

MACHINE LEARNING LAB

August 2, 2020

Manish Singh CS3
Roll No. 17/466



Computer Science and Engineering Department
Rajasthan Technical University, Kota

Contents

1	Problem Statement	3
2	ID3 Algorithm	4
3	Training Set	5
4	Program	7
5	Output	9
6	Result	20

1 Problem Statement

Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
(*id3.csv*)(*id3_test1.csv*)(*pima - indians - diabetes.csv*)

2 ID3 Algorithm

ID3 stands for Iterative Dichotomiser 3

It is a classification algorithm that follows a greedy approach by selecting a best attribute that yields maximum Information Gain(IG) or minimum Entropy(H).

Algorithm:

1. Calculate entropy for dataset.
2. For each attribute/feature
 1. Calculate entropy for all its categorical values.
 2. Calculate information gain for the feature.
3. Find the feature with maximum information gain.
4. Repeat it until we get the desired tree.

3 Training Set

1. ID3 - Dataset

	Outlook	Temperature	Humidity	Wind	Answer
0	sunny	hot	high	weak	no
1	sunny	hot	high	strong	no
2	overcast	hot	high	weak	yes
3	rain	mild	high	weak	yes
4	rain	cool	normal	weak	yes
5	rain	cool	normal	strong	no
6	overcast	cool	normal	strong	yes
7	sunny	mild	high	weak	no
8	sunny	cool	normal	weak	yes
9	rain	mild	normal	weak	yes
10	sunny	mild	normal	strong	yes
11	overcast	mild	high	strong	yes
12	overcast	hot	normal	weak	yes
13	rain	mild	high	strong	no

2. ID3 Test- Dataset

	Outlook	Temperature	Humidity	Wind
0	rain	cool	normal	strong
1	sunny	mild	normal	strong

3. PIMA Indian Diabetes- Dataset

	pregnancies	Glucose	BloodPressure	SkinThickness	Insuline	BMI	DiabeticPed	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1
10	4	110	92	0	0	37.6	0.191	30	0
11	10	168	74	0	0	38.0	0.537	34	1
12	10	139	80	0	0	27.1	1.441	57	0
13	1	189	60	23	846	30.1	0.398	59	1
14	5	166	72	19	175	25.8	0.587	51	1
15	7	100	0	0	0	30.0	0.484	32	1
16	0	118	84	47	230	45.8	0.551	31	1
17	7	107	74	0	0	29.6	0.254	31	1
18	1	103	30	38	83	43.3	0.183	33	0
19	1	115	70	30	96	34.6	0.529	32	1
20	3	126	88	41	235	39.3	0.704	27	0
21	8	99	84	0	0	35.4	0.388	50	0
22	7	196	90	0	0	39.8	0.451	41	1
23	9	119	80	35	0	29.0	0.263	29	1
24	11	143	94	33	146	36.6	0.254	51	1

