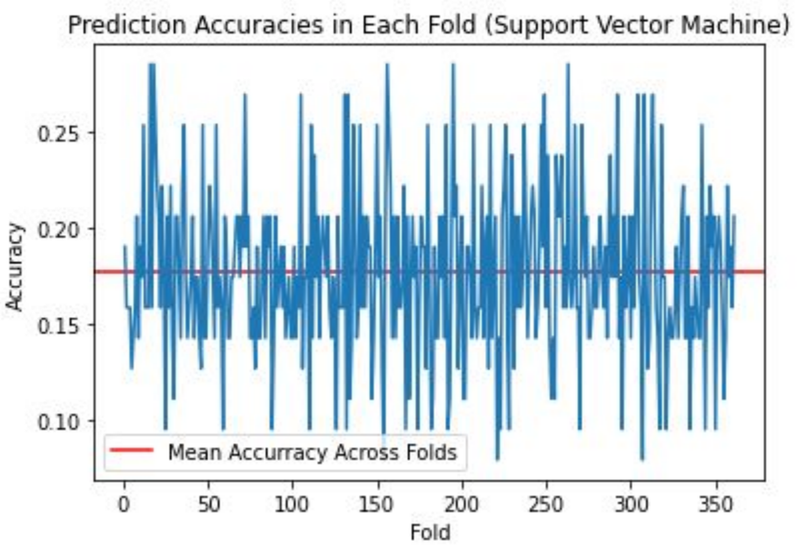


Plant Species

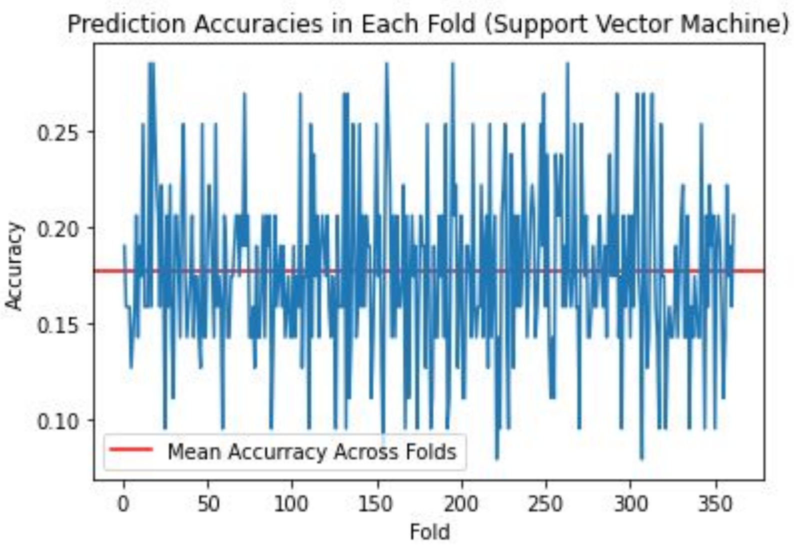
13 Unique Species, 250 samples total, Average 19 samples per label.

{'POPR', 'ELSC', 'POLE', 'TRSP', 'ELEL', 'ELTR', 'POST', 'FETH', 'FEBR', 'FESA', 'POAL', 'ACNE', 'ACLE'}

Continuous Measurement



Binary Matrix



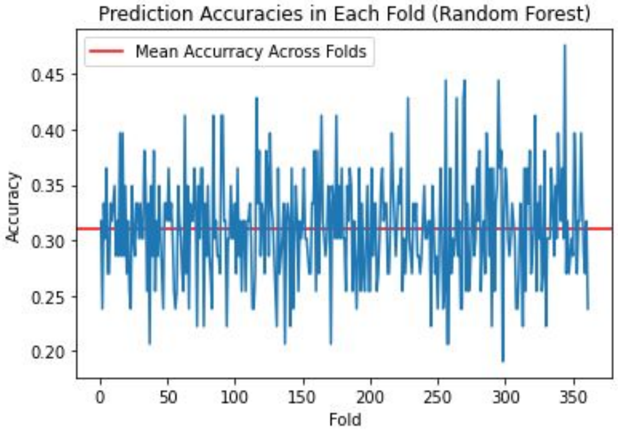
Plant Species (grouped by genus)

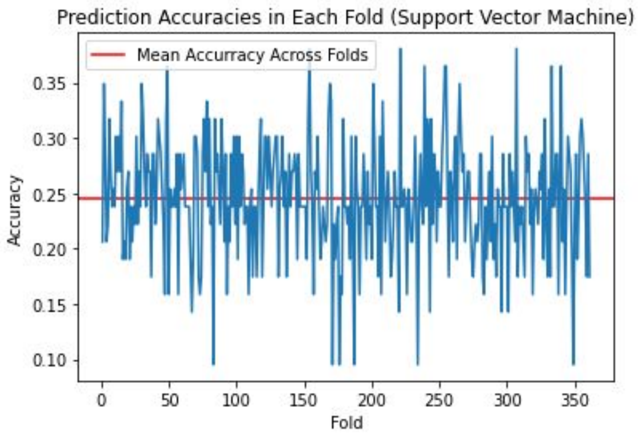
{'PO', 'TR', 'AC', 'FE', 'EL'}

Random Guess Accuracy: 20%

Mean prediction Accuracy using SVM: 24.57%

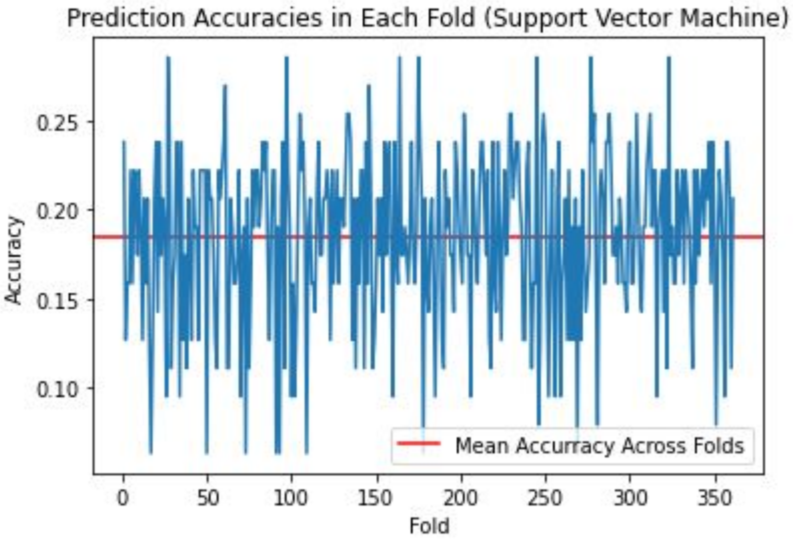
Mean prediction Accuracy using Random Forest: 31%





Gradient

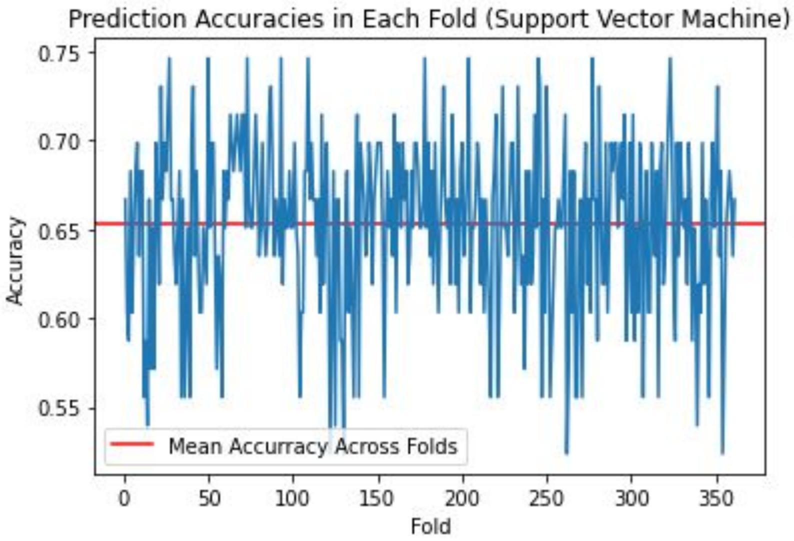
{'Avery', 'Treasury', 'Ruby', 'HunterHill', 'RP', 'Cinnamon', 'KP', 'JC2', 'Teocali', 'JC1'}

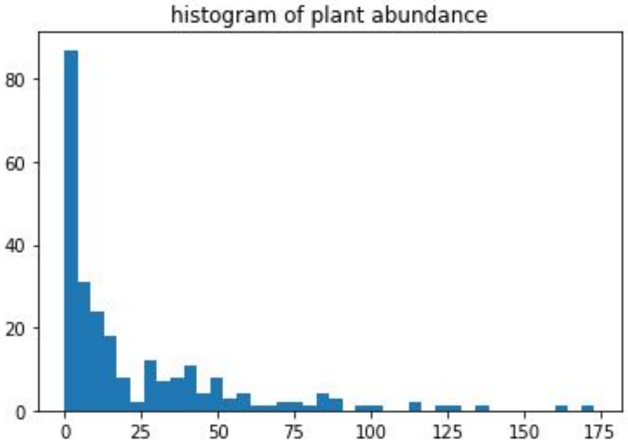


Plant Abundance (Ignoring Species and Genus)

Continuous measurements, generalized into 2 class: {Abundant, Not Abundant}

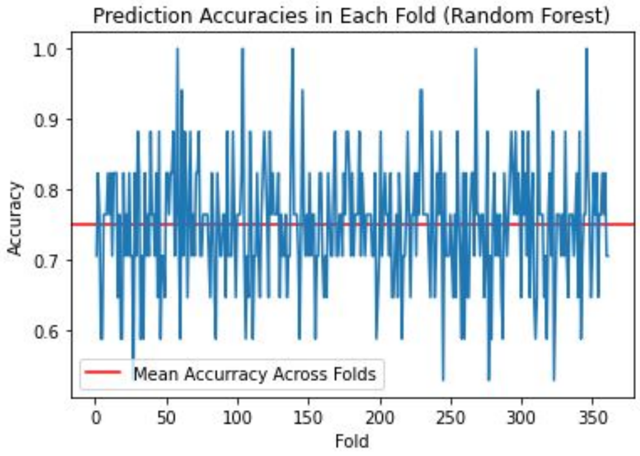
Currently using 5 as the cutoff.



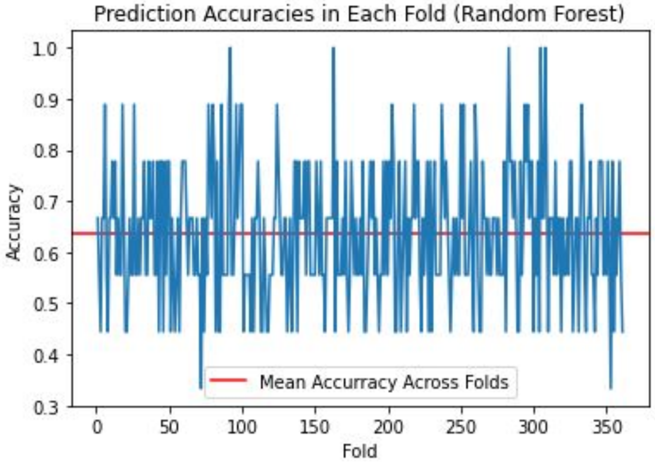


Plant Abundance by Genus

PO

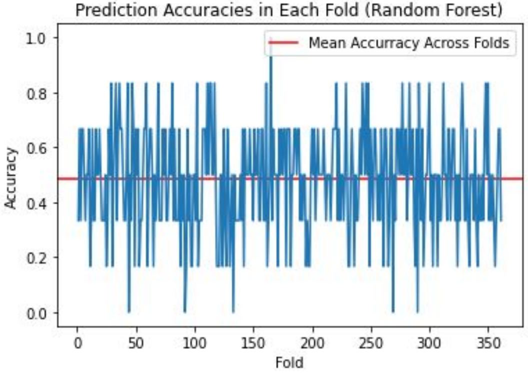


TR

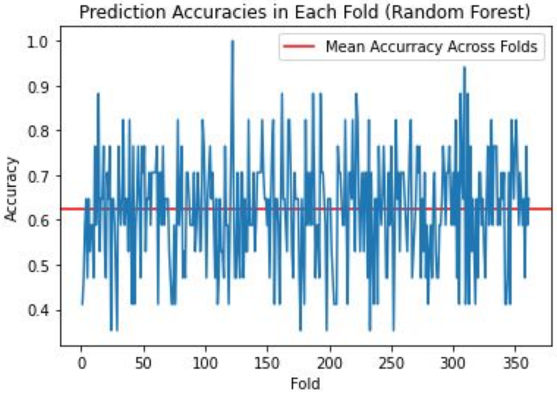


Plant Abundance by Genus

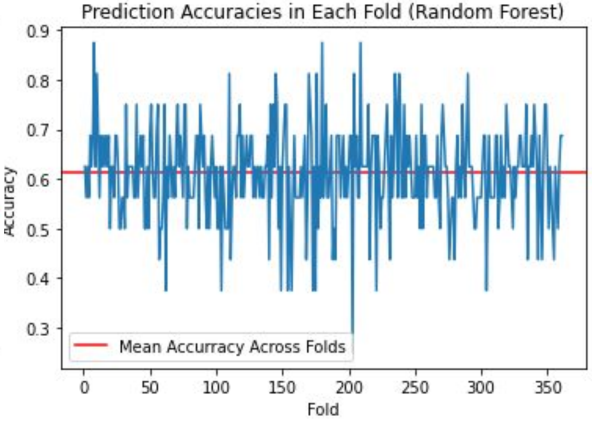
AC



FE



EL



Plant Abundance by Genus

Mean Accuracy Prediction by Genus:

