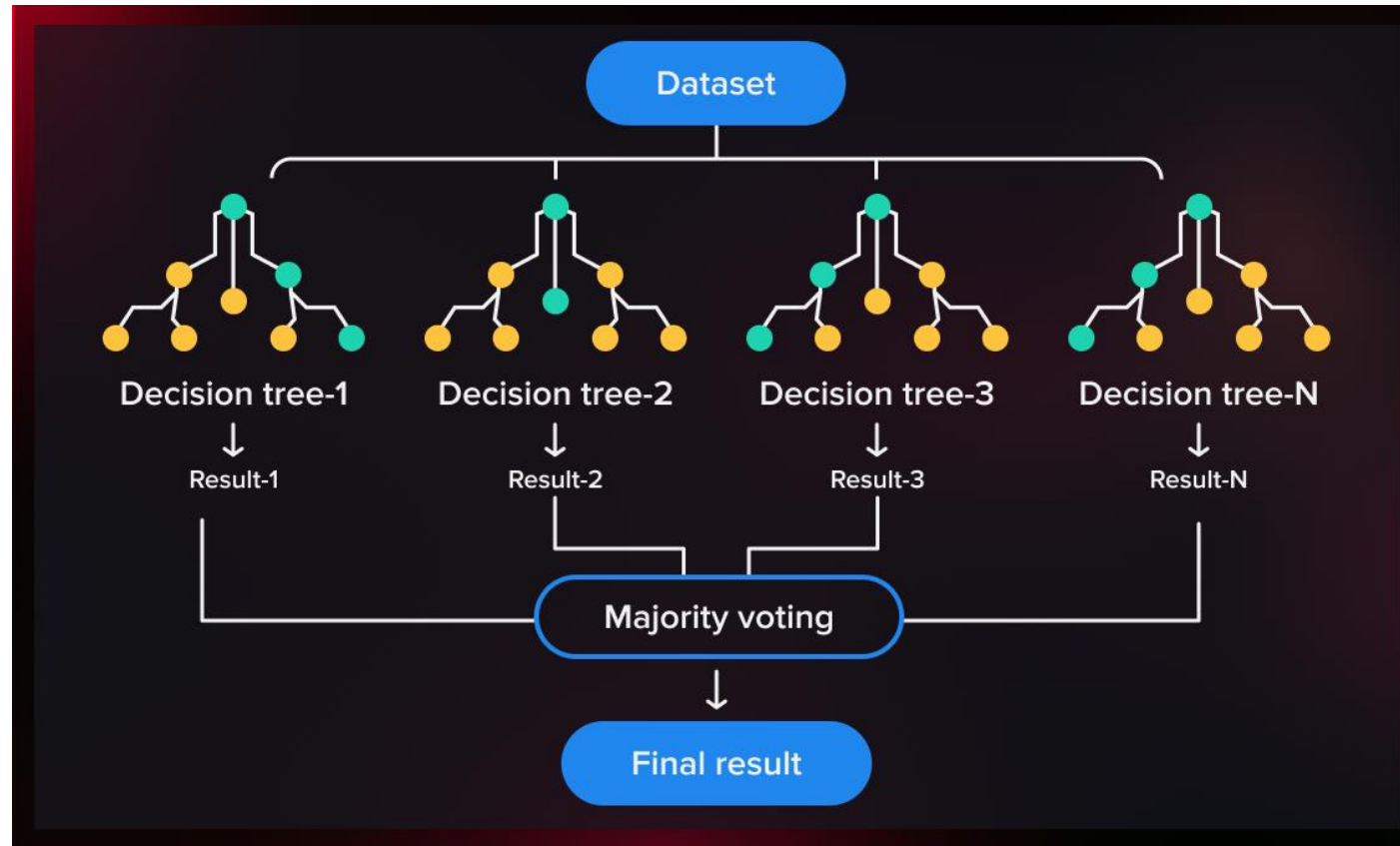# Ensemble Random Forest

UTKARSH GAIKWAD

STARTING IN FEW MINUTES

Ensemble Techniques

➢ Bagging

➢ Boosting

Random Forest (Bagging Ensemble)