4 Program

```
import numpy as np
eps = np.finfo(float).eps
from numpy import log2 as log
import pandas as pd
import pprint

df = pd.read_csv("id3.csv")
def find_entropy(df):
    Class = df.keys()[-1]
    entropy = 0
    values = df[Class].unique()
    for value in values:
        fraction = df[Class].value_counts()[value]/len(df[Class])
        entropy += -fraction*np.log2(fraction)
    return entropy

def find_entropy_attribute(df,attribute):
    Class = df.keys()[-1]
    target_variables = df[Class].unique()
    variables = df[attribute].unique()
    entropy2 = 0
    for variable in variables:
        entropy = 0
        for target_variable in target_variables:
            num = len(df[attribute][df[attribute]==variable][df[Class] ==target_variable])
            den = len(df[attribute][df[attribute]==variable])
            fraction = num/(den+eps)
            entropy += -fraction*log(fraction+eps)
        fraction2 = den/len(df)
        entropy2 += -fraction2*entropy
    return abs(entropy2)

def find_winner(df):
    Entropy_att = []
    IG = []
    for key in df.keys()[:-1]:
        Entropy_att.append(find_entropy_attribute(df,key))
        IG.append(find_entropy(df)-find_entropy_attribute(df,key))
    return df.keys()[:-1][np.argmax(IG)]

def get_sub_table(df, node,value):
    return df[df[node] == value].reset_index(drop=True)

def buildTree(df,tree=None):
    Class = df.keys()[-1]
    node = find_winner(df)
    attValue = np.unique(df[node])
    if tree is None:
        tree={}
        tree[node] = {}
    for value in attValue:
        sub_table = get_sub_table(df,node,value)
        clValue,counts = np.unique(sub_table[Class],return_counts=True)
        if len(counts)==1:
```

```

        tree[node][value] = clValue[0]
    else:
        tree[node][value] = buildTree(sub_table)

    return tree

print("***20, "ID3 Algorithm to build Decision Tree","**20)
print("\n\nDecision Tree for the dataset ID3: \n")
t=buildTree(df)
pprint.pprint(t)
print("\n\n\n"+"**50)
print("\n\nDecision Tree for the dataset ID3 Test: \n")
df1 = pd.read_csv("id3_test_1.csv")
t_test = buildTree(df1)
pprint.pprint(t_test)
print("\n\n\n"+"**50)
print("\n\nDecision Tree for the dataset PIMA Indian Diabetes: \n")
df2 = pd.read_csv("pima_indian_diabetes.csv")
t_pima = buildTree(df2)
pprint.pprint(t_pima)

```


5 Output

```
***** ID3 Algorithm to build Decision Tree *****
```

```
Decision Tree for the dataset ID3:
```

```
{'Outlook': {'overcast': 'yes',  
             'rain': {'Wind': {'strong': 'no', 'weak': 'yes'}},  
             'sunny': {'Humidity': {'high': 'no', 'normal': 'yes'}}}}
```

```
*****
```

```
Decision Tree for the dataset ID3 Test:
```

```
{'Outlook': {'rain': 'strong', 'sunny': 'strong'}}
```

Decision Tree for the dataset PIMA Indian Diabetes:

```
{'DiabeticPed': {0.078: 0,
0.084: 0,
0.085: 0,
0.08800000000000001: {'pregnancies': {1: 1, 2: 0}},
0.08900000000000001: 0,
0.092: 0,
0.096: 0,
0.1: 0,
0.10099999999999999: 0,
0.102: 0,
0.107: 0,
0.10800000000000001: 0,
0.115: 0,
0.11800000000000001: 0,
0.121: {'pregnancies': {3: 0, 6: 1}},
0.122: 0,
0.12300000000000001: 0,
0.126: 0,
0.127: 1,
0.128: {'pregnancies': {2: 0, 7: 1}},
0.129: 1,
0.13: 0,
0.133: 0,
0.134: 0,
0.135: 1,
0.136: 0,
0.13699999999999998: {'pregnancies': {8: 1, 12: 0}},
0.138: 0,
0.14: 0,
0.141: {'pregnancies': {0: 1, 2: 0, 10: 1}},
0.142: 0,
0.14300000000000002: 0,
0.14400000000000002: 0,
0.145: 0,
0.147: 0,
0.14800000000000002: {'pregnancies': {2: 0, 4: 0, 9: 1}},
0.149: 0,
0.15: {'pregnancies': {3: 0, 4: 1}},
0.151: {'pregnancies': {3: 0, 5: 1, 6: 0}},
0.153: {'pregnancies': {5: 0, 15: 1}},
0.154: 0,
0.155: 0,
0.156: 0,
0.157: 0,
0.158: {'Glucose': {108: 0, 197: 1}},
0.159: 0,
0.16: 0,
0.161: {'pregnancies': {4: 1, 7: 0}},
0.162: 0,
0.163: 1,
0.16399999999999998: 0,
0.165: {'pregnancies': {1: 0, 7: 1, 8: 1}},
0.166: 0,
0.16699999999999998: 0,
0.17: 0,
0.171: 0,
0.17300000000000001: 0,
```

```

0.16699999999999998: 0,
0.17: 0,
0.171: 0,
0.17300000000000001: 0,
0.174: 0,
0.175: 0,
0.17600000000000002: 0,
0.177: 0,
0.17800000000000002: {'Glucose': {106: 0, 147: 1, 162: 1}},
0.179: 0,
0.18: {'pregnancies': {6: 1, 13: 0}},
0.18100000000000002: 0,
0.182: 0,
0.183: {'pregnancies': {1: 0, 8: 1}},
0.18600000000000003: 0,
0.187: 0,
0.188: 0,
0.18899999999999997: 0,
0.19: 0,
0.191: 0,
0.192: 0,
0.19399999999999998: 0,
0.196: 1,
0.19699999999999998: {'Glucose': {105: 0,
                                106: 0,
                                112: 1,
                                123: 0}},
0.198: 0,
0.19899999999999998: 1,
0.2: 0,
0.201: 0,
0.203: {'pregnancies': {0: 1, 5: 0}},
0.204: 0,
0.205: {'pregnancies': {0: 1, 1: 0, 10: 1}},
0.20600000000000002: 0,
0.207: 0,
0.209: {'pregnancies': {5: 1, 7: 0}},
0.21: 0,
0.212: 1,
0.215: 0,
0.217: 0,
0.218: 0,
0.21899999999999997: {'pregnancies': {1: 0, 3: 1}},
0.22: 1,
0.221: 0,
0.222: 1,
0.223: 0,
0.225: 0,
0.226: 1,
0.22699999999999998: 1,
0.22899999999999998: 0,
0.23: {'pregnancies': {4: 0, 9: 1}},
0.231: 0,
0.23199999999999998: 1,
0.233: {'pregnancies': {0: 0, 1: 1}},
0.23399999999999999: {'Glucose': {91: 0, 133: 1}},
0.235: {'pregnancies': {2: 0, 4: 1, 7: 0}},
0.23600000000000002: 0,
0.237: 0,
0.23800000000000002: {'Glucose': {86: 0,
                                96: 0,