PO: 0.751

TR: 0.638

FE: 0.623

EL: 0.6125346260387812

AC: 0.483

POPR

Sample Size is small (20), and extremely unbalanced:

SVM prediction with all ASVs:

Real Labels:

['Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Not Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Not Abundant']

Predictions:

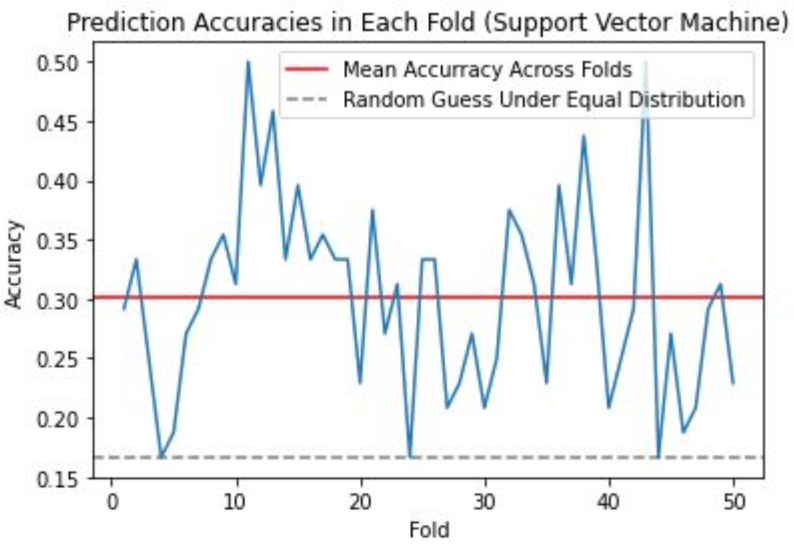
['Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant', 'Abundant']

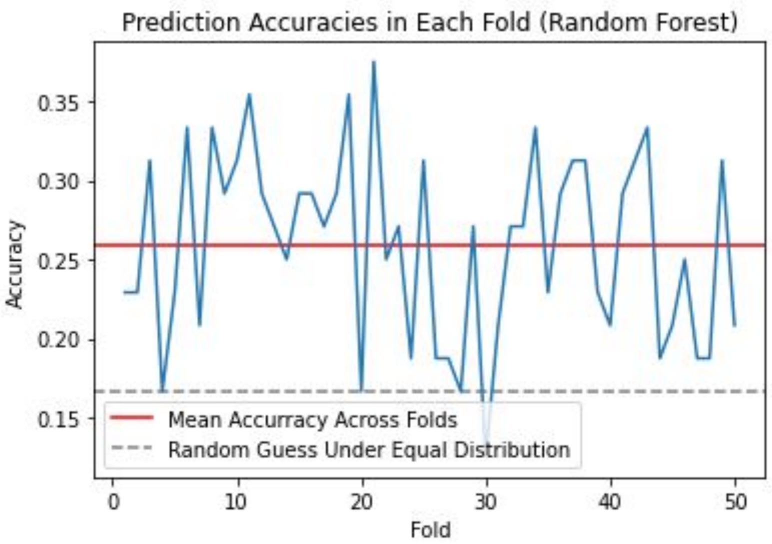
Species Prediction (One VS Others)

SVM, NB and RF all yield meaningless predictions (no true positives among all species and samples). With one false positive in the prediction of ELTR nad FETH.

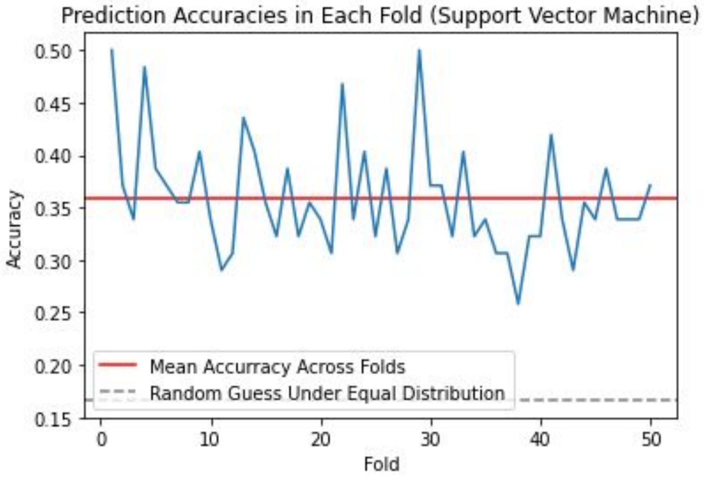
Most predictions are completely all negatives.

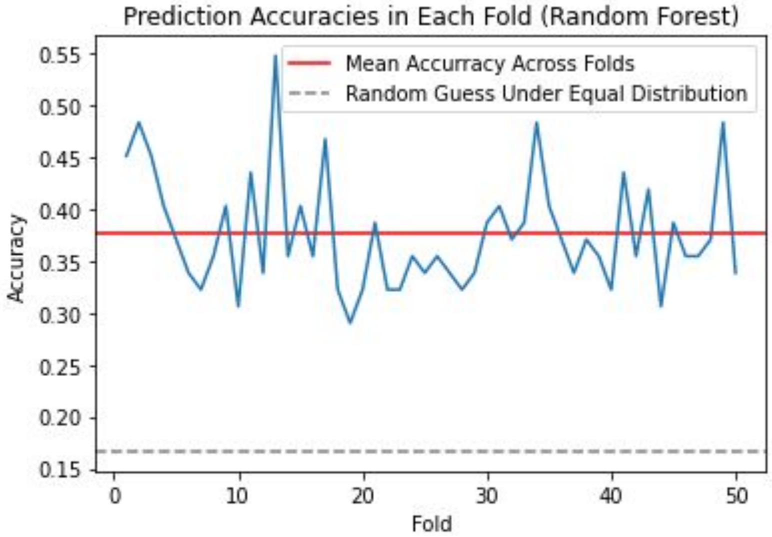
Leaves-Gradient



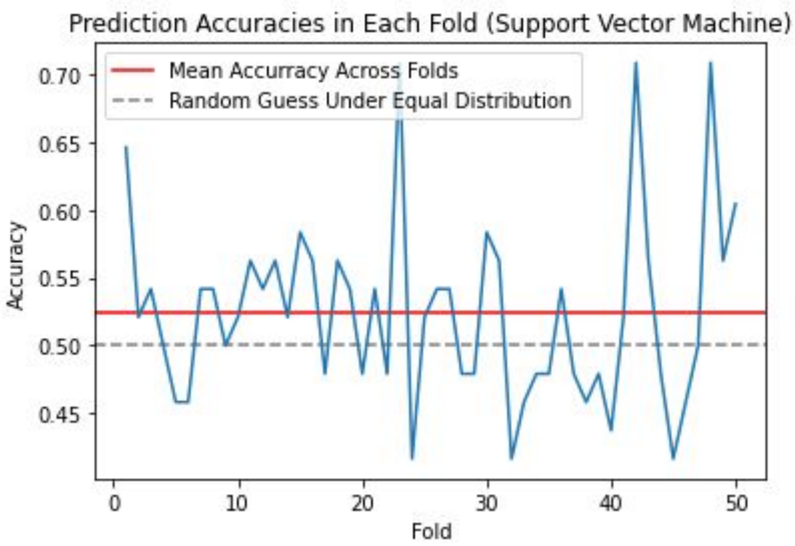


Root-Gradient

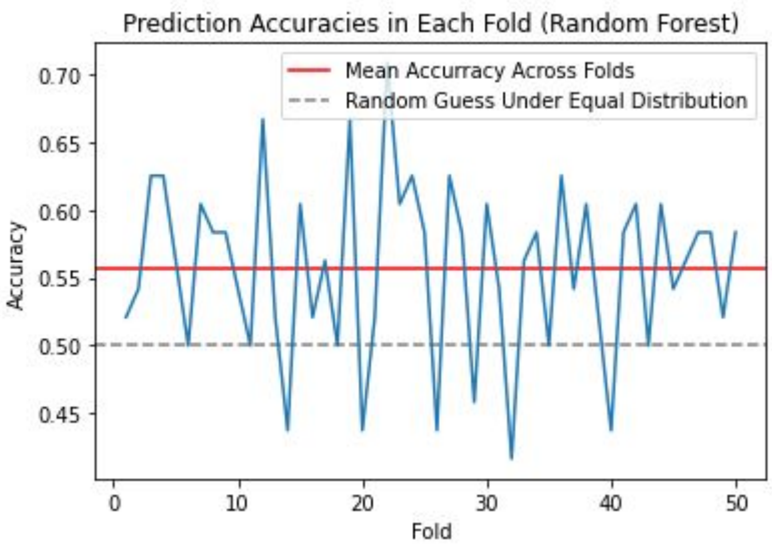




Leave-Richness

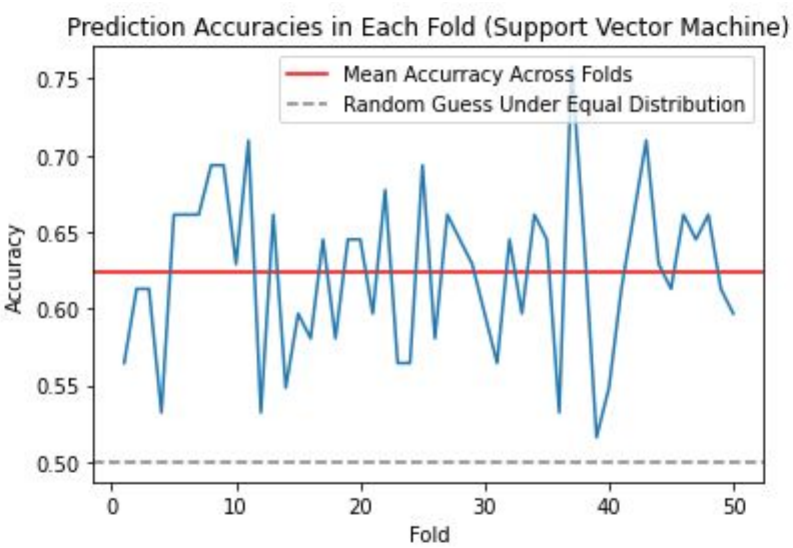


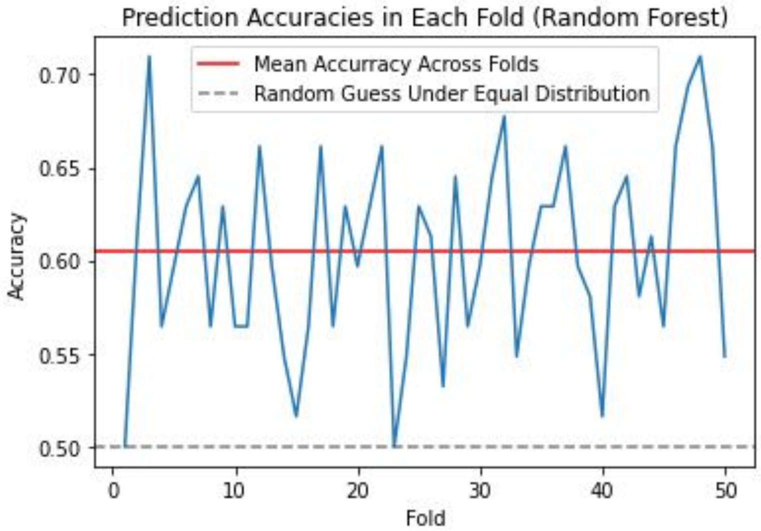
Prediction Accuracies in Each Fold (Support Vector Machine)



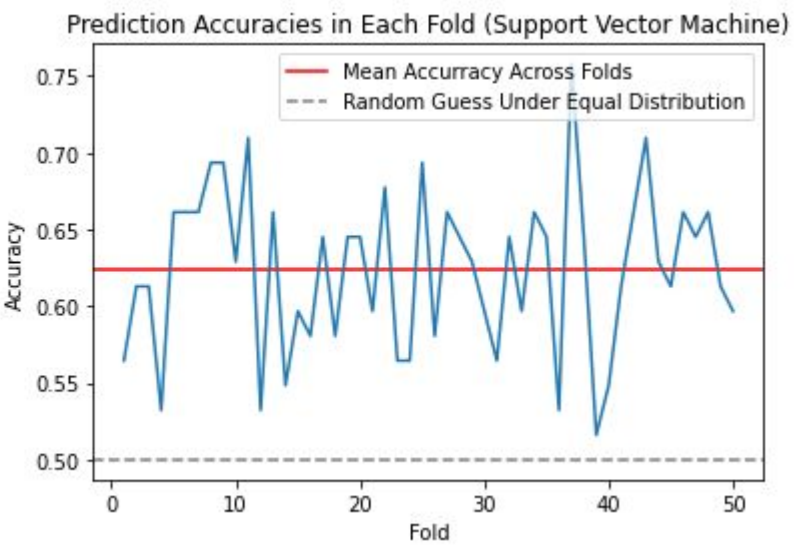
Prediction Accuracies in Each Fold (Random Forest)

