**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]);

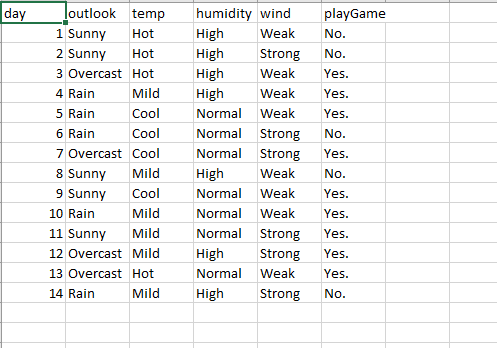
output << attributes[i]<<"," << gain << "," << gini<< endl;

}

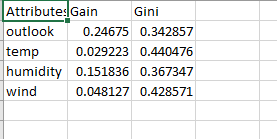
}

**Result:**

**Input.csv:**

****

**Output.csv:**

****