Experiment no. 12

Name: Sonali Dattatray Kaingade

PRN: 21620002

Title: Attribute for classification ,Write a program to find

A. Gain

B. Gini index

For categorical and numerical values.

code:

```
#include <bits/stdc++.h>

using namespace std;


vector<string>attributes;

map<string,int>class_attri_cnt;

map<string,unordered_set<string>>values_in_attributes;

map<string,int>attribute_cnt;

map<string,map<string,int>>attribute_yn_cnt;


double calculateGain(string attri_name, double entropy){
   double info = 0.0;
   unordered_set<string>sets = values_in_attributes[attri_name];
   for(auto it : sets){
      double total = class_attri_cnt["Yes"]+class_attri_cnt["No"];
      double cnt_attri = attribute_cnt[it];
      double cnt_attri_yes = attribute_yn_cnt[it]["Yes"];
      double cnt_attri_no = attribute_yn_cnt[it]["No"];
```

```
        if(cnt_attri_yes != 0){
            info += cnt_attri/total*(-cnt_attri_yes/cnt_attri * log2(cnt_attri_yes/cnt_attri));
        }
        if(cnt_attri_no != 0){
          info += cnt_attri/total*( -cnt_attri_no/cnt_attri * log2(cnt_attri_no/cnt_attri)) ;
        }



    }
    double gain = entropy - info;
    return gain;
}

double calculategini(string attri_name){
    double gini = 0.0;
    unordered_set<string>sets = values_in_attributes[attri_name];
    //cout << sets.size() << endl;
    for(auto it : sets){
        double total = class_attri_cnt["Yes"]+class_attri_cnt["No"];
        double cnt_attri = attribute_cnt[it];
        double cnt_attri_yes = attribute_yn_cnt[it]["Yes"];
        double cnt_attri_no = attribute_yn_cnt[it]["No"];
```

```
        gini += cnt_attri/total * (1-(cnt_attri_yes/cnt_attri)*(cnt_attri_yes/cnt_attri)-
(cnt_attri_no/cnt_attri)*(cnt_attri_no/cnt_attri));


        }

        return gini;

}


int main(){

   ifstream input("info-gain.csv");


   string line, day, outlook, temp, humidity, wind, playGame;

     int j = 0;

   while(getline(input,line)){


        stringstream str(line);


        getline(str,day,',');

        getline(str,outlook,',');

        getline(str,temp,',');

        getline(str,humidity,',');

        getline(str,wind,',');

        getline(str,playGame,'.');



       if(j == 0){

        j++;
```

```
 attributes.push_back(day);
 attributes.push_back(outlook);
 attributes.push_back(temp);
 attributes.push_back(humidity);
 attributes.push_back(wind);
 attributes.push_back(playGame);
 continue;
}

class_attri_cnt[playGame]++;

values_in_attributes["outlook"].insert(outlook);
values_in_attributes["temp"].insert(temp);
values_in_attributes["humidity"].insert(humidity);
values_in_attributes["wind"].insert(wind);

attribute_cnt[outlook]++;
attribute_cnt[temp]++;
attribute_cnt[humidity]++;
attribute_cnt[wind]++;

attribute_yn_cnt[outlook][playGame]++;
attribute_yn_cnt[temp][playGame]++;
attribute_yn_cnt[humidity][playGame]++;
attribute_yn_cnt[wind][playGame]++;
```

```
    }


    double yes_cnt = class_attri_cnt["Yes"];
    double no_cnt = class_attri_cnt["No"];
    double total = yes_cnt + no_cnt;
    //cout << yes_cnt << no_cnt << total << endl;


    double entropy_ca = (-yes_cnt/total * log2(yes_cnt/total)-no_cnt/total *
log2(no_cnt/total));//entropy of class attribute
    cout <<  "Entropy of class attribute is: " << entropy_ca << endl;


    ofstream output("gain_output.csv");
    output << "Attributes"<<"," << "Gain" << "," << "Gini"<< endl;
    for(int i = 1; i<5; i++){
        double gain = calculateGain(attributes[i],entropy_ca);
        double gini = calculategini(attributes[i]);
        output << attributes[i]<<"," << gain << "," << gini<< endl;
    }
}
```

Result:

Input.csv:

day	outlook	temp	humidity	wind	playGame
1	Sunny	Hot	High	Weak	No.
2	Sunny	Hot	High	Strong	No.
3	Overcast	Hot	High	Weak	Yes.
4	Rain	Mild	High	Weak	Yes.
5	Rain	Cool	Normal	Weak	Yes.
6	Rain	Cool	Normal	Strong	No.
7	Overcast	Cool	Normal	Strong	Yes.
8	Sunny	Mild	High	Weak	No.
9	Sunny	Cool	Normal	Weak	Yes.
10	Rain	Mild	Normal	Weak	Yes.
11	Sunny	Mild	Normal	Strong	Yes.
12	Overcast	Mild	High	Strong	Yes.
13	Overcast	Hot	Normal	Weak	Yes.
14	Rain	Mild	High	Strong	No.

Output.csv:

Attributes	Gain	Gini
outlook	0.24675	0.342857
temp	0.029223	0.440476
humidity	0.151836	0.367347
wind	0.048127	0.428571