Diabetic -1
Non Diabetic - 3

Decision Tree vs Random Forest
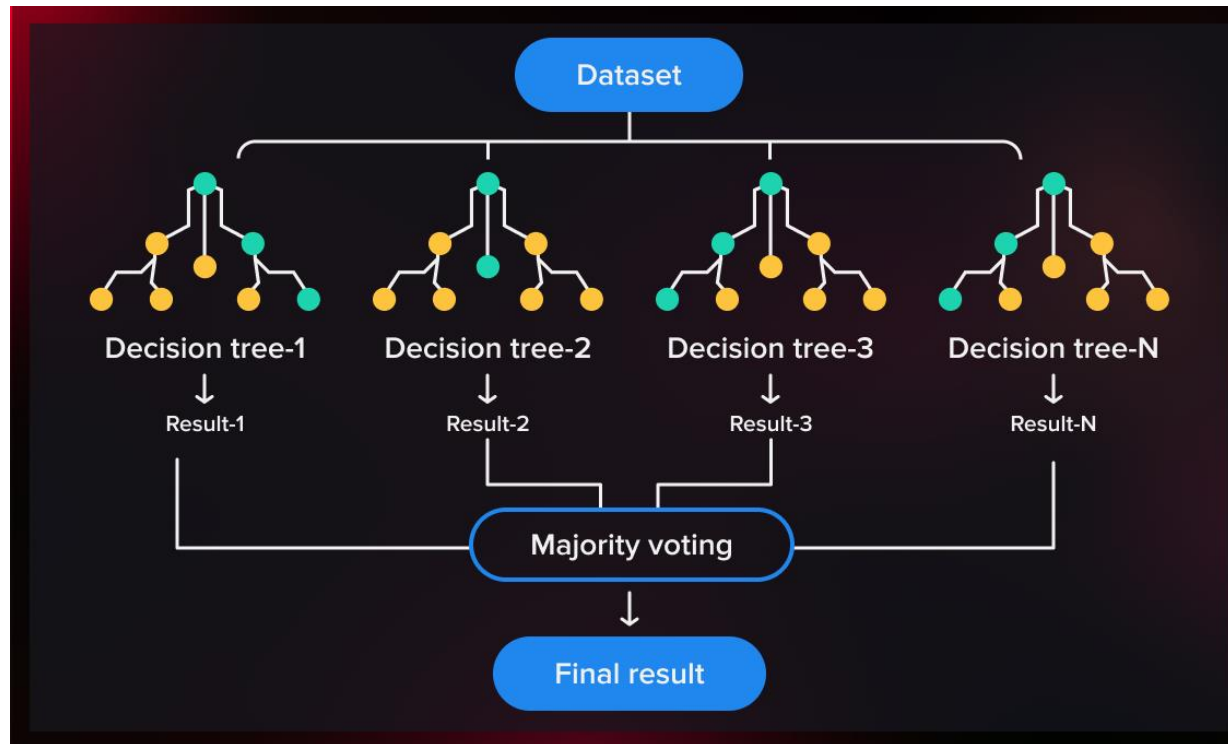
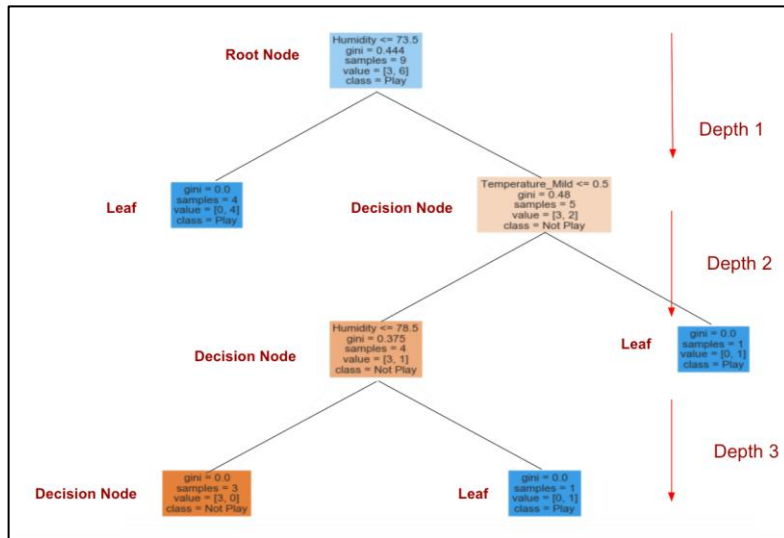|	Decision Tree	Random Forest
Interpretability	Easy to interpret	Hard to interpret
Accuracy	Accuracy can vary	Highly accurate
Overfitting	Likely to overfit data	Unlikely to overfit data
Outliers	Can be highly affected by outliers	Robust against outliers
Computation	Quick to build	Slow to build (computationally intensive)

Overfitting
Training – 0.98
Testing – 0.6

N-estimators

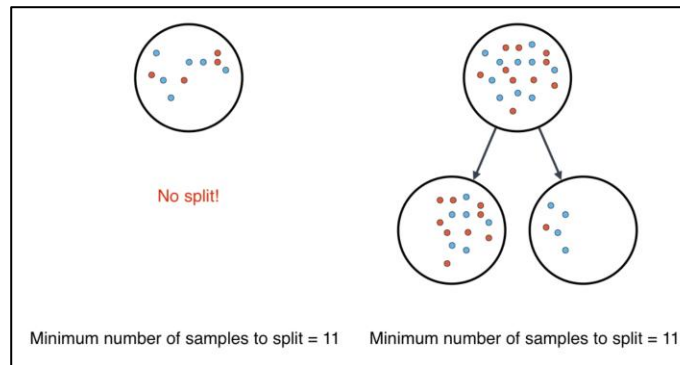How many decision trees to take ? In Below case it is 4 estimators
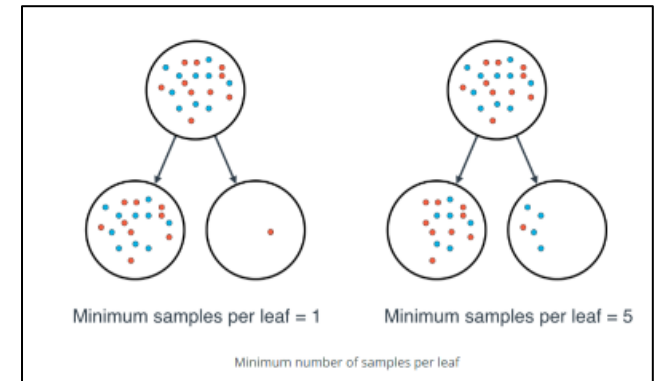
Parameters for Random Forest

Max Depth

Min Samples Split

Min Samples Leaf

Regression

Classification

Criterion : squared_error, absolute_error

Criterion : gini, entropy

Some more classification metrics

Scoring Metrics

https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter

```
In [57]:    1  import sklearn
            2  sklearn.metrics.get_scorer_names()

Out[57]: ['accuracy',
          'adjusted_mutual_info_score',
          'adjusted_rand_score',
          'average_precision',
          'balanced_accuracy',
          'completeness_score',
          'explained_variance',
          'f1',
          'f1_macro',
          'f1_micro',
```

Thank you

PING ME ON SKYPE FOR ANY QUERIES