Root-Richness

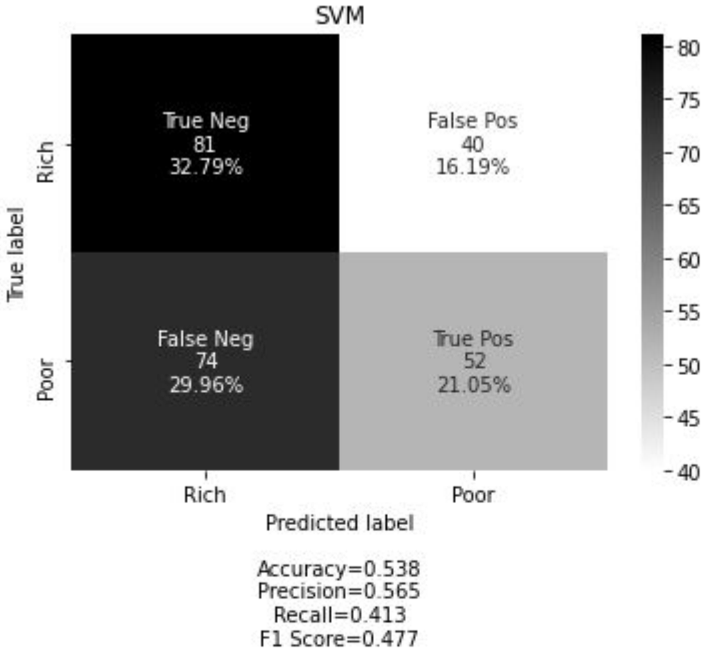




Richness using total-N (seeing minor signal so far)

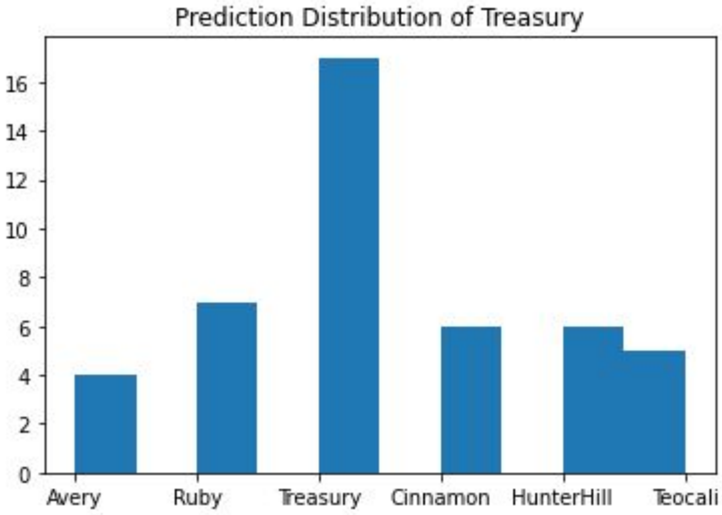


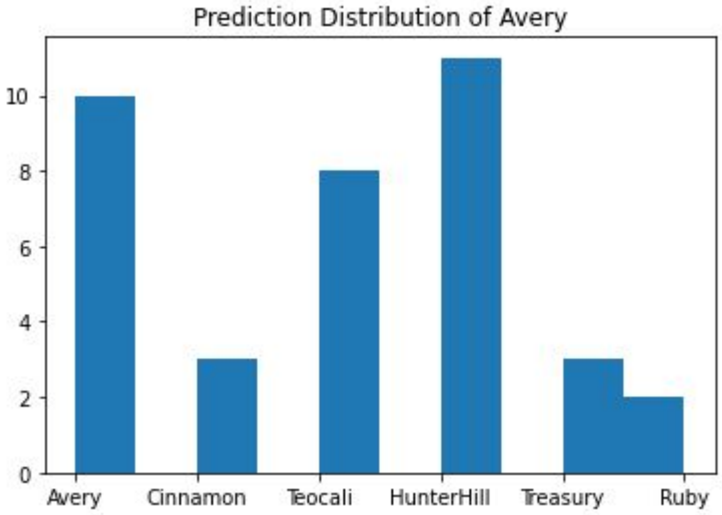
Using Top 30 Features



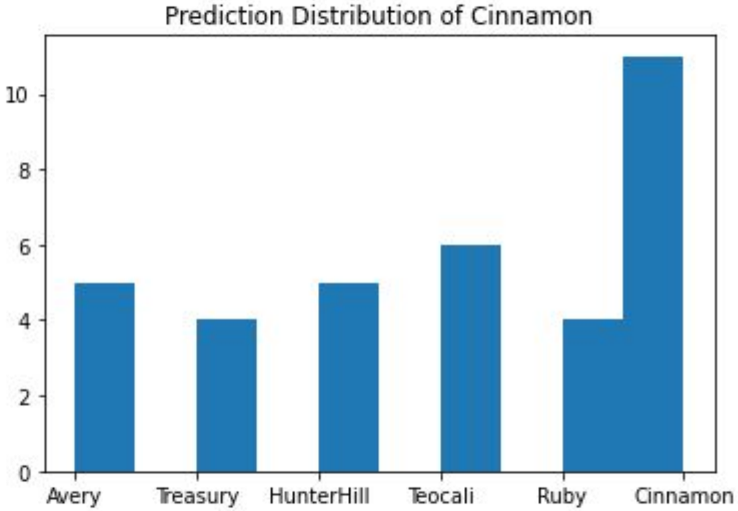
Accuracy=0.538
Precision=0.565
Recall=0.413
F1 Score=0.477

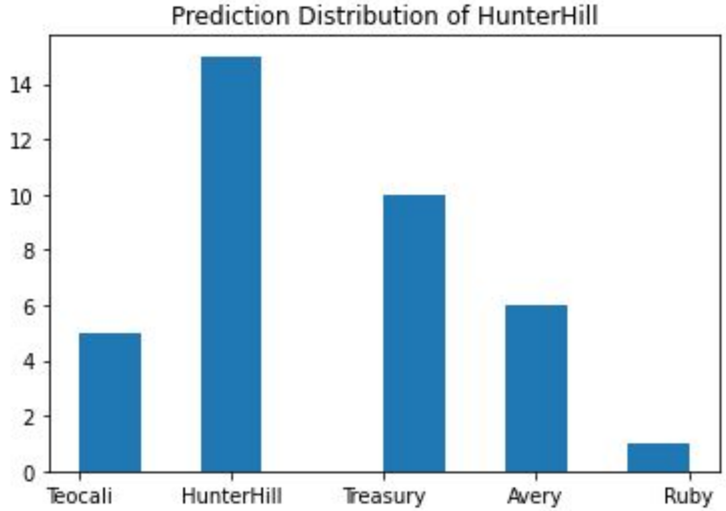
Gradient Prediction By Groups



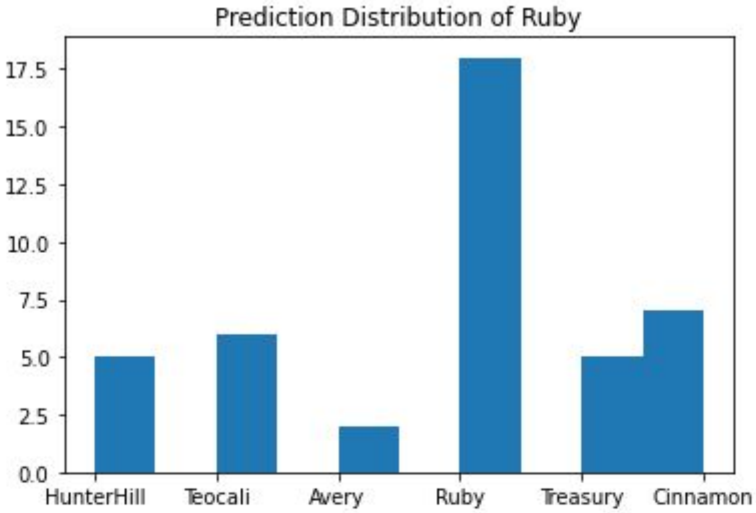


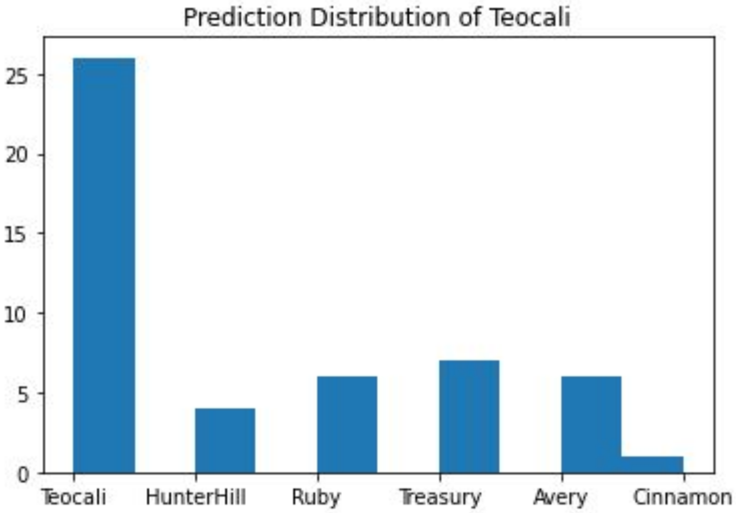
Gradient Prediction By Species



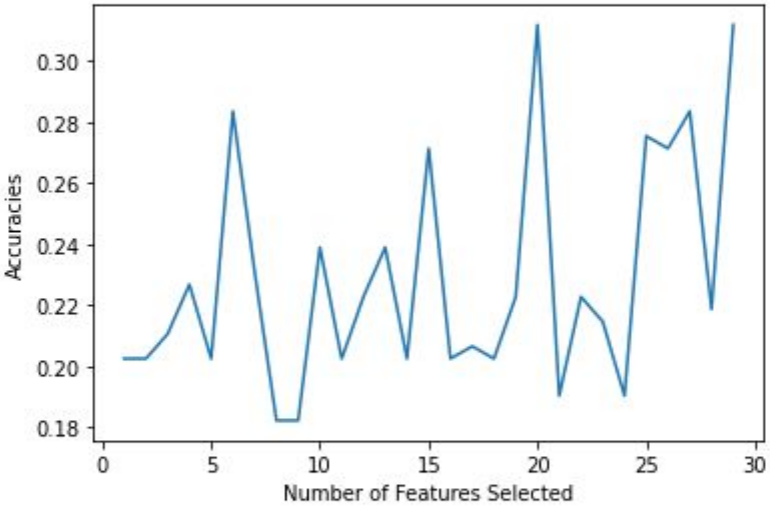


Gradient Prediction By Species



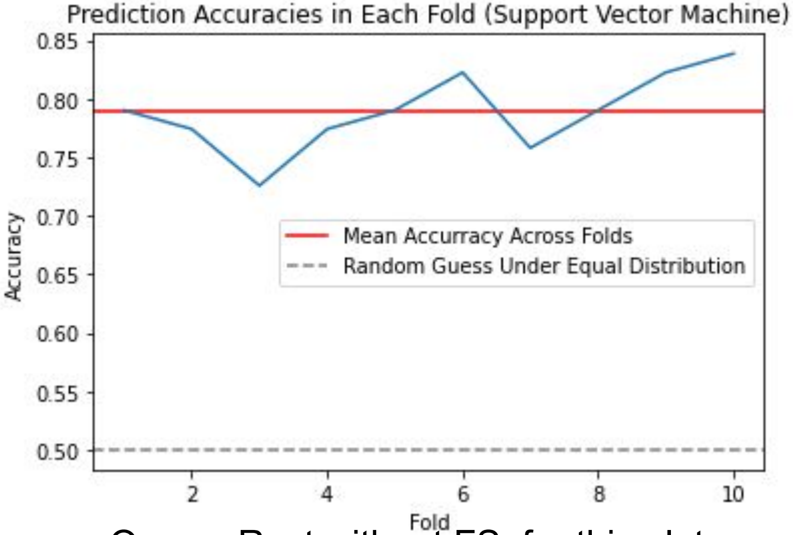


Feature Selection on Gradient



Treasury

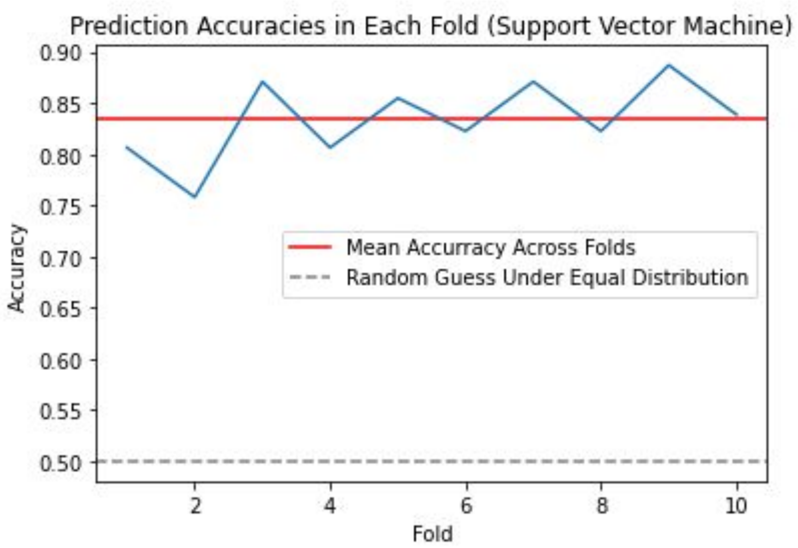
Mean Accuracy Across Folds 0.7887096774193548



One vs Rest without FS, for this plot, its binary labels of Treasury vs Not Treasury

