



Accuracy vs depth graph

At maxdepth = 15, accuracy is maximum

Train error tends to 100, but test error slightly decreases or becomes constant after maxdepth = 20

1) Decision Tree:

```
writes = 0
|god = 0
|that = 0
|bible = 0
|bill = 0
|if = 0
|book = 0
|cheers = 0
|snm = 0
|sole = 0 : 2
|sole = 1 : 1
```

|snm = 1 : 1
|cheers = 1 : 1
|book = 1
|to = 0 : 1
|to = 1 : 2
|if = 1
|thanks = 0
|you = 0
|tao = 0
|addition = 0 : 2
|addition = 1 : 1
|tao = 1 : 1
|you = 1
|there = 0
|time = 0
|your = 0
|for = 0
|circle = 0
|mp = 0 : 1
|mp = 1 : 2
|circle = 1 : 2
|for = 1 : 2
|your = 1 : 1
|time = 1 : 1
|there = 1 : 2
|thanks = 1 : 2
|bill = 1 : 1
|bible = 1 : 1
|that = 1
|wrote = 0
|people = 0
|religious = 0
|an = 0
|he = 0
|face = 0
|tammy = 0
|claiming = 0
|doubt = 0 : 2
|doubt = 1 : 1
|claiming = 1 : 1

|tammy = 1 : 1
|face = 1 : 1
|he = 1
|graphic = 0 : 1
|graphic = 1 : 2
|an = 1
|can = 0
|how = 0
|know = 0
|will = 0
|port = 0 : 1
|port = 1 : 2
|will = 1 : 2
|know = 1 : 2
|how = 1 : 1
|can = 1
|dwyer = 0
|coming = 0 : 2
|coming = 1 : 1
|dwyer = 1 : 1
|religious = 1 : 1
|people = 1
|windows = 0 : 1
|windows = 1 : 2
|wrote = 1
|came = 0
|ve = 0
|online = 0 : 1
|online = 1 : 2
|ve = 1 : 2
|came = 1 : 2
|god = 1
|use = 0 : 1
|use = 1
|will = 0 : 2
|will = 1 : 1
writes = 1
|graphics = 0
|image = 0
|that = 0

```
|keith = 0
|god = 0
|who = 0
|have = 0
|am = 0
|time = 0
|by = 0
|with = 0
|you = 0
|july = 0
|mac = 0
|au = 0
|get = 0
|password = 0 : 1
|password = 1 : 2
|get = 1 : 2
|au = 1 : 2
|mac = 1 : 2
|july = 1 : 2
|you = 1 : 1
|with = 1 : 2
|by = 1 : 1
|time = 1 : 2
|am = 1 : 2
|have = 1
|rosenau = 0
|your = 0 : 2
|your = 1
|or = 0 : 2
|or = 1 : 1
|rosenau = 1 : 1
|who = 1
|disk = 0 : 1
|disk = 1 : 2
|god = 1 : 1
|keith = 1 : 1
|that = 1
|program = 0
|comp = 0
|csd = 0
```

```

|slow = 0
|uchicago = 0
|acm = 0
|tiff = 0
|algorithm = 0
|format = 0
|in = 0
|just = 0
|tracing = 0 : 1
|tracing = 1 : 2
|just = 1
|says = 0 : 2
|says = 1 : 1
|in = 1
|don = 0 : 1
|don = 1
|with = 0
|what = 0
|either = 0
|rahul = 0
|here = 0 : 1
|here = 1 : 2
|rahul = 1 : 2
|either = 1 : 2
|what = 1 : 1
|with = 1 : 1
|format = 1 : 2
|algorithm = 1 : 2
|tiff = 1 : 2
|acm = 1 : 2
|uchicago = 1 : 2
|slow = 1 : 2
|csd = 1 : 2
|comp = 1 : 2
|program = 1 : 2
|image = 1 : 2
|graphics = 1 : 2

```

2) yes overfitting occurs, after around maxdepth = 20

3) Accuracy:

For maxdepth = 15, maximum accuracy

Decision tree implementation-

Using info gain training set accuracy : 99.8114985862394

Using info gain test set accuracy : 82.74398868458275

scikit_learn implementation

using info_gain accuracy : 83.30975954738331