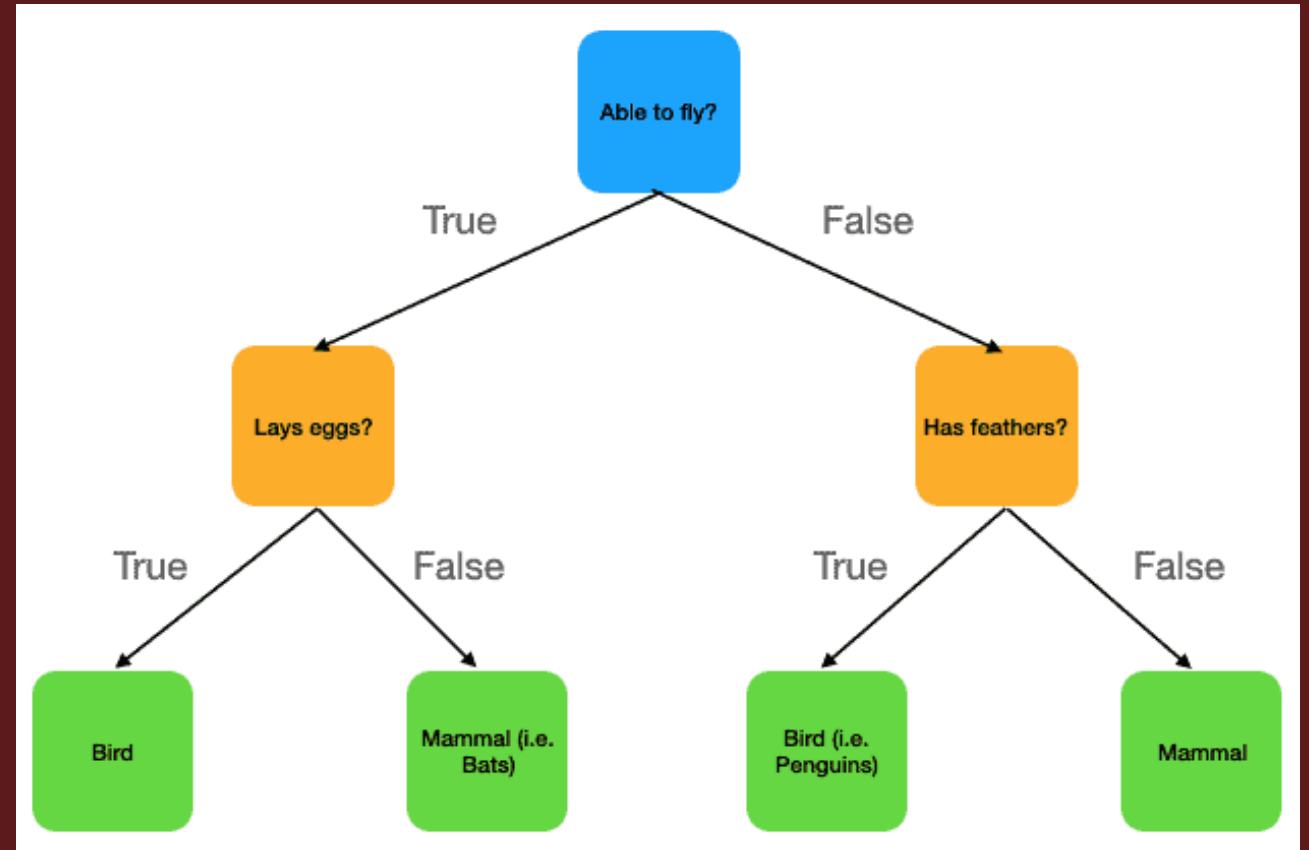# THE RANDOM FOREST ALGORITHM

DECISION TREES

A decision tree splits the data using decision nodes until we are left with "leaf nodes", or target value.

RANDOM FOREST

____

"Random forest is a commonly-used machine learning algorithm, trademarked by Leo Breiman and Adele Cutler, that combines the output of multiple decision trees to reach a single result." - What Is Random Forest? | IBM

id	x_0	x_1	x_2	x_3	x_4	y
0	4.3	4.9	4.1	4.7	5.5	0
1	3.9	6.1	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1

