ALI HELMUT

1301154246

1. CREATE A FUNCTION TO CALCULATE PERFORMANCE OF CLASSIFIER THAT IMPLEMENTS F1-MICRO AVERAGE, F1- MACRO AVERAGE, AND SIMPLE ACCURACY (I.E. NUMBER OF CORRECT PREDICTION DIVIDED NUMBER OF DATA). (HINTS: GIVE AN INPUT PARAMETER OF THE FUNCTION THAT INFORM THE FUNCTION WHEN IT SHOULD USE EITHER F1-MICRO AVERAGE, F1-MACRO AVERAGE, OR SIMPLE ACCURACY). [5 POINTS]

```
function [ perfom] = getPerfome( dataTrain , dataTest , jenisOutput )
%jenis: 1=micro , 2=macro ,3=simple accu
    a = unique(dataTrain(:,3));
    hasil = []; %tp fn fp tn
    for i=1 : length(a)
        tp = 0;
        tn = 0;
        fp = 0;
        fn = 0;
        for j=1 : length(dataTrain)
            if(dataTrain(j,3) == a(i))
                if(dataTest(j,3) == dataTrain(j,3))
                    tp = tp+1;
                end
                if(dataTest(j,3) ~= dataTrain(j,3))
                    fn = fn+1;
                end
            else
                if(a(i) == dataTrain(j,3))
                    fp = fp+1;
                end
                if(a(i) ~= dataTrain(j,3))
                    tn = tn+1;
                end
            end
        end
        hasil = [ hasil; tp fn fp tn];
    end

    hasilPerKelas = []; % recall precision f1
    for i=1 : length(hasil);
           recall           = hasil(i,1) / ( hasil(i,1) + hasil(i,2));
           precision        = hasil(i,1) / ( hasil(i,1) + hasil(i,3));
           f1               = ( 2 * recall * precision ) / ( precision +
recall);
           hasilPerKelas    = [hasilPerKelas; recall precision f1];
    end

    if(jenisOutput == 2)
       perfom = mean(hasilPerKelas(:,3));
    elseif (jenisOutput == 1)
        r = mean(hasilPerKelas(:,1));
        p = mean(hasilPerKelas(:,2))
        perfom      = ( 2 * r * p ) / ( r + p);
    elseif (jenisOutput == 3)
        tp = 0;
        for i=1 : length(hasil)
            tp = tp + hasil(i,1);
        end
```

```
        perfom = tp / length(dataTrain(:,3));
    end
end
```

2. (45 POINTS) IN THIS PROBLEM WE IMPLEMENT NAIVE BAYES FOR CLASSIFICATION.

(A) LOAD THE SELECTED DATA SET. VISUALIZE ALL DATA POINTS USING SCATTER PLOT. FROM YOUR SCATTER PLOT, ONE COULD EASILY DISTINGUISH EACH CLASS. (HINT: USE ATTRIBUTE 1 AS X -AXIS, ATTRIBUTE 2 AS Y -AXIS. USE DIFFERENT COLOR AND/OR DIFFERENT SYMBOL FOR EACH CLASS LABEL.[5 POINTS]