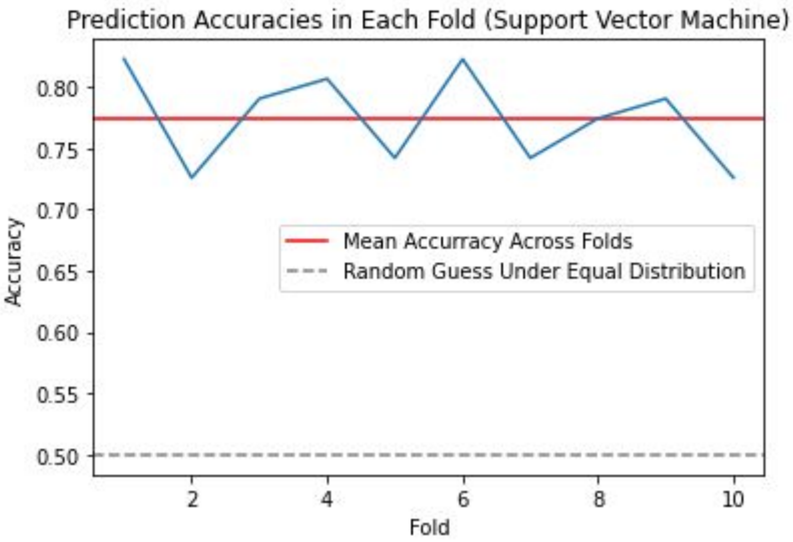
Cinnamon :

Mean Accuracy Across Folds 0.8338709677419354



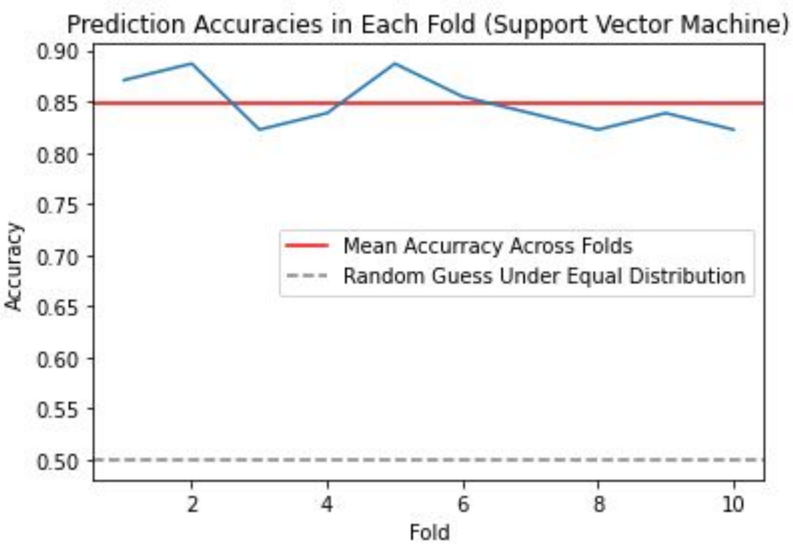
Avery

Mean Accuracy Across Folds 0.7741935483870968



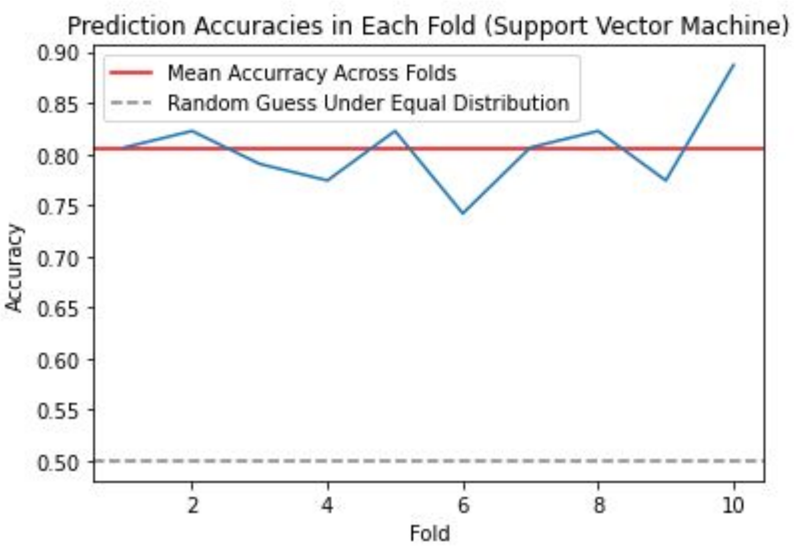
Hunterhill

Mean Accuracy Across Folds 0.8483870967741935



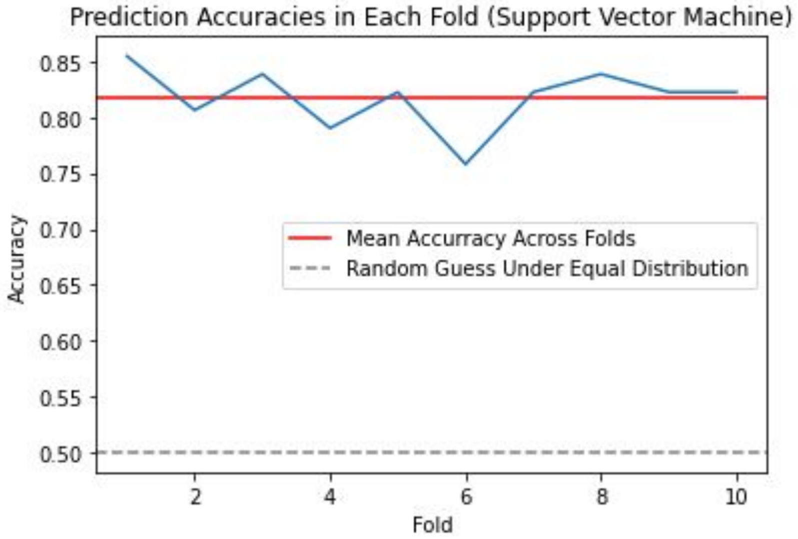
Teocali

Mean Accuracy Across Folds 0.8048387096774194

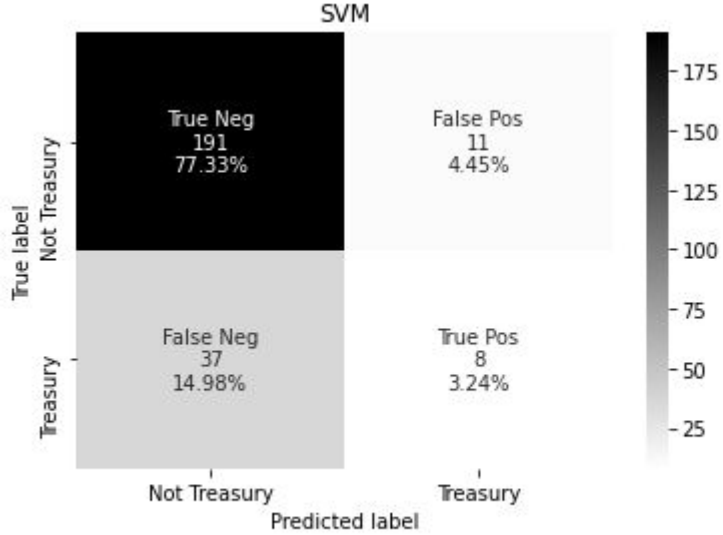


Ruby

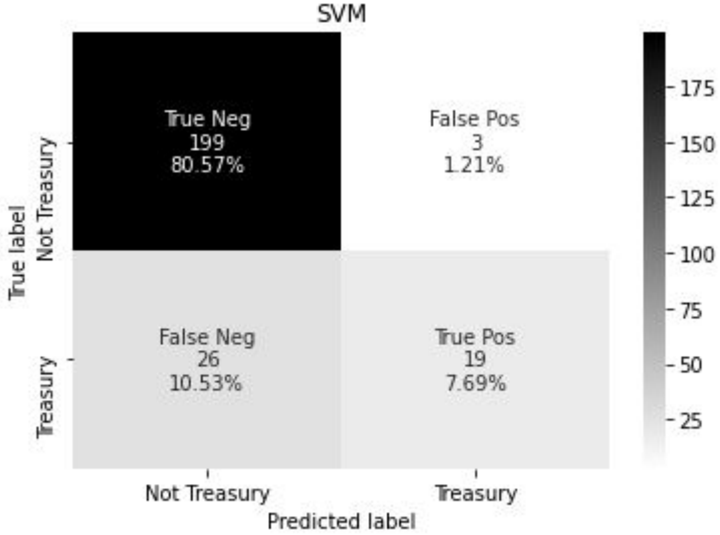
Mean Accuracy Across Folds 0.8177419354838711



Different Number of features impact different gradient prediction precision Top 30 features (left) vs Top 100 features(right) for Treasury, select K best with Chi2 scoring function.