```

```

0.23800000000000002: {'Glucose': {86: 0,
                                   96: 0,
                                   102: 0,
                                   134: 1,
                                   156: 1}},
0.239: 1,
0.24: 1,
0.24100000000000002: 1,
0.243: 0,
0.244: 0,
0.245: {'pregnancies': {5: 0, 6: 1, 10: 0, 13: 0}},
0.24600000000000002: 0,
0.247: {'pregnancies': {0: 1, 6: 0}},
0.248: {'pregnancies': {1: 0, 3: 1}},
0.249: 0,
0.251: 0,
0.252: 0,
0.253: 0,
0.254: {'Glucose': {107: 1,
                    122: 0,
                    124: 1,
                    143: 1,
                    161: 0,
                    187: 1}},
0.255: 0,
0.256: 0,
0.257: {'pregnancies': {2: 0, 7: 1, 13: 1}},
0.258: {'pregnancies': {0: 1, 1: 0, 3: 1, 5: 0, 7: 1, 10: 0}},
0.259: {'Glucose': {108: 1, 119: 0, 120: 1, 121: 0, 165: 0}},
0.26: {'pregnancies': {1: 0, 6: 0, 9: 1, 11: 1}},
0.261: {'Glucose': {81: 0, 112: 1, 115: 1, 118: 0, 121: 0}},
0.262: 0,
0.263: {'pregnancies': {0: 0, 5: 0, 9: 1, 10: 0}},
0.264: 1,
0.265: 0,
0.267: 0,
0.268: {'Glucose': {73: 0, 128: 0, 130: 0, 163: 1, 169: 1}},
0.26899999999999996: 0,
0.27: {'Glucose': {78: 0, 131: 1, 133: 1, 152: 0}},
0.271: 0,
0.272: 1,
0.27699999999999997: 1,
0.278: 1,
0.27899999999999997: 0,
0.28: 0,
0.282: {'pregnancies': {1: 1, 9: 0}},
0.28300000000000003: 0,
0.284: 0,
0.285: 0,
0.28600000000000003: {'Glucose': {106: 0, 136: 1}},
0.287: 0,
0.289: 0,
0.29: 0,
0.292: 0,
0.293: {'pregnancies': {1: 1, 3: 0}},
0.294: 0,
0.295: 0,
0.29600000000000004: 1,
0.297: 1,
0.299: 0,
0.3: 0

