Tree Health Results

Results for Paper

Best Models

Model	Accuracy	Macro F1	Weighted F1
RF Classifier	**0.917 (0.010)**	**0.851 (0.100)**	**0.917 (0.010)**
GB Classifier	0.911 (0.008)	0.843 (0.101)	0.911 (0.008)

5 fold train / test splits, on all data, oversampling, using manual / grid search best params for each model

Other Models
(all Oversampling,
one run)

1,008,859
Samples

3,091
Samples

Model	Accuracy	Macro F1	Weighted F1
Linear Regressor	0.01 (R2 score)	-	-
RF Regressor	0.42 (R2 score)	-	-
Logistic Regressor	0.19	0.12	0.12
RF Classifier	0.81	0.76	0.81
RF Classifier + Accessories	0.79	0.71	0.79
RF Classifier + Accessories	0.57	0.29	0.56
RF Classifier + PLUTO	0.92	0.89	0.92
RF Classifier + Accessories + PLUTO	**0.94**	**0.92**	**0.94**
GB Classifier + Accessories + PLUTO	**0.94**	**0.92**	**0.94**
SVM + Accessories + PLUTO	0.84	0.64	0.88

All Results

With features:
Latitude, longitude, zipcode, species, diameter

Logistic Regression Unbalanced

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	603
1.0	0.00	0.00	0.00	5173
2.0	0.52	0.85	0.64	103354
3.0	0.49	0.17	0.25	92642
accuracy			0.51	201772
macro avg	0.25	0.26	0.22	201772
weighted avg	0.49	0.51	0.45	201772

Logistic Regression Oversampled

	precision	recall	f1-score	support
0.0	0.00	0.60	0.01	603
1.0	0.02	0.11	0.04	5173
2.0	0.56	0.36	0.43	103354
3.0	0.00	0.00	0.00	92642
accuracy			0.19	201772
macro avg	0.15	0.27	0.12	201772
weighted avg	0.29	0.19	0.22	201772

Logistic Regression Undersampled

	precision	recall	f1-score	support
0.0	0.00	0.60	0.01	603
1.0	0.03	0.40	0.05	5173
2.0	0.45	0.06	0.10	103354
3.0	0.00	0.00	0.00	92642
accuracy			0.04	201772
macro avg	0.12	0.26	0.04	201772
weighted avg	0.23	0.04	0.05	201772

Random Forest Classifier Unbalanced

	precision	recall	f1-score	support
0.0	0.79	0.38	0.51	603
1.0	0.75	0.47	0.57	5173
2.0	0.78	0.80	0.79	103354
3.0	0.77	0.76	0.77	92642
accuracy			0.77	201772
macro avg	0.77	0.60	0.66	201772
weighted avg	0.77	0.77	0.77	201772

Random Forest Classifier Undersampled

	precision	recall	f1-score	support
0.0	0.02	0.78	0.04	603
1.0	0.06	0.55	0.11	5173
2.0	0.57	0.34	0.42	103354
3.0	0.55	0.41	0.47	92642
accuracy			0.38	201772
macro avg	0.30	0.52	0.26	201772
weighted avg	0.55	0.38	0.44	201772

Random Forest Classifier Oversampled

	precision	recall	f1-score	support
0.0	0.84	0.57	0.68	603
1.0	0.75	0.66	0.70	5173
2.0	0.81	0.82	0.82	103354
3.0	0.80	0.80	0.80	92642
accuracy			0.81	201772
macro avg	0.80	0.71	0.75	201772
weighted avg	0.80	0.81	0.80	201772

Random Forest Classifier Oversampled (100 trees)

	precision	recall	f1-score	support
0.0	0.84	0.57	0.68	603
1.0	0.74	0.65	0.70	5173
2.0	0.81	0.82	0.82	103354
3.0	0.80	0.79	0.80	92642
accuracy			0.80	201772
macro avg	0.80	0.71	0.75	201772
weighted avg	0.80	0.80	0.80	201772

Random Forest Classifier Oversampled (300 trees)

	precision	recall	f1-score	support
0.0	0.83	0.58	0.68	603
1.0	0.75	0.66	0.70	5173
2.0	0.82	0.82	0.82	103354
3.0	0.80	0.80	0.80	92642
accuracy			0.81	201772
macro avg	0.80	0.71	0.75	201772
weighted avg	0.81	0.81	0.81	201772

Random Forest Entropy 1000 trees (grid search best one)

	precision	recall	f1-score	support
0.0	0.84	0.59	0.69	603
1.0	0.76	0.67	0.71	5173
2.0	0.82	0.83	0.82	103354
3.0	0.80	0.80	0.80	92642
accuracy			0.81	201772
macro avg	0.80	0.72	0.76	201772
weighted avg	0.81	0.81	0.81	201772

With features:
Latitude, longitude, zipcode, species, diameter, sidewalk cracks, wires

Random Forest Classifier Oversampled (100 trees)

With new data

	precision	recall	f1-score	support
0.0	0.85	0.47	0.61	603
1.0	0.73	0.56	0.63	5173
2.0	0.80	0.81	0.80	103354
3.0	0.78	0.78	0.78	92642
accuracy			0.79	201772
macro avg	0.79	0.66	0.71	201772
weighted avg	0.79	0.79	0.79	201772

Random Forest Classifier Oversampled (500 trees) With new data

	precision	recall	f1-score	support
0.0	0.85	0.46	0.60	603
1.0	0.73	0.55	0.63	5173
2.0	0.80	0.80	0.80	103354
3.0	0.78	0.79	0.78	92642
accuracy			0.79	201772
macro avg	0.79	0.65	0.70	201772
weighted avg	0.79	0.79	0.79	201772