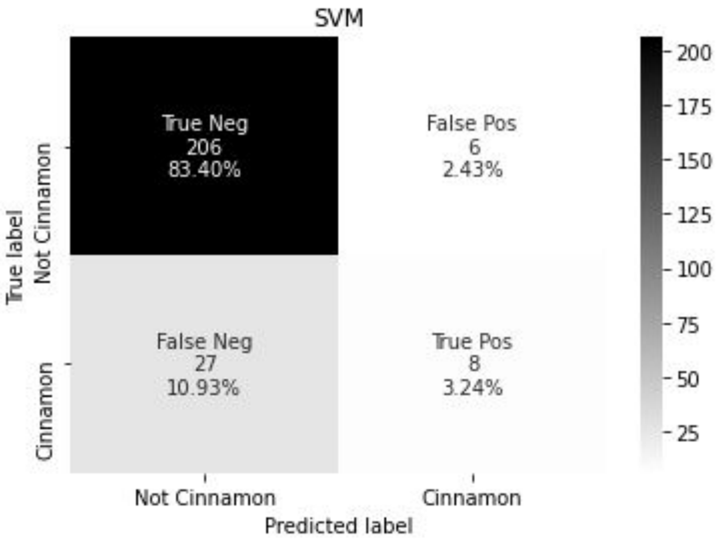


Accuracy=0.806
Precision=0.421
Recall=0.178
F1 Score=0.250

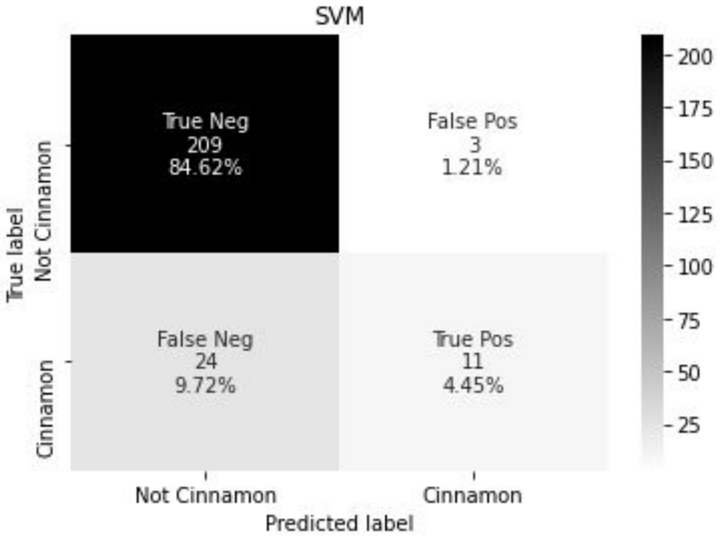


Accuracy=0.883
Precision=0.864
Recall=0.422
F1 Score=0.567

Cinnamon

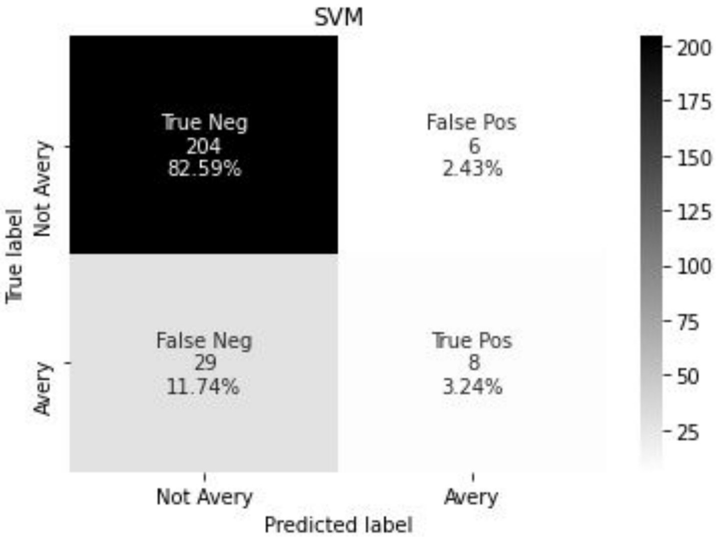


Accuracy=0.866
Precision=0.571
Recall=0.229
F1 Score=0.327

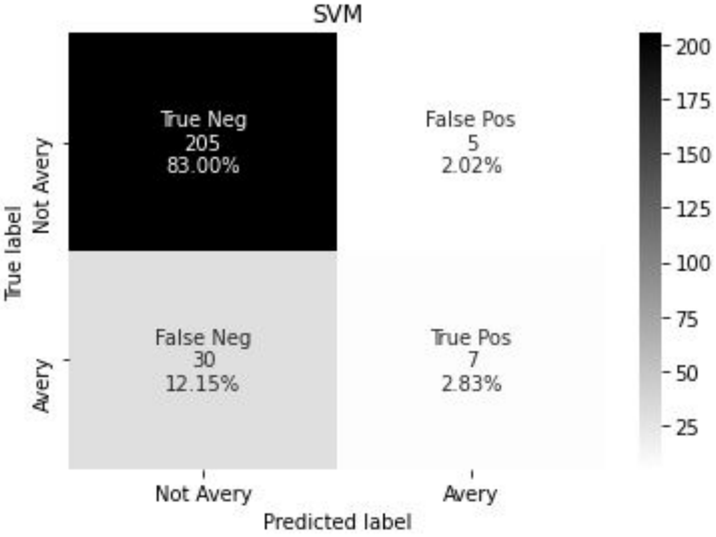


Accuracy=0.891
Precision=0.786
Recall=0.314
F1 Score=0.449

Avery

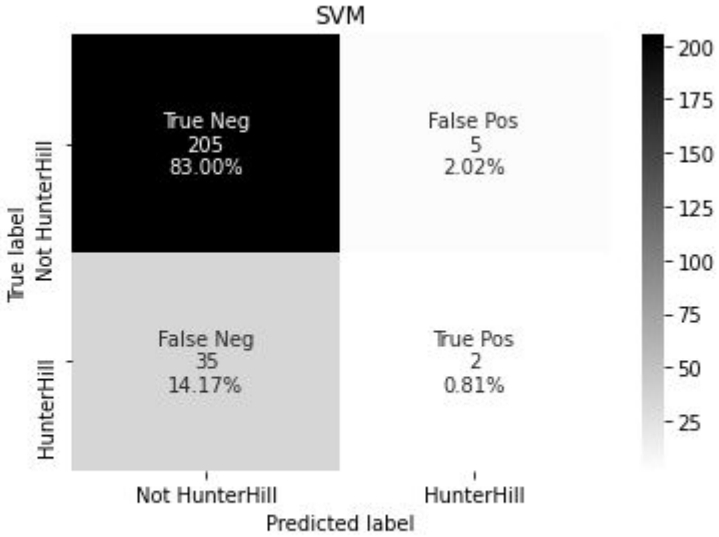


Accuracy=0.858
Precision=0.571
Recall=0.216
F1 Score=0.314

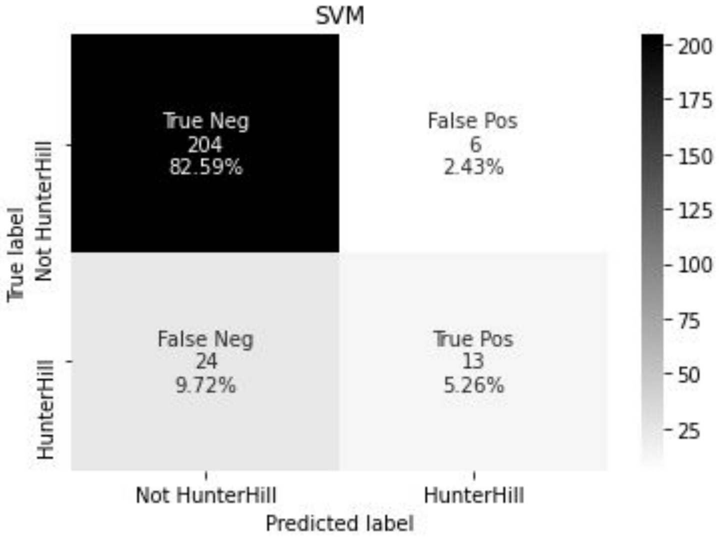


Accuracy=0.858
Precision=0.583
Recall=0.189
F1 Score=0.286

Hunterhill

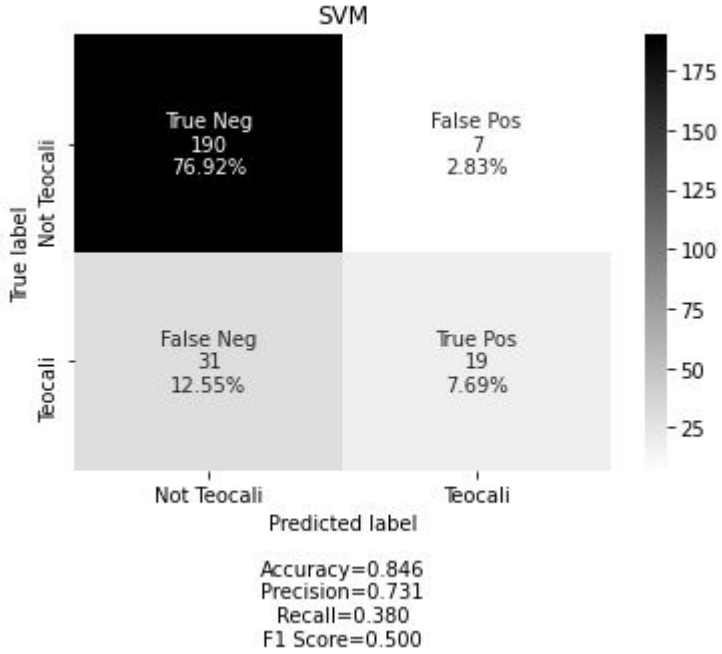


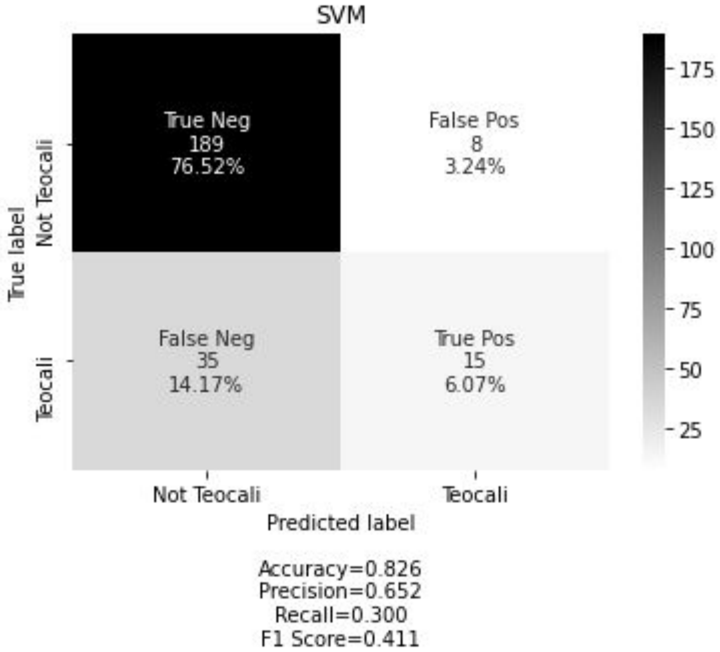
Accuracy=0.838
Precision=0.286
Recall=0.054
F1 Score=0.091



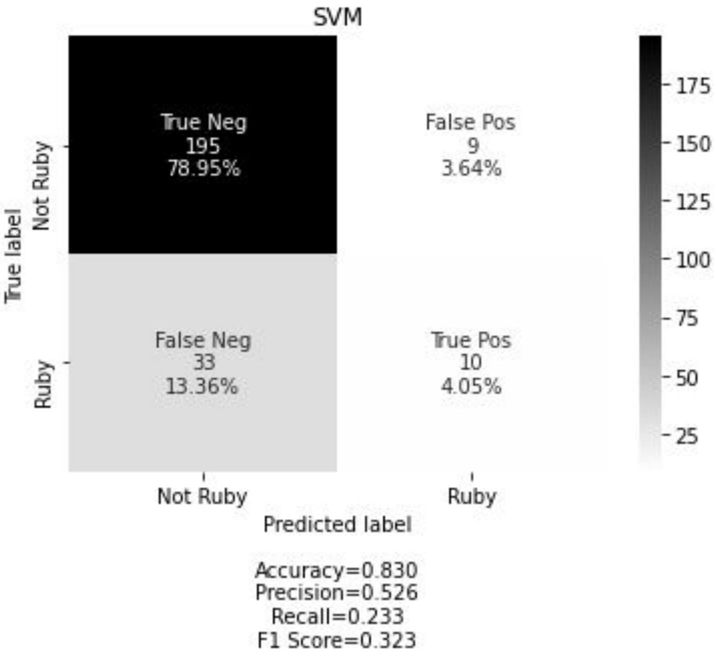
Accuracy=0.879
Precision=0.684
Recall=0.351
F1 Score=0.464

Teocali

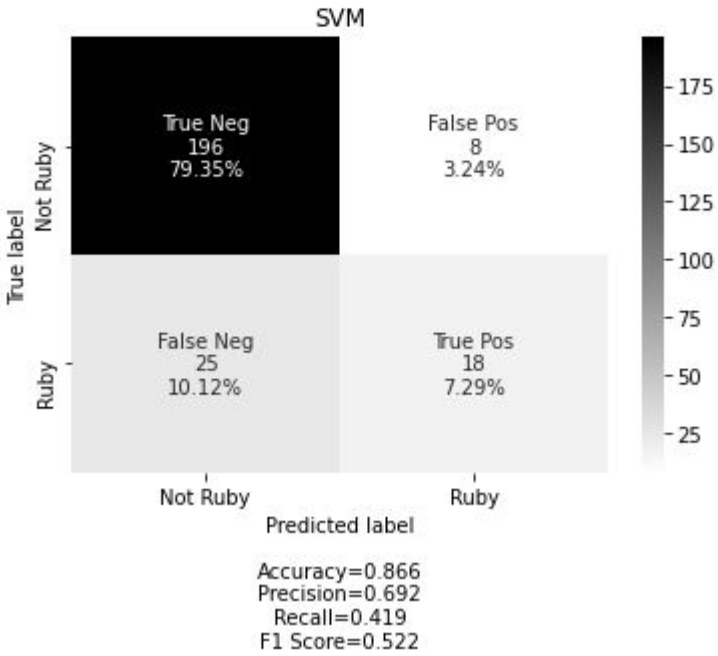

SVM
True Neg
190
76.92%
False Pos
7
2.83%
False Neg
31
12.55%
True Pos
19
7.69%
Not Teocali
Teocali
True label
Not Teocali
Teocali
Predicted label
175
150
125
100
75
50
25
Accuracy=0.846
Precision=0.731
Recall=0.380
F1 Score=0.500


SVM
True Neg
189
76.52%
False Pos
8
3.24%
False Neg
35
14.17%
True Pos
15
6.07%
Not Teocali
Teocali
True label
Not Teocali
Teocali
Predicted label
175
150
125
100
75
50
25
Accuracy=0.826
Precision=0.652
Recall=0.300
F1 Score=0.411

Ruby

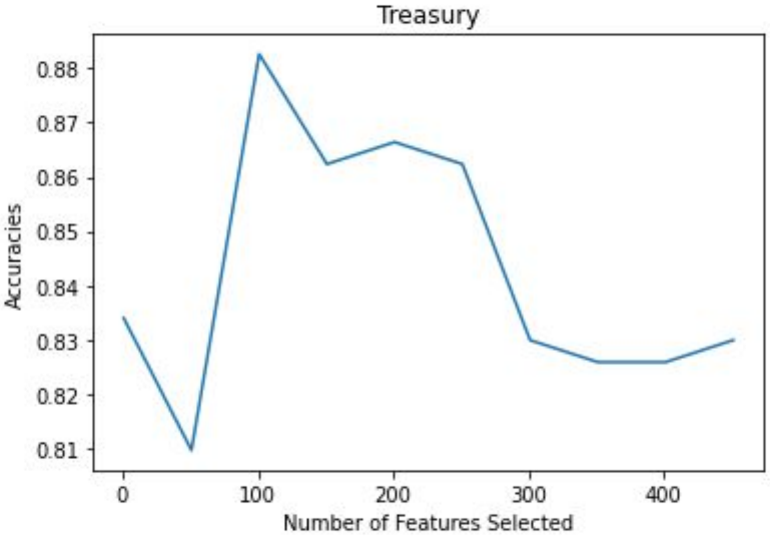


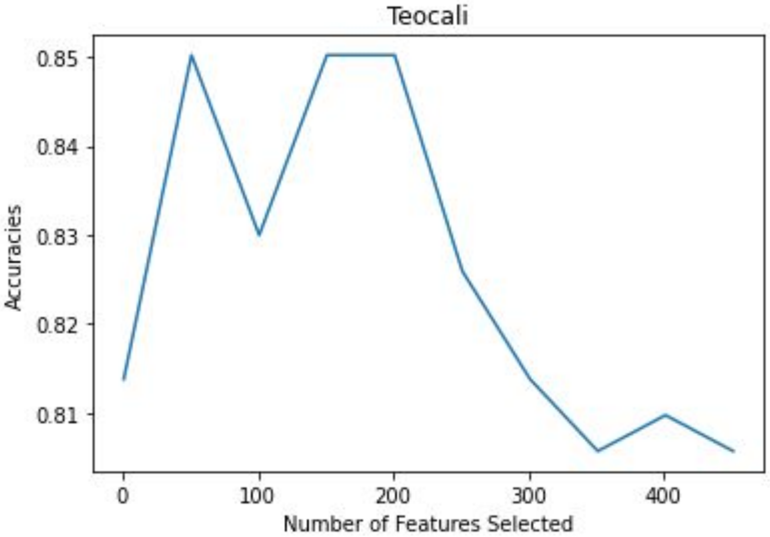
Accuracy=0.830
Precision=0.526
Recall=0.233
F1 Score=0.323