```

```

0.299: 0,
0.3: 0,
0.302: 1,
0.303: 0,
0.304: 0,
0.305: 0,
0.306: 0,
0.307: 0,
0.313: 0,
0.314: {'pregnancies': {0: 1, 3: 0}},
0.315: 0,
0.317: 0,
0.318: 0,
0.319: 1,
0.32299999999999995: {'pregnancies': {1: 0, 3: 1}},
0.324: 0,
0.325: 1,
0.326: {'pregnancies': {2: 0, 10: 1}},
0.32799999999999996: 1,
0.32899999999999996: 0,
0.33: 0,
0.331: 1,
0.332: 0,
0.33399999999999996: {'pregnancies': {0: 1, 1: 0}},
0.335: 1,
0.336: 0,
0.337: {'pregnancies': {2: 1, 5: 0, 7: 1}},
0.33799999999999997: 0,
0.34: {'pregnancies': {2: 0, 4: 0, 5: 1}},
0.341: 0,
0.342: 0,
0.34299999999999997: {'pregnancies': {2: 0, 5: 1}},
0.344: 1,
0.345: 1,
0.34600000000000003: 1,
0.34700000000000003: 0,
0.349: {'BloodPressure': {60: 1, 68: 1, 88: 0}},
0.35100000000000003: 0,
0.35200000000000004: 0,
0.355: 1,
0.35600000000000004: {'pregnancies': {3: 1, 6: 0}},
0.358: 1,
0.361: 1,
0.36200000000000004: 0,
0.364: {'pregnancies': {0: 1, 5: 0}},
0.365: {'pregnancies': {0: 1, 1: 0}},
0.366: 0,
0.368: {'pregnancies': {2: 0, 6: 1}},
0.37: 0,
0.371: 1,
0.374: 0,
0.375: 0,
0.376: 1,
0.37799999999999995: {'pregnancies': {5: 1, 12: 0}},
0.38: {'pregnancies': {4: 1, 9: 0}},
0.381: 0,
0.382: 0,
0.38299999999999995: 1,
0.385: 0,
0.38799999999999996: 0,
0.388: 0

```

```

0.385: 0,
0.38799999999999996: 0,
0.389: 0,
0.391: 0,
0.39299999999999996: 0,
0.39399999999999996: 0,
0.395: 1,
0.396: 0,
0.39799999999999996: 1,
0.39899999999999997: 0,
0.4: 0,
0.401: 0,
0.402: 1,
0.40299999999999997: 1,
0.40399999999999997: 0,
0.40700000000000003: 0,
0.408: 1,
0.409: 0,
0.41100000000000003: 0,
0.41200000000000003: {'pregnancies': {1: 0, 14: 1}},
0.415: 0,
0.41600000000000004: 0,
0.41700000000000004: 0,
0.419: 0,
0.42: 0,
0.42100000000000004: 0,
0.42200000000000004: {'Glucose': {71: 0, 88: 0, 144: 1}},
0.423: 1,
0.426: 0,
0.42700000000000005: 0,
0.43: 0,
0.431: 1,
0.43200000000000005: 0,
0.433: {'pregnancies': {0: 0, 2: 1}},
0.434: 0,
0.435: 1,
0.439: {'pregnancies': {5: 0, 7: 1}},
0.441: 1,
0.44299999999999995: {'Glucose': {97: 0, 123: 0, 154: 1}},
0.444: 0,
0.446: 0,
0.447: 1,
0.451: 1,
0.452: {'pregnancies': {0: 0, 3: 0, 5: 1}},
0.45299999999999996: 0,
0.45399999999999996: 0,
0.455: {'pregnancies': {0: 1, 2: 0}},
0.457: 0,
0.46: 0,
0.46299999999999997: 0,
0.46399999999999997: 0,
0.465: 1,
0.466: 0,
0.467: 1,
0.47100000000000003: 0,
0.47200000000000003: 0,
0.479: 1,
0.48200000000000004: 0,
0.483: 0,
0.484: 1,
0.485: 0,
0.48799999999999996: 0,

