Student: Bradley Dodds

Class: CPSC 483

Due Date: 11/27/2019

Project 1

Project Description: This project will demonstrate the ID3 algorithm. It can be used for "most" or "different" data sets. The program will calculate the entropy and information gain at each step and stage.

The program will read different datasets by text file and implement the ID3 algorithm. This project will use the ID3 algorithm on two separate text files: *data.txt, data1.txt.*

Files used: main.cpp, header.h, data.txt, data1.txt

Programming Language: c++

Class created:

```
class ID3
```

Functions used:

```
//Read data file and recognize attributes and values
ID3.readFile(FILE);
ID3.recognizeFeatures();
ID3.output_recognizedFeatures();

//Convert original data set to binary for manipulation
ID3.convertBinaryVector();
ID3.output_initialDataVector();
ID3.output_binaryVector();

//Calculations-calculate entropy and information gain
ID3.entropy();
ID3.information_gain();
```

Output: Decision Trees Power Point Data Set

```
Humidity Wind Play
Total attributes: 3
Each attribute has 14 features

Humidity
Total different values: 2
Recognized values: high normal

Wind
Total different values: 2
Recognized values: weak strong

Play
Total different values: 2
Recognized values: no yes


INITIAL DATA VECTOR
Humidity Wind Play
high weak no
high strong no
high weak yes
high weak yes
normal weak yes
normal strong no
normal strong yes
high weak no
normal weak yes
normal weak yes
normal strong yes
high strong yes
normal weak yes
high strong no


NEW BINARY VECTOR
1 1 1
1 0 1
1 1 0
1 1 0
0 1 0
0 0 1
0 0 0
1 1 1
0 1 0
0 1 0
0 0 0
1 0 0
0 1 0
1 0 1

ENTROPY
```

```
instances: 14
positives: 5
negatives: 9
E(S) = 0.940286

INFORMATION GAIN
Humidity
P(S high) = 0.5
P(S normal) = 0.5
E(S high) = 0.985228
E(S normal) = 0.591673
IG(S,Humidity) = 0.151836

Wind
P(S weak) = 0.571429
P(S strong) = 0.428571
E(S weak) = 0.811278
E(S strong) = 1
IG(S,Wind) = 0.048127

Play
P(S no) = 0.357143
P(S yes) = 0.642857
E(S no) = 0
E(S yes) = 0
IG(S,Play) = 0.940286


Process finished with exit code 0
```

Output: Project Data Set

```
HAS-a-JOB    HAS-an-INSURANCE    VOTES  ACTION
Total attributes: 4
Each attribute has 10 features

HAS-a-JOB
Total different values: 2
Recognized values: yes no

HAS-an-INSURANCE
Total different values: 2
Recognized values: yes no

VOTES
Total different values: 2
Recognized values: yes no

ACTION
Total different values: 2
Recognized values: leave-alone force-into



INITIAL DATA VECTOR
HAS-a-JOB HAS-an-INSURANCE VOTES ACTION
yes yes yes leave-alone
yes no yes leave-alone
yes no no force-into
no no yes leave-alone
no no no force-into
yes yes yes leave-alone
yes no yes leave-alone
```

yes no no force-into

no no yes leave-alone

no no no force-into

NEW BINARY VECTOR

1 1 1 1

1 0 1 1

1 0 0 0

0 0 1 1

0 0 0 0

1 1 1 1

1 0 1 1

1 0 0 0

0 0 1 1

0 0 0 0

ENTROPY

instances: 10

positives: 6

negatives: 4

E(S) = 0.970951

INFORMATION GAIN

HAS-a-JOB

P(S yes) = 0.6

P(S no) = 0.4

E(S yes) = 0.918296

E(S no) = 1

IG(S,HAS-a-JOB) = 0.0199731

HAS-an-INSURANCE

P(S yes) = 0.2

P(S no) = 0.8

E(S yes) = 0

E(S no) = 1

IG(S,HAS-an-INSURANCE) = 0.170951

VOTES

P(S yes) = 0.6

P(S no) = 0.4

E(S yes) = 0

E(S no) = 0

IG(S,VOTES) = 0.970951

ACTION

P(S leave-alone) = 0.6

P(S force-into) = 0.4

E(S leave-alone) = 0

E(S force-into) = 0

IG(S,ACTION) = 0.970951

Process finished with exit code 0