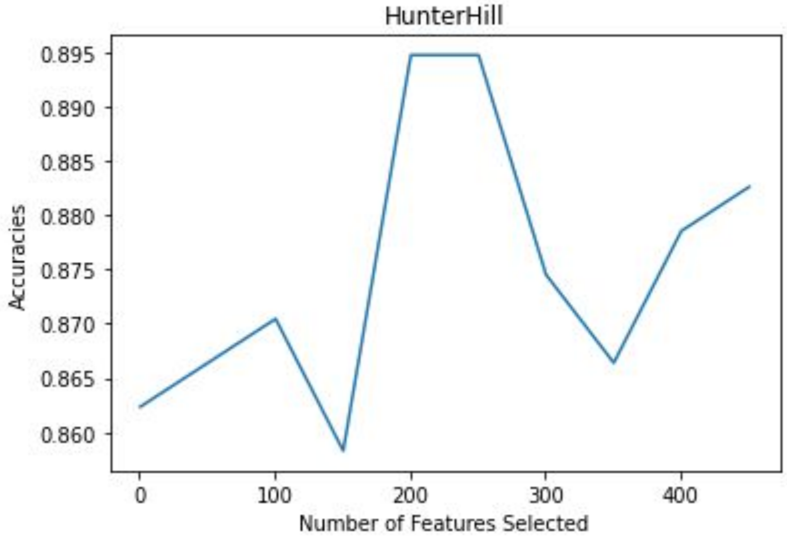


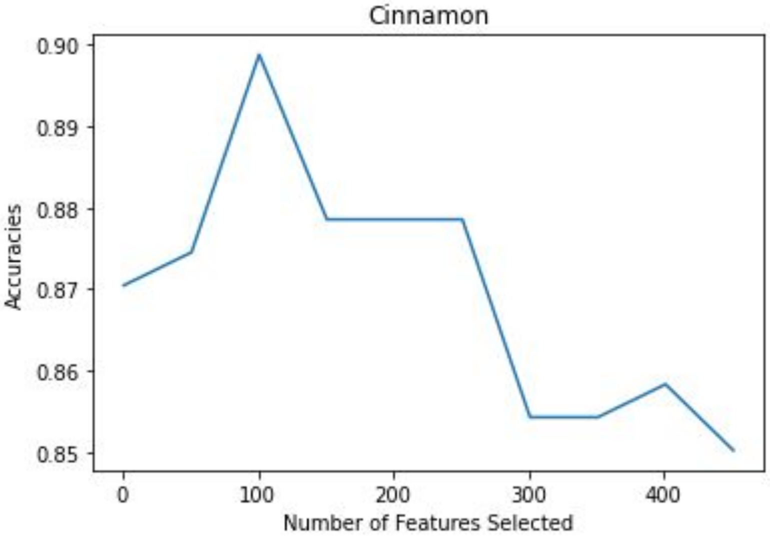
Accuracy=0.866
Precision=0.692
Recall=0.419
F1 Score=0.522

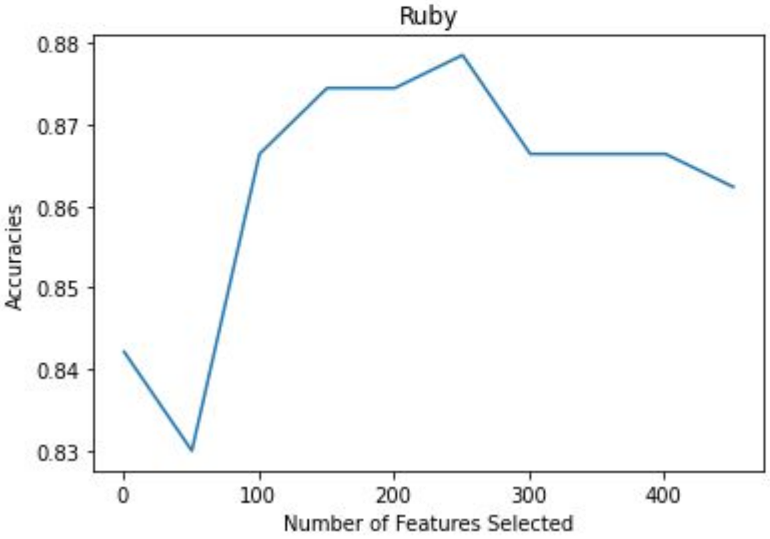
Number of Top Features Selected (increment of 50)

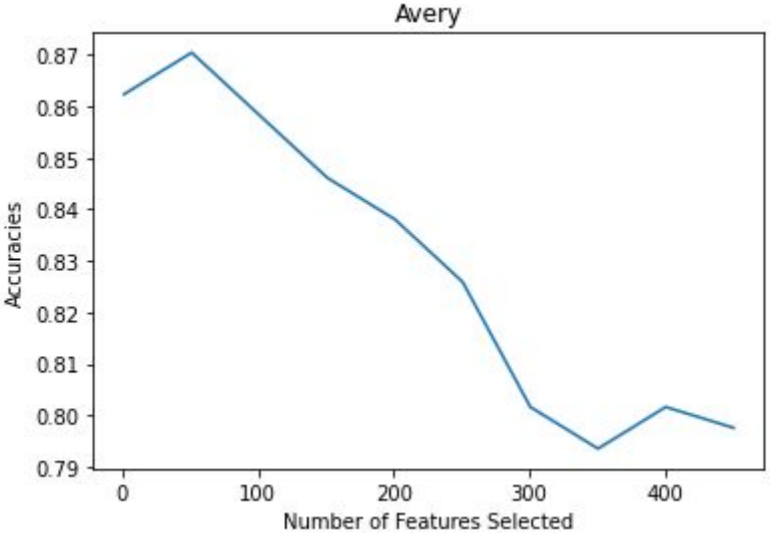



Teocali
0.85
0.84
0.83
0.82
0.81
Accuracies
0
100
200
300
400
Number of Features Selected

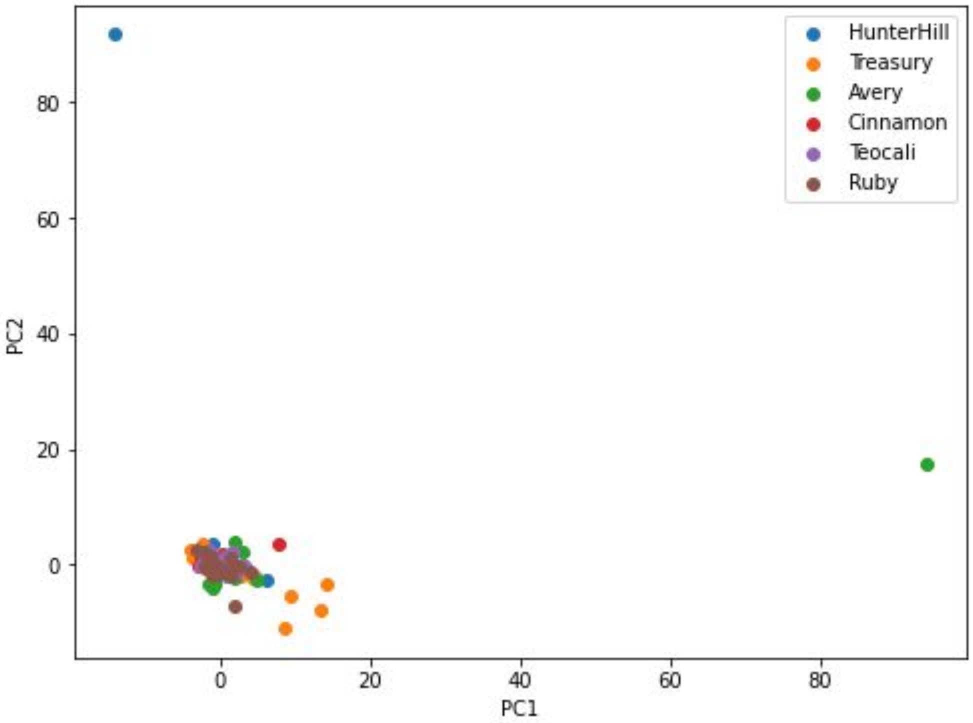

HunterHill
Accuracies
0.895
0.890
0.885
0.880
0.875
0.870
0.865
0.860
0
100
200
300
400
Number of Features Selected


Cinnamon
Accuracies
0.90
0.89
0.88
0.87
0.86
0.85
0
100
200
300
400
Number of Features Selected

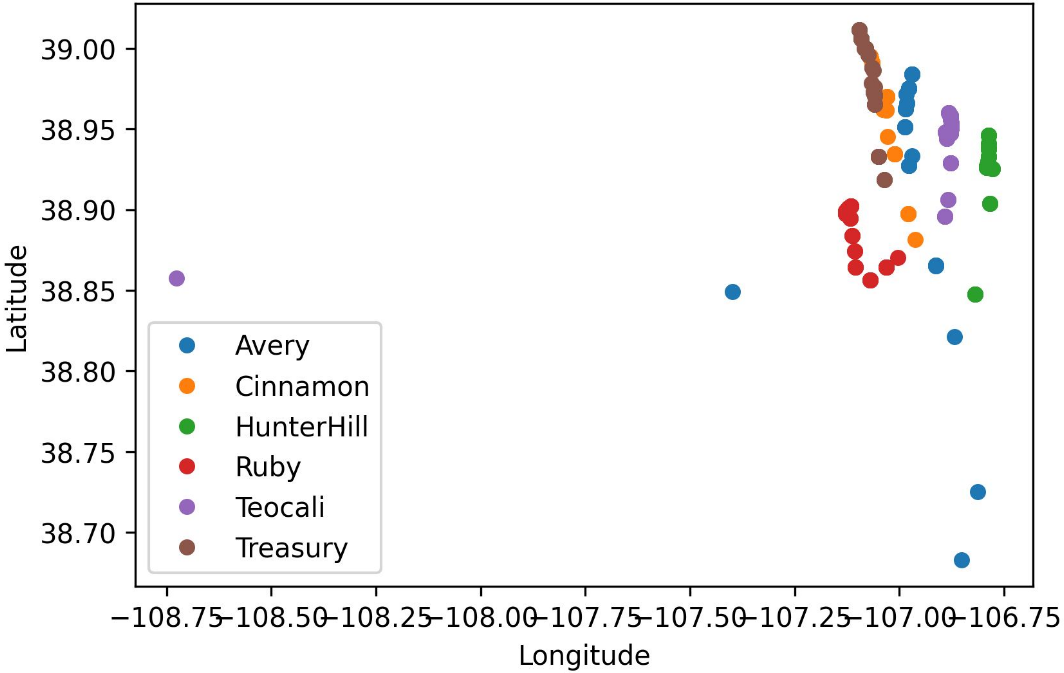

Ruby
0.88
0.87
0.86
0.85
0.84
0.83
Accuracies
0
100
200
300
400
Number of Features Selected


Avery
Accuracies
0.87
0.86
0.85
0.84
0.83
0.82
0.81
0.80
0.79
0
100
200
300
400
Number of Features Selected

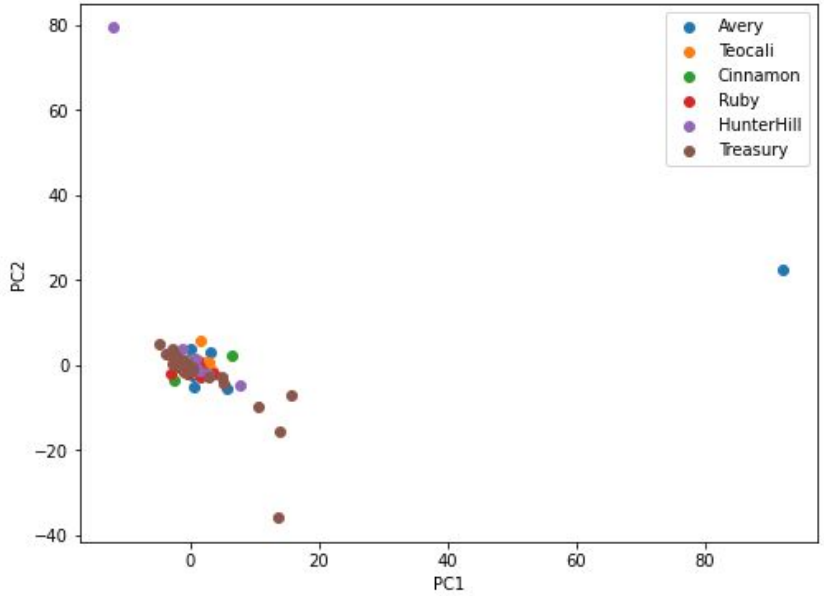
PCA Plot (Not much Variance Explained, as expected,about 1.5 percent variance explained on PC1 and PC2)



