# **Random Forest Classifier**

## Experiment 1

RandomForestClassifier(random\_state=0, criterion='gini', n\_estimators=390)

The accuracy for Random Forest: 0.8094944512946979

F-Measure for Random Forest: 0.8094944512946979

precision recall f1-score support

AF Other 0.61 0.45 0.52 91

Animal 0.92 0.77 0.84 75

Conductor 0.48 0.33 0.39 39

Connection 0.70 0.89 0.79 227

Crossarm 0.80 0.96 0.87 116

Dug up 0.90 0.90 0.90 99

Fuse 0.85 0.87 0.86 105

Installation 0.00 0.00 0.00 7

Lightning 0.96 0.82 0.89 33

OH Cable 0.81 0.37 0.51 59

Other 0.83 0.88 0.85 315

Pole 0.81 0.51 0.63 84

Trees 0.87 0.92 0.89 138

UG Cable 0.00 0.00 0.00 6

Vehicle 0.86 0.93 0.89 228

accuracy 0.81 1622

macro avg 0.69 0.64 0.66 1622

weighted avg 0.80 0.81 0.80 1622

Chart, scatter chart

Description automatically generated

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

A picture containing line chart

Description automatically generated

Chart, line chart

Description automatically generated

## Experiment 2

RandomForestClassifier(random\_state=0, criterion='gini', n\_estimators=1000, max\_depth=8, min\_samples\_leaf=5)

The accuracy for Random Forest: 0.7213316892725031

F-Measure for Random Forest: 0.7213316892725032

precision recall f1-score support

AF Other 1.00 0.04 0.08 91

Animal 0.96 0.57 0.72 75

Conductor 0.00 0.00 0.00 39

Connection 0.56 0.93 0.70 227

Crossarm 0.67 0.97 0.79 116

Dug up 0.93 0.76 0.83 99

Fuse 0.86 0.85 0.85 105

Installation 0.00 0.00 0.00 7

Lightning 1.00 0.42 0.60 33

OH Cable 1.00 0.02 0.03 59

Other 0.65 0.91 0.76 315

Pole 0.86 0.21 0.34 84

Trees 0.85 0.88 0.87 138

UG Cable 0.00 0.00 0.00 6

Vehicle 0.88 0.85 0.87 228

accuracy 0.72 1622

macro avg 0.68 0.49 0.50 1622

weighted avg 0.76 0.72 0.67 1622

Chart, scatter chart

Description automatically generated

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Experiement 3:

RandomForestClassifier(random\_state=0, criterion='gini', n\_estimators=1000, max\_depth=20, min\_samples\_leaf=2)

The accuracy for Random Forest: 0.8002466091245376

F-Measure for Random Forest: 0.8002466091245376

precision recall f1-score support

AF Other 0.62 0.40 0.48 91

Animal 0.95 0.76 0.84 75

Conductor 0.67 0.26 0.37 39

Connection 0.67 0.90 0.77 227

Crossarm 0.79 0.94 0.86 116

Dug up 0.93 0.90 0.91 99

Fuse 0.85 0.87 0.86 105

Installation 0.00 0.00 0.00 7

Lightning 0.90 0.85 0.88 33

OH Cable 0.80 0.34 0.48 59

Other 0.81 0.87 0.84 315

Pole 0.79 0.54 0.64 84

Trees 0.85 0.92 0.89 138

UG Cable 0.00 0.00 0.00 6

Vehicle 0.85 0.91 0.88 228

accuracy 0.80 1622

macro avg 0.70 0.63 0.65 1622

weighted avg 0.79 0.80 0.79 1622

A picture containing chart

Description automatically generated

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

Chart, line chart

Description automatically generated

## Using SMOTE

The accuracy for Random Forest: 0.7324290998766955

F-Measure for Random Forest: 0.7324290998766955

precision recall f1-score support

AF Other 0.49 0.40 0.44 91

Animal 0.87 0.77 0.82 75

Conductor 0.34 0.56 0.42 39

Connection 0.83 0.55 0.66 227

Crossarm 0.79 0.89 0.84 116

Dug up 0.72 0.96 0.82 99

Fuse 0.80 0.86 0.83 105

Installation 0.06 0.43 0.11 7

Lightning 0.72 0.94 0.82 33

OH Cable 0.46 0.71 0.56 59

Other 0.92 0.69 0.79 315

Pole 0.55 0.56 0.55 84

Trees 0.89 0.92 0.90 138

UG Cable 0.28 0.83 0.42 6

Vehicle 0.82 0.82 0.82 228

accuracy 0.73 1622

macro avg 0.64 0.73 0.65 1622

weighted avg 0.78 0.73 0.74 1622

A picture containing background pattern

Description automatically generated

Chart

Description automatically generated with medium confidence

Chart, bar chart

Description automatically generated

Chart

Description automatically generated with low confidence

Chart, line chart

Description automatically generated

To avoid overfitting:

I would agree with @Falcon w.r.t. the dataset size. It's likely that the main problem is the small size of the dataset. If possible, the best thing you can do is get more data, the more data (generally) the less likely it is to overfit, as random patterns that appear predictive start to get drowned out as the dataset size increases.

That said, I would look at the following params:

1. n\_estimators: @Falcon is wrong, in general the more trees the **less likely** the algorithm is to overfit. So try **increasing this**. The lower this number, the closer the model is to a decision tree, with a restricted feature set.
2. max\_features: try reducing this number (try 30-50% of the number of features). This determines how many features each tree is randomly assigned. The smaller, the less likely to overfit, but too small will start to introduce under fitting.
3. max\_depth: Experiment with this. This will reduce the complexity of the learned models, lowering over fitting risk. Try starting small, say 5-10, and increasing you get the best result.
4. min\_samples\_leaf: Try setting this to values greater than one. This has a similar effect to the max\_depth parameter, it means the branch will stop splitting once the leaves have that number of samples each.

**Final Experiment:**

The accuracy for Random Forest: 0.7540073982737361

F-Measure for Random Forest: 0.754007398273736

Recall for Random Forest: 0.754007398273736

A screenshot of a computer

Description automatically generated with low confidence

Chart, scatter chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, bar chart

Description automatically generated

A picture containing line chart

Description automatically generated

Chart, line chart

Description automatically generated

**Comment**: The accuracy, f-measure and recall of 75% which is converged. It has higher variance and lower bias. It has performed good in training data; however, it doesn’t generalize well on test data.