```

```

0.467: 1,
0.47100000000000003: 0,
0.47200000000000003: 0,
0.479: 1,
0.48200000000000004: 0,
0.483: 0,
0.484: 1,
0.485: 0,
0.48700000000000004: 0,
0.488: 0,
0.491: 0,
0.493: 0,
0.495: 0,
0.496: {'Glucose': {88: 1, 102: 0, 116: 0}},
0.49700000000000005: 0,
0.498: 0,
0.499: 0,
0.501: 0,
0.502: 1,
0.503: 1,
0.507: 0,
0.509: 0,
0.51: 1,
0.512: 0,
0.514: {'pregnancies': {1: 0, 5: 1}},
0.515: 0,
0.516: 1,
0.52: {'pregnancies': {0: 0, 2: 0, 4: 1}},
0.525: 0,
0.526: 0,
0.527: 0,
0.528: {'pregnancies': {1: 0, 12: 1}},
0.529: 1,
0.532: 0,
0.534: 1,
0.536: {'pregnancies': {1: 0, 4: 1}},
0.537: 1,
0.539: 1,
0.542: 1,
0.5429999999999999: 1,
0.545: 0,
0.546: 0,
0.547: 0,
0.5489999999999999: 1,
0.551: {'pregnancies': {0: 1, 1: 0, 3: 0, 8: 0}},
0.5539999999999999: 1,
0.557: {'pregnancies': {3: 0, 11: 1}},
0.5589999999999999: 0,
0.56: 0,
0.561: 0,
0.564: 0,
0.565: 1,
0.569: 1,
0.57100000000000001: 0,
0.57200000000000001: 0,
0.575: 1,
0.578: 1,
0.58: 0,
0.58200000000000001: 0,
0.583: {'pregnancies': {1: 0, 5: 1, 13: 1}},
0.586: {'pregnancies': {2: 0, 7: 1}},

```



```

0.5870000000000001: {'Glucose': {44: 0, 91: 0, 166: 1}},
0.588: 1,
0.591: 0,
0.593: 1,
0.595: 0,
0.597: 0,
0.598: 0,
0.6: 0,
0.601: 0,
0.605: {'pregnancies': {0: 0, 8: 1}},
0.607: 0,
0.61: 0,
0.612: 0,
0.613: 1,
0.614: 0,
0.615: 1,
0.619: 0,
0.624: 0,
0.626: 0,
0.627: 1,
0.629: 0,
0.63: 1,
0.631: 0,
0.637: {'pregnancies': {2: 0, 9: 1}},
0.64: {'pregnancies': {5: 0, 8: 1}},
0.645: 1,
0.6459999999999999: 1,
0.647: 0,
0.649: 0,
0.652: 1,
0.654: 0,
0.655: 0,
0.6579999999999999: 0,
0.66: {'pregnancies': {0: 0, 5: 1}},
0.6609999999999999: 1,
0.665: 1,
0.6659999999999999: 0,
0.672: 1,
0.6729999999999999: 0,
0.674: {'pregnancies': {2: 1, 6: 0}},
0.677: 0,
0.6779999999999999: {'pregnancies': {1: 0, 3: 1}},
0.68: 0,
0.682: 1,
0.6859999999999999: {'Glucose': {107: 0, 179: 1}},
0.687: {'Glucose': {124: 1, 135: 0, 142: 0, 153: 0}},
0.6920000000000001: {'Glucose': {122: 1,
                                130: 0,
                                150: 1,
                                151: 0}},
0.693: 1,
0.695: 0,
0.696: 0,
0.698: 0,
0.6990000000000001: 0,
0.7020000000000001: 1,
0.703: 0,
0.7040000000000001: 0,
0.705: 0,
0.7090000000000001: 0,