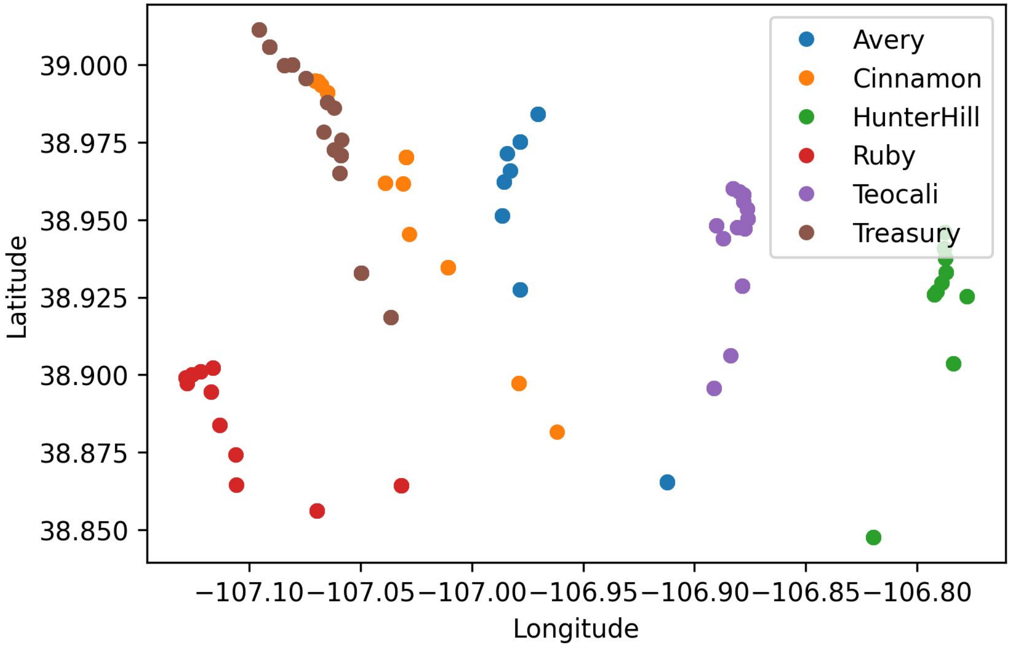
Sample Map (one odd Teocali Sample) (old version with ELEL)

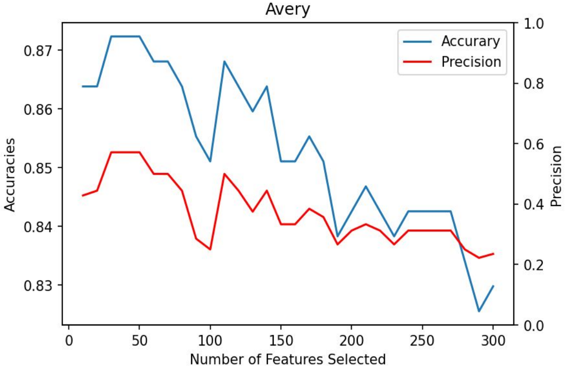


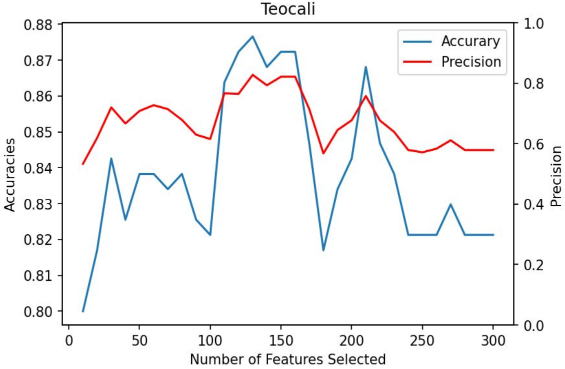
Updated PCA (slightly better, [0.01696325 0.01507139])

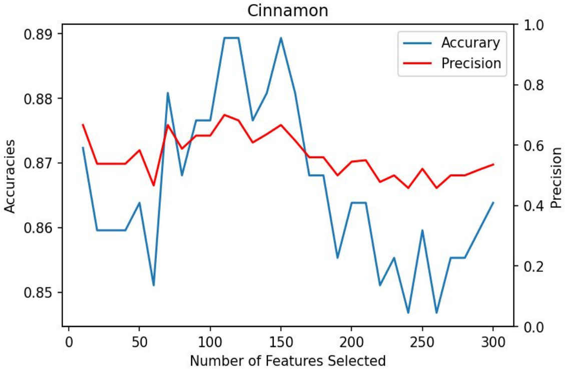


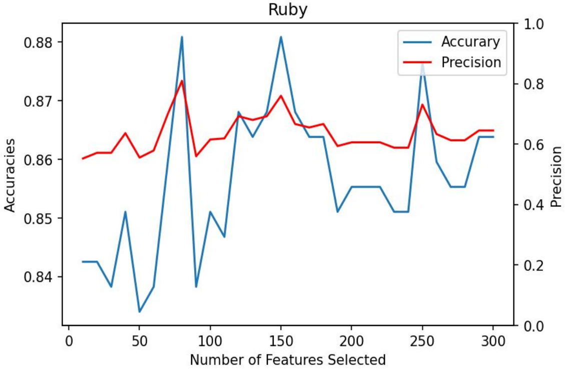
Updated Map (removed ELEL from data)

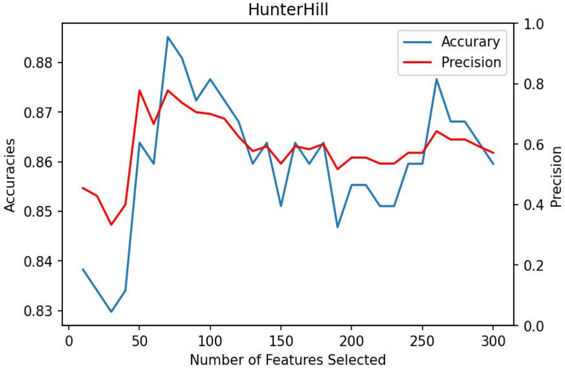


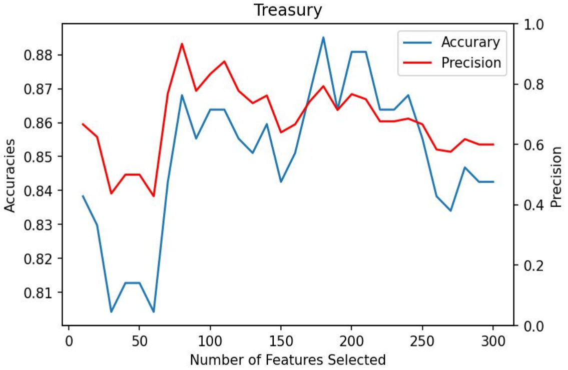

Avery
Accurary
Precision
Accuracies
Precision
Number of Features Selected
0.87
0.86
0.85
0.84
0.83
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300

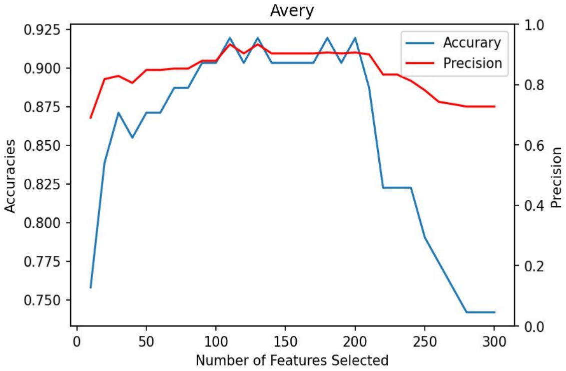

Teocali
Accurary
Precision
Accuracies
Precision
Number of Features Selected
0.88
0.87
0.86
0.85
0.84
0.83
0.82
0.81
0.80
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300

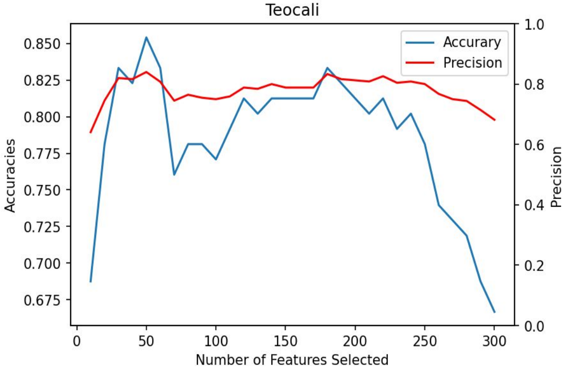

Cinnamon
Accurary
Precision
Accuracies
Precision
Number of Features Selected
0.89
0.88
0.87
0.86
0.85
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300

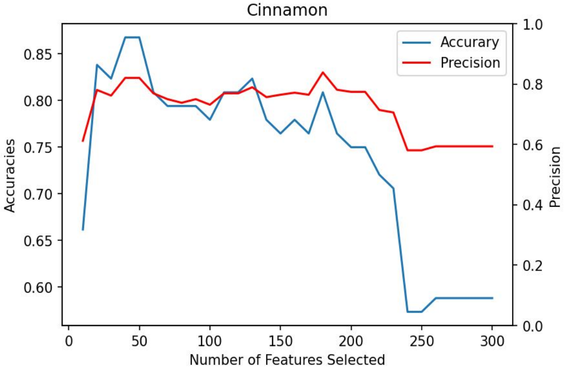

Ruby
Accurary
Precision
Accuracies
Precision
Number of Features Selected
0.88
0.87
0.86
0.85
0.84
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300

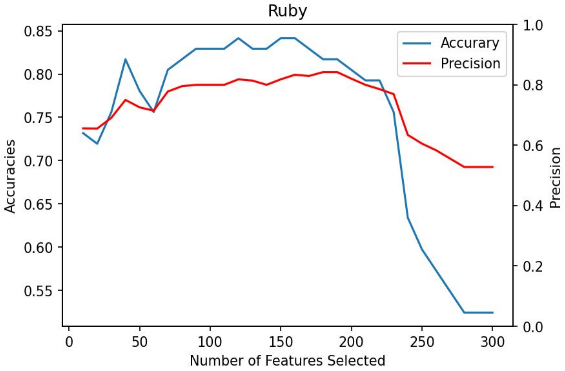

HunterHill
Accurary
Precision
Accuracies
Precision
Number of Features Selected
0.88
0.87
0.86
0.85
0.84
0.83
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300

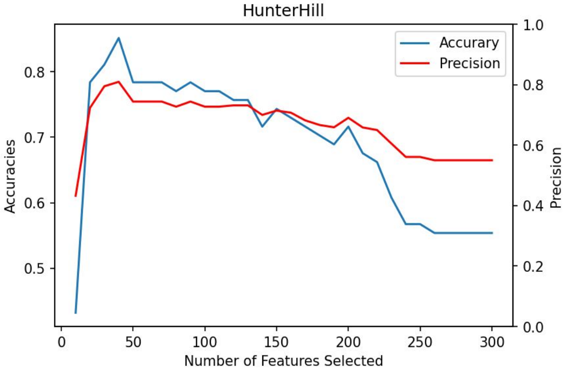

Treasury
Accurary
Precision
Accuracies
Precision
Number of Features Selected
0.88
0.87
0.86
0.85
0.84
0.83
0.82
0.81
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300

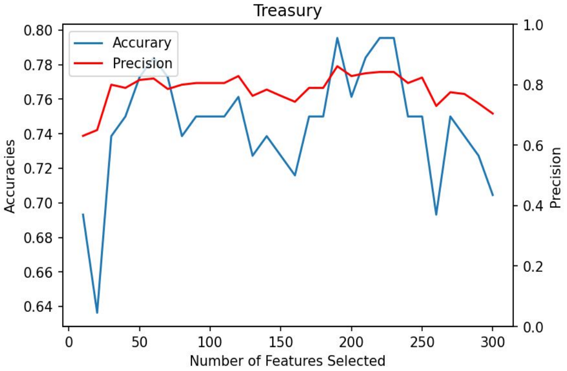

Avery
0.925
0.900
0.875
0.850
0.825
0.800
0.775
0.750
Accuracies
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0
50
100
150
200
250
300
Number of Features Selected
Accurary
Precision


Teocali
0.850
0.825
0.800
0.775
0.750
0.725
0.700
0.675
Accuracies
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0
50
100
150
200
250
300
Number of Features Selected
Accurary
Precision

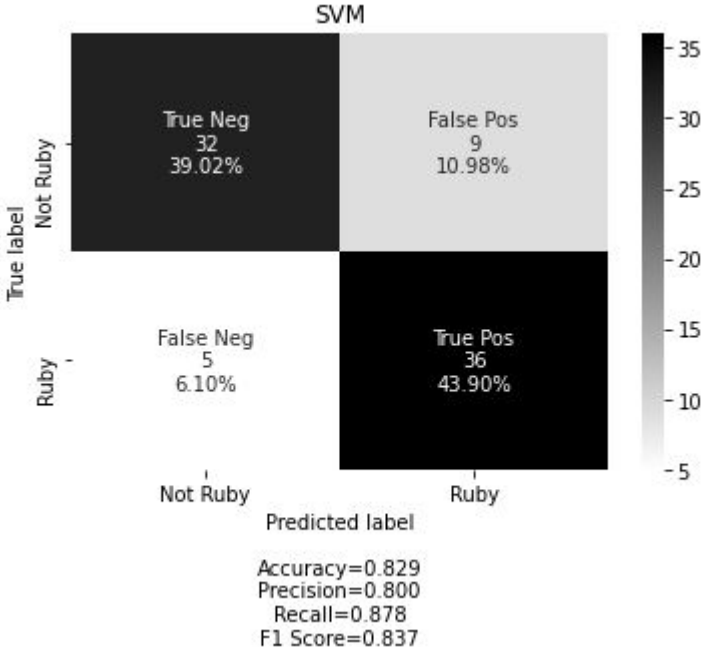

Cinnamon
0.85
0.80
0.75
0.70
0.65
0.60
Accuracies
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0
50
100
150
200
250
300
Number of Features Selected
Accurary
Precision


Ruby
0.85
0.80
0.75
0.70
0.65
0.60
0.55
Accuracies
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0
50
100
150
200
250
300
Number of Features Selected
Accurary
Precision

