AIM: Implementation of ID3

THEORY:

ID3 stands for Iterative Dichotomiser 3 and is named such because the algorithm iteratively (repeatedly) dichotomizes(divides) features into two or more groups at each step.
Invented by Ross Quinlan, ID3 uses a top-down greedy approach to build a decision tree. In simple words, the top-down approach means that we start building the tree from the top and the greedy approach means that at each iteration we select the best feature at the present moment to create a node. Most generally ID3 is only used for classification problems with nominal features only.

1) IMPORTING LIBRARIES:

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
from sklearn.preprocessing import LabelEncoder #if data is string thn use it to preprocess the data
from sklearn.tree import DecisionTreeClassifier
```

2) READING DATASET:

```
df=pd.read_csv('glass.csv')
df
```

	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
0	1.52101	13.64	4.49	1.10	71.78	0.06	8.75	0.00	0.0	1
1	1.51761	13.89	3.60	1.36	72.73	0.48	7.83	0.00	0.0	1
2	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0.00	0.0	1
3	1.51766	13.21	3.69	1.29	72.61	0.57	8.22	0.00	0.0	1
4	1.51742	13.27	3.62	1.24	73.08	0.55	8.07	0.00	0.0	1
...	...	...	...	...	...	...	...	...	...	...
209	1.51623	14.14	0.00	2.88	72.61	0.08	9.18	1.06	0.0	7
210	1.51685	14.92	0.00	1.99	73.06	0.00	8.40	1.59	0.0	7
211	1.52065	14.36	0.00	2.02	73.42	0.00	8.44	1.64	0.0	7
212	1.51651	14.38	0.00	1.94	73.61	0.00	8.48	1.57	0.0	7
213	1.51711	14.23	0.00	2.08	73.36	0.00	8.62	1.67	0.0	7

214 rows × 10 columns

3) EXTRACTING FEATURES:

```
x=df.iloc[:,:-1]
```

```
y=df.iloc[:,9]
```

4) **MAKING MODEL:**

```
dt=DecisionTreeClassifier(criterion="entropy")
dt
```

5) **SPLITTING DATASET INTO TRAINING , TESTING & PREDICTING VALUES:**

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = .3, random_state=0)
dt.fit(X_train, y_train)
y_pred = dt.predict(X_test)
y_pred
```

```
array([7, 1, 2, 6, 5, 2, 1, 2, 2, 2, 1, 1, 2, 2, 2, 7, 3, 2, 3, 2, 5, 1,
       7, 7, 7, 1, 7, 1, 1, 2, 1, 3, 2, 7, 2, 1, 1, 3, 1, 7, 2, 6, 2, 1,
       2, 1, 1, 2, 1, 2, 1, 7, 7, 1, 2, 1, 1, 2, 1, 3, 1, 1, 1, 6, 1])
```

6) **CONFUSION MATRIX:**

```
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test, y_pred))
```

```
[[17  3  1  0  0  0]
 [ 6 15  1  0  1  3]
 [ 2  2  3  0  0  0]
 [ 0  0  0  2  0  0]
 [ 0  0  0  0  2  0]
 [ 0  0  0  0  0  7]]
```

CONCLUSION:

From this practical, I have learned about implementation of ID3 algorithm in python.