id
2
0
2
4
5
5

id
2
1
3
1
4
4

id
4
1
3
0
0
2

A FOREST OF DECISION TREES

x_0, x_1

x_2, x_3

x_2, x_4

$x_1 \le 4.9$

$x_0 \le 4.3$ 1

0   1

$x_3 \le 4.6$

1   0

$x_2 \le 4.1$

0   $x_2 \le 4.7$

1   0

Number of trees

N=1

N=10

N=100
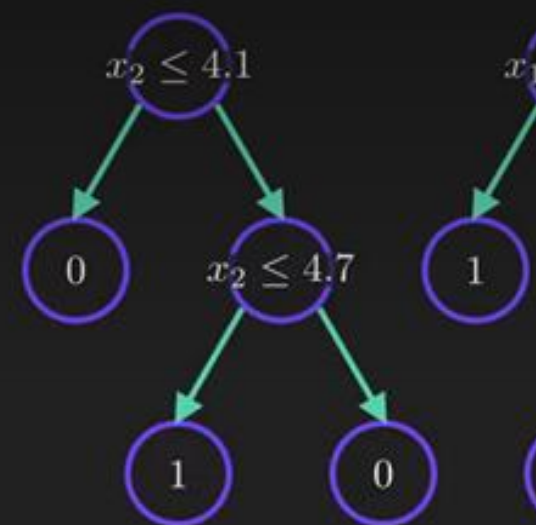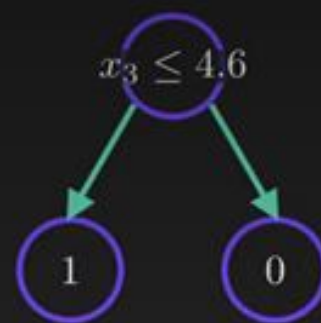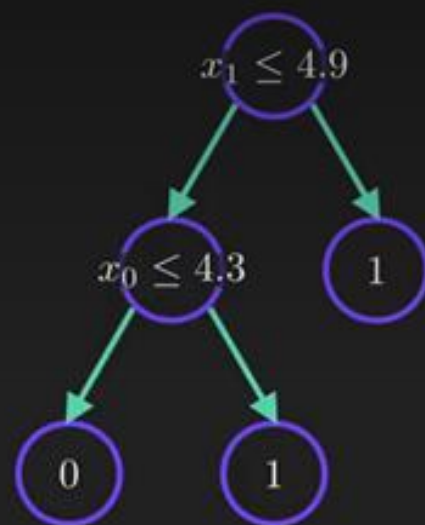
```
Actual vs. Predicted:
     Actual  Predicted
319       0          0
552       1          1
508       1          1
492       0          0
999       1          1
..      ...        ...
399       1          1
823       1          1
437       1          1
830       1          1
562       0          0

[200 rows x 2 columns]
Model Accuracy: 98.50%
```

```
     Actual  Predicted
931       1          1
748       0          0
692       0          0
506       1          1
117       0          0
..      ...        ...
546       0          0
81        1          1
123       1          1
460       0          0
924       1          1

[200 rows x 2 columns]
Model Accuracy: 99.50%
Confusion Matrix:
[[ 48   0]
 [  1 151]]
```

```
     Actual  Predicted
868       1          1
467       0          0
86        1          1
845       1          1
237       1          1
..      ...        ...
448       1          1
762       1          1
471       1          1
226       1          1
937       1          1

[200 rows x 2 columns]
Model Accuracy: 100.00%
Confusion Matrix:
[[ 49   0]
 [  0 151]]
```

```
score:0.975, n_estimator:1
score:0.99, n_estimator:2
score:1.0, n_estimator:3
score:1.0, n_estimator:4
score:1.0, n_estimator:5
score:0.995, n_estimator:6
score:1.0, n_estimator:7
score:1.0, n_estimator:8
score:1.0, n_estimator:9
score:1.0, n_estimator:10
score:1.0, n_estimator:11
score:1.0, n_estimator:12
score:1.0, n_estimator:13
score:1.0, n_estimator:14
score:1.0, n_estimator:15
score:1.0, n_estimator:16
score:1.0, n_estimator:17
score:1.0, n_estimator:18
score:1.0, n_estimator:19
score:1.0, n_estimator:20
score:1.0, n_estimator:21
score:1.0, n_estimator:22
score:1.0, n_estimator:23
score:1.0, n_estimator:24
score:1.0, n_estimator:25
```

OTHER PARAMETERS

$$Gini\ Index = 1 - \sum_{i=1}^{n}(P_i)^2$$

- *Split quality criterion: Gini*

  While designing the decision tree, the features possessing the least value of the Gini Index would get preferred.

- Max depth=none

- Min samples leaf=1

- Max features=sqrt

- Bootstrap=true

INITIAL DATASET

	A	B	C	D
1	Cell_Load	RSRP	RSRQ	Handover_Decision
2	0.25	81.78	12.92	0
3	0.64	87.11	16.27	1
4	0.79	80.32	13.04	1
5	0.41	65.97	11.06	1
6	0.11	79.52	11.92	1
7	0.22	69.9	11.24	1
8	0.78	99.33	19.37	1
9	0.35	81.5	12.24	1
10	0.89	60.84	9.47	1
11	0.34	63.57	7.06	1
12	0.75	65.75	11.42	1
13	0.75	64.2	7.81	1
14	0.32	81.4	13.74	0
15	0.58	51.88	9.58	1
16	0.85	61.07	7.58	1
17	0.9	68.57	14.53	1
18	0.46	98.88	16.02	0
19	0.02	64.38	7.65	1
20	0.73	74.05	13.65	1
21	0.73	68.96	13.16	1
22	0.7	63.69	6.61	1
23	0.65	81.25	14.9	1
24	0.16	53.81	6.35	1
25	0.2	91.62	19.66	0

...1000 samples

Train model using
random forest algorithm
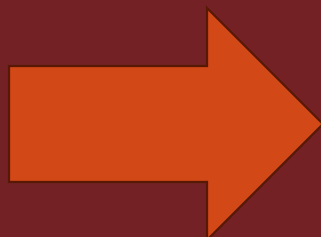
Trained_model.pkl

```
[1 1 1 1 0 1 1 0 1]
Predicted first handover decision : 1_
```

	A	B	C	D
1	Cell_Load	RSRP	RSRQ	Handover_Decision
2	0.64	73.33	13.93	
3	0.75	95.99	19.09	
4	0.91	76.52	14.97	
5	0.95	93.7	19.88	
6	0.13	98.28	16	
7	0.25	80.27	11.01	
8	0.46	56.63	9.29	
9	0.22	95.88	16.86	
10	0.58	50.64	5.87	

10 samples