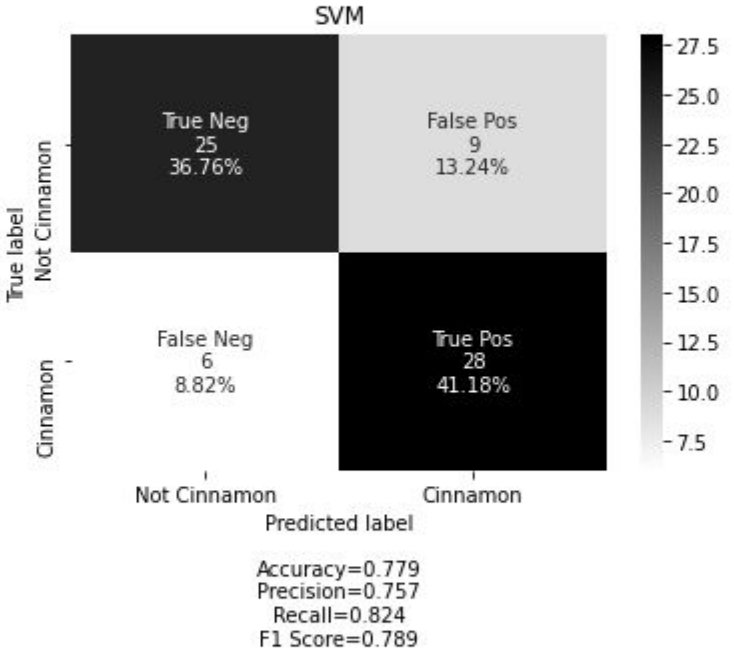

HunterHill
0.8
0.7
0.6
0.5
Accuracies
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0
50
100
150
200
250
300
Number of Features Selected
Accurary
Precision


Treasury
0.80
0.78
0.76
0.74
0.72
0.70
0.68
0.66
0.64
Accuracies
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0
50
100
150
200
250
300
Number of Features Selected
Accurary
Precision

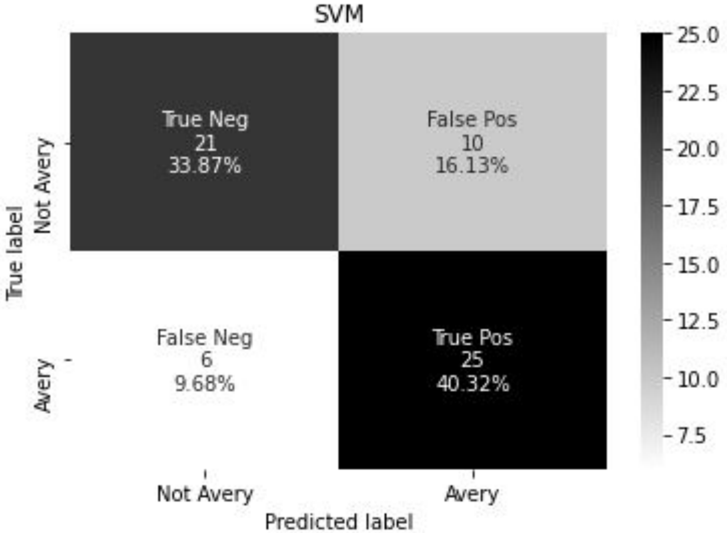
RUBY



CINNAMON

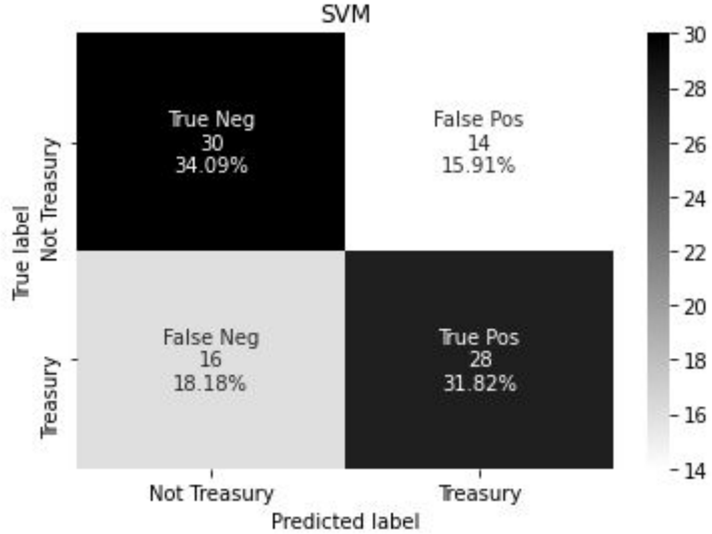


Avery



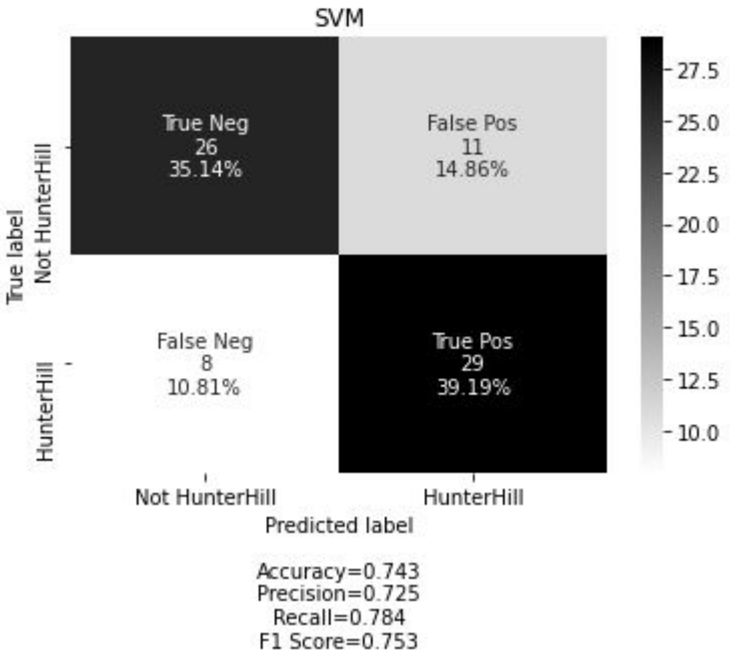
Accuracy=0.742
Precision=0.714
Recall=0.806
F1 Score=0.758

Treasury



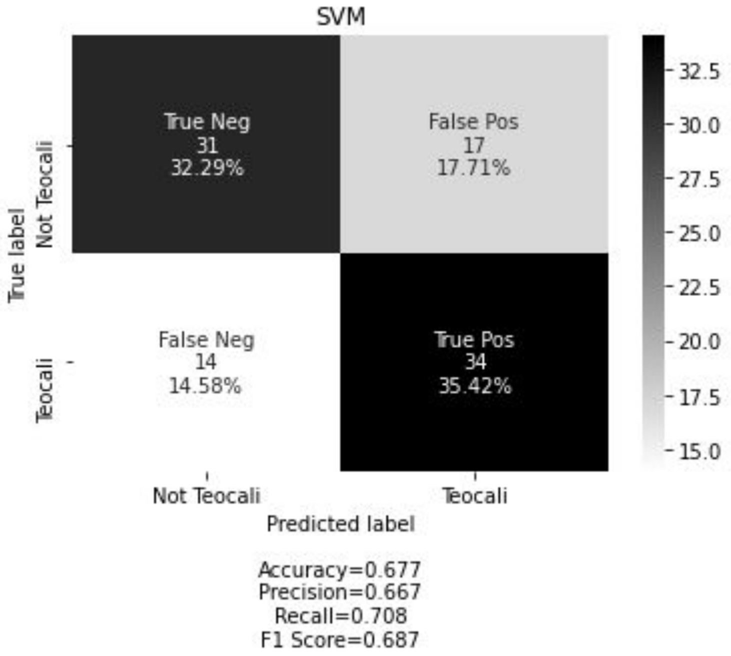
Accuracy=0.659
Precision=0.667
Recall=0.636
F1 Score=0.651

Hunterhill



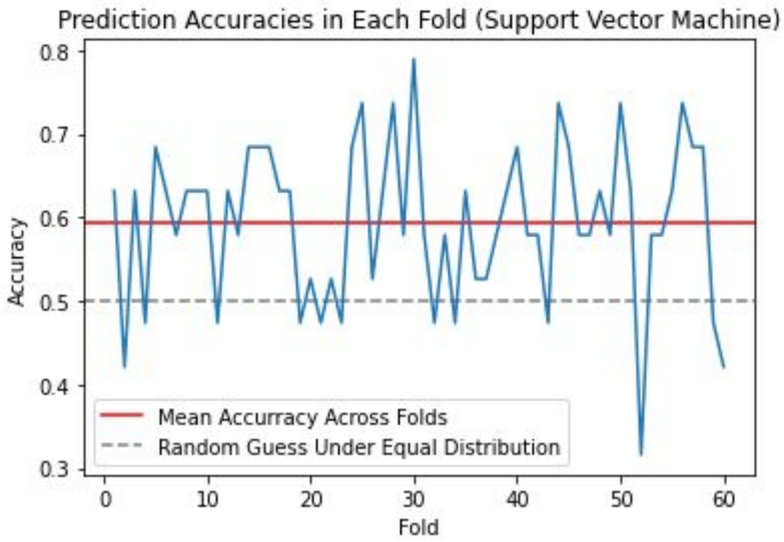
Accuracy=0.743
Precision=0.725
Recall=0.784
F1 Score=0.753

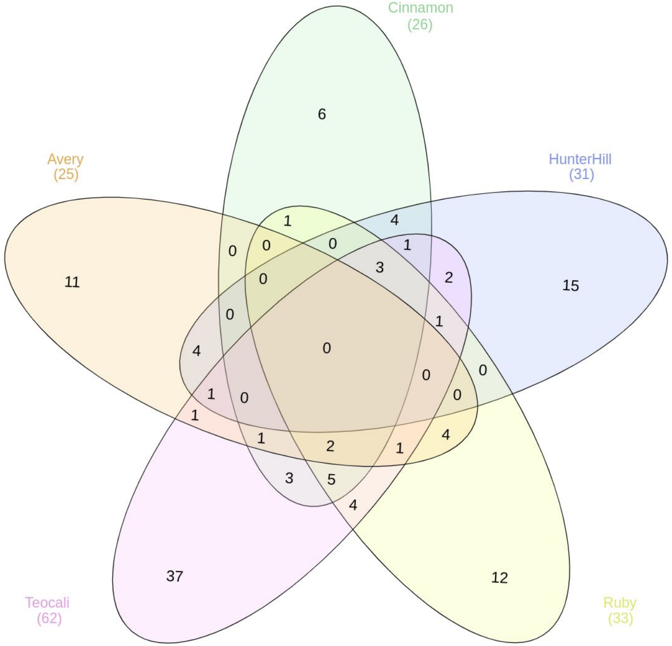
Teocali

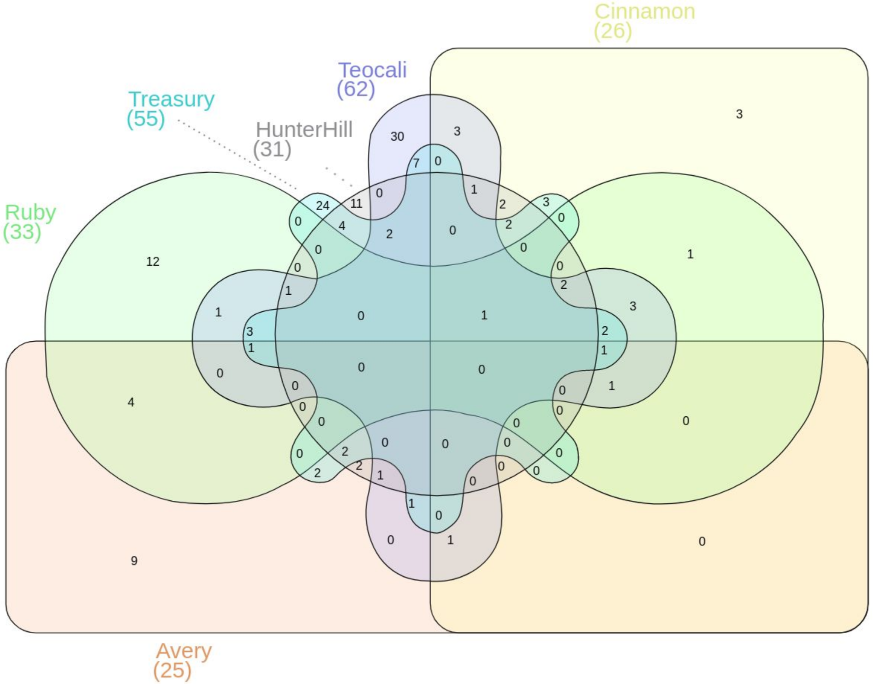


Accuracy=0.677
Precision=0.667
Recall=0.708
F1 Score=0.687

After including 20 percent noise




Cinnamon
(26)
Avery
(25)
HunterHill
(31)
Teocali
(62)
Ruby
(33)
6
11
15
37
12
1
4
0
0
0
1
0
3
2
0
0
1
0
4
0
1
0
0
4
1
0
1
2
1
3
5
4


Cinnamon
(26)
Teocali
(62)
Treasury
(55)
HunterHill
(31)
Ruby
(33)
Avery
(25)
30
3
3
7
0
24
11
0
1
2
3
0
4
2
0
2
0
12
1
0
0
0
2
1
1
1
3
3
0
1
2
1
3
1
0
0
0
4
0
1
0
0
0
0
0
0
0
2
0
0
0
0
2
2
1
0
0
0
0
1
0
9
0
1
0