```



```

0.703: 0,
0.7040000000000001: 0,
0.705: 0,
0.7090000000000001: 0,
0.711: 1,
0.7170000000000001: 0,
0.718: 0,
0.7190000000000001: 1,
0.721: 1,
0.722: 1,
0.725: 0,
0.727: {'Glucose': {0: 1, 107: 0}},
0.73: 0,
0.731: 1,
0.732: 1,
0.733: 0,
0.7340000000000001: 1,
0.735: 0,
0.738: 0,
0.741: 1,
0.742: 1,
0.743: 1,
0.7440000000000001: 0,
0.745: 1,
0.748: 0,
0.757: 1,
0.759: 1,
0.7609999999999999: {'pregnancies': {2: 0, 3: 1}},
0.7659999999999999: 0,
0.767: 0,
0.7709999999999999: 1,
0.773: 0,
0.785: 1,
0.787: {'pregnancies': {0: 0, 7: 1}},
0.8009999999999999: 0,
0.8029999999999999: 1,
0.804: 0,
0.805: 1,
0.8079999999999999: 1,
0.813: 0,
0.816: 0,
0.8170000000000001: 1,
0.821: 0,
0.825: 1,
0.826: 1,
0.828: 0,
0.831: 1,
0.8320000000000001: 0,
0.833: 0,
0.8390000000000001: {'pregnancies': {0: 1, 6: 0}},
0.84: 0,
0.845: 0,
0.851: 1,
0.855: 1,
0.856: 0,
0.867: 1,
0.871: 1,
0.8740000000000001: 0,
0.875: 1,
0.878: 0,
- - -

```

```

0.871: 1,
0.8740000000000001: 0,
0.875: 1,
0.878: 0,
0.88: 0,
0.8809999999999999: 0,
0.8859999999999999: 0,
0.892: 0,
0.893: 1,
0.904: 0,
0.905: 1,
0.917: 0,
0.925: 1,
0.9259999999999999: 1,
0.93: 0,
0.932: 0,
0.9329999999999999: 1,
0.9440000000000001: 0,
0.9470000000000001: 0,
0.9490000000000001: 0,
0.955: 1,
0.956: 1,
0.9620000000000001: {'pregnancies': {0: 0, 1: 1}},
0.966: 0,
0.968: {'pregnancies': {2: 0, 3: 1}},
0.97: 1,
0.997: 0,
1.001: 1,
1.021: 0,
1.022: 0,
1.034: 1,
1.057: 1,
1.072: 1,
1.0759999999999998: 0,
1.095: 0,
1.0959999999999999: 0,
1.101: 0,
1.114: 1,
1.127: 1,
1.136: 1,
1.138: 0,
1.1440000000000001: 1,
1.1540000000000001: 1,
1.159: 0,
1.162: 0,
1.1740000000000002: 0,
1.182: 1,
1.189: 1,
1.1909999999999998: 1,
1.213: 1,
1.222: 1,
1.224: 1,
1.251: 0,
1.258: 1,
1.268: 0,
1.2819999999999998: 1,
1.2919999999999998: 1,
1.318: 1,
1.321: 1,
1.3530000000000002: 1,
1.39: 1,

```

```

1.001: 1,
1.021: 0,
1.022: 0,
1.034: 1,
1.057: 1,
1.072: 1,
1.0759999999999998: 0,
1.095: 0,
1.0959999999999999: 0,
1.101: 0,
1.114: 1,
1.127: 1,
1.136: 1,
1.138: 0,
1.1440000000000001: 1,
1.1540000000000001: 1,
1.159: 0,
1.162: 0,
1.1740000000000002: 0,
1.182: 1,
1.189: 1,
1.1909999999999998: 1,
1.213: 1,
1.222: 1,
1.224: 1,
1.251: 0,
1.258: 1,
1.268: 0,
1.2819999999999998: 1,
1.2919999999999998: 1,
1.318: 1,
1.321: 1,
1.3530000000000002: 1,
1.39: 1,
1.391: 1,
1.3940000000000001: 1,
1.4: 0,
1.4409999999999998: 0,
1.4609999999999999: 0,
1.476: 0,
1.6: 0,
1.6980000000000002: 0,
1.699: 0,
1.7309999999999999: 0,
1.781: 0,
1.893: 1,
2.137: 1,
2.2880000000000003: 1,
2.329: 0,
2.42: 1}}

```

6 Result

We have successfully generated the Decision Tree using ID3 algorithm for the given Data set.