Random Forest Oversampled, 1000 trees
Non-pluto data, but only 3092 samples

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	1
1.0	0.00	0.00	0.00	17
2.0	0.58	0.61	0.60	313
3.0	0.56	0.56	0.56	288
accuracy			0.57	619
macro avg	0.29	0.29	0.29	619
weighted avg	0.56	0.57	0.56	619

With features:
Latitude, longitude, zipcode, species, diameter, sidewalk cracks, wires
+ PLUTO (numfloors, landuse)

Num samples now much lower: 3092

Random Forest Classifier Oversampled 500 trees

With PLUTO data, not sidewalk or wires

	precision	recall	f1-score	support
1.0	0.84	0.80	0.82	20
2.0	0.92	0.95	0.94	363
3.0	0.92	0.89	0.91	236
accuracy			0.92	619
macro avg	0.90	0.88	0.89	619
weighted avg	0.92	0.92	0.92	619

Random Forest Classifier Oversampled 500 trees

With PLUTO, wires and sidewalk data

	precision	recall	f1-score	support
1.0	0.94	0.85	0.89	20
2.0	0.93	0.95	0.94	363
3.0	0.93	0.90	0.92	236
accuracy			0.93	619
macro avg	0.93	0.90	0.92	619
weighted avg	0.93	0.93	0.93	619

Random Forest Classifier Oversampled 1000 trees

With PLUTO, wires and sidewalk data (best so far)

	precision	recall	f1-score	support
1.0	0.94	0.85	0.89	20
2.0	0.94	0.96	0.95	363
3.0	0.94	0.92	0.93	236
accuracy			0.94	619
macro avg	0.94	0.91	0.92	619
weighted avg	0.94	0.94	0.94	619

Just to add here, this is with a different random state split, which includes one dead tree in the test data which the model does not predict:

	precision	recall	f1-score	support

Random Forest Classifier Oversampled 500 trees

With PLUTO and sidewalk, wires; without lat and lon bin

	precision	recall	f1-score	support
1.0	0.94	0.75	0.83	20
2.0	0.90	0.93	0.92	363
3.0	0.90	0.87	0.88	236
accuracy			0.90	619
macro avg	0.91	0.85	0.88	619
weighted avg	0.90	0.90	0.90	619

Random Forest Classifier Oversampled 500 trees

With PLUTO and sidewalk, wires; without lat and lon bin or zipcode

	precision	recall	f1-score	support
1.0	0.94	0.75	0.83	20
2.0	0.90	0.90	0.90	363
3.0	0.84	0.86	0.85	236
accuracy			0.88	619
macro avg	0.89	0.83	0.86	619
weighted avg	0.88	0.88	0.88	619

Random Forest Classifier Oversampled 500 trees

With sidewalk, wires; without lat and lon bin or zipcode

	precision	recall	f1-score	support
0.0	0.00	0.33	0.01	603
1.0	0.04	0.38	0.06	5173
2.0	0.59	0.22	0.32	103354
3.0	0.58	0.40	0.47	92642
accuracy			0.30	201772
macro avg	0.30	0.33	0.22	201772
weighted avg	0.57	0.30	0.38	201772

Gradient Booster, Oversampled

```
GradientBoostingClassifier( n_estimators=500, learning_rate=1.0, max_depth=8, random_state=0)
```

All data

	precision	recall	f1-score	support
1.0	0.94	0.85	0.89	20
2.0	0.94	0.96	0.95	363
3.0	0.94	0.91	0.92	236
accuracy			0.94	619
macro avg	0.94	0.91	0.92	619
weighted avg	0.94	0.94	0.94	619

SVM Oversampled All data

```
svm = SVC(gamma='auto', C=3, random_state=0)
```

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	0
1.0	0.88	0.75	0.81	20
2.0	0.88	0.92	0.90	363
3.0	0.88	0.82	0.85	236
accuracy			0.88	619
macro avg	0.66	0.62	0.64	619
weighted avg	0.88	0.88	0.88	619

Final Results, All Features, Oversampled

Random Forest 0

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	1
1.0	0.82	0.92	0.87	25
2.0	0.92	0.90	0.91	350
3.0	0.87	0.89	0.88	243
accuracy			0.90	619
macro avg	0.65	0.68	0.66	619
weighted avg	0.90	0.90	0.90	619

Gradient Boosting 0

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	1
1.0	0.82	0.92	0.87	25
2.0	0.92	0.90	0.91	350
3.0	0.87	0.89	0.88	243
accuracy			0.90	619
macro avg	0.65	0.68	0.67	619

Random Forest 50 trees with wires and sidewalk data

	precision	recall	f1-score	support
0.0	0.84	0.47	0.60	603
1.0	0.72	0.57	0.64	5173
2.0	0.80	0.81	0.80	103354
3.0	0.78	0.78	0.78	92642
accuracy			0.79	201772
macro avg	0.78	0.66	0.71	201772
weighted avg	0.79	0.79	0.79	201772

```
'mean_fit_time': array([79.52576151, 86.13717837]),
 'std_fit_time': array([1.09218088, 0.73392889]),
 'mean_score_time': array([4.85432987, 4.86475325]),
 'std_score_time': array([0.07725953, 0.07512442]),
 'param_criterion': masked_array(data=['gini', 'entropy'],
              mask=[False, False],
        fill_value='?',
             dtype=object),
 'param_n_estimators': masked_array(data=[50, 50],
              mask=[False, False],
        fill_value='?',
             dtype=object),
 'params': [{'criterion': 'gini', 'n_estimators': 50},
  {'criterion': 'entropy', 'n_estimators': 50}],
 'split0_test_score': array([0.90293727, 0.90295842]),
```