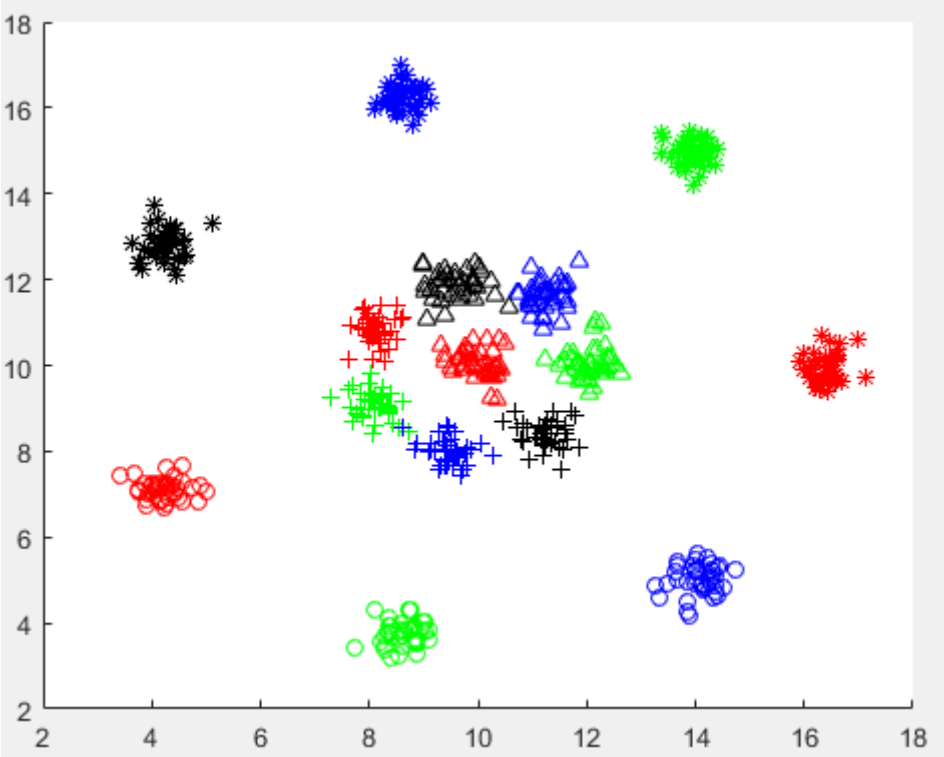


Figure 1 data train

```
function plotinHasil (datas)

a = datas;

b = (datas(:,1:2));

for i=1:size(a,1)

   if a(i,3) == 1

      scatter(b(i,1), b(i,2), '^', '.r');

      hold on;
```

```
elseif a(i,3) == 2
   scatter(b(i,1), b(i,2), '^', '.g');
   hold on;
elseif a(i,3) == 3
   scatter(b(i,1), b(i,2), '^','.b');
   hold on;
elseif a(i,3) == 4
   scatter(b(i,1), b(i,2), '^','.k');
   hold on;
elseif a(i,3) == 5
   scatter(b(i,1), b(i,2), '+','.r');
   hold on;
elseif a(i,3) == 6
   scatter(b(i,1), b(i,2), '+','.g');
   hold on;
elseif a(i,3) == 7
   scatter(b(i,1), b(i,2), '+','.b');
   hold on;
elseif a(i,3) == 8
   scatter(b(i,1), b(i,2), '+','.k');
   hold on;
elseif a(i,3) == 9
   scatter(b(i,1), b(i,2), '*','.r');
   hold on;
elseif a(i,3) == 10
   scatter(b(i,1), b(i,2), '*','.g');
   hold on;
elseif a(i,3) == 11
   scatter(b(i,1), b(i,2), '*','.b');
   hold on;
elseif a(i,3) == 12
   scatter(b(i,1), b(i,2), '*','.k');
   hold on;
elseif a(i,3) == 13
   scatter(b(i,1), b(i,2), 'o','.r');
   hold on;
elseif a(i,3) == 14
   scatter(b(i,1), b(i,2), 'o','.g');
   hold on;
elseif a(i,3) == 15
   scatter(b(i,1), b(i,2), 'o','.b');
   hold on;
end

end
```

(B) APPLY NAIVE BAYES CLASSIFIER ON THE SELECTED DATA SET. YOUR CODES HAVE TO CLEARLY CONTAIN

I. FUNCTION FOR LEARNING THAT IMPLEMENTS THE CALCULATION OF PRIOR AND LIKELIHOOD PROBABILITIES. [10 POINTS]

```
function [ outputPrior ] = getPrior( data )
a = unique(data);
b = [];
total = 0;
for i=1 : length(a)
        d =0;
        for j=1 : length(data)
                if(data(j) == a(i))
                d = d + 1;
                end
        end
        b = [ b ; a(i) d/length(data) d ];
end
outputPrior = b;
end


function [ classOutput ] = getLikeliHood( datInput, meanAtribute , Variance )
likelihood = [];
for i=1 : 15
        likelihood = [ likelihood ;(1/(Variance(i,1)*sqrt(44/7)))* exp(-1*(datInput(1,1)-
        meanAtribute(i,1)).^(2))/(2*(Variance(i,1).^(2))), (1/(Variance(i,2)*sqrt(44/7)))*
        exp(-1*(datInput(1,2)-meanAtribute(i,2)).^(2))/(2*(Variance(i,2).^(2)))];
end
classOutput = likelihood;
end
```

II. FUNCTION FOR PREDICTING/CLASSIFYING DATA THAT IMPLEMENTS THE CALCULATION OF POSTERIOR PROBABILITIES. IT HAD BETTER FOR YOU TO IMPLEMENT THE LOGARITHM (LOG) FUNCTION.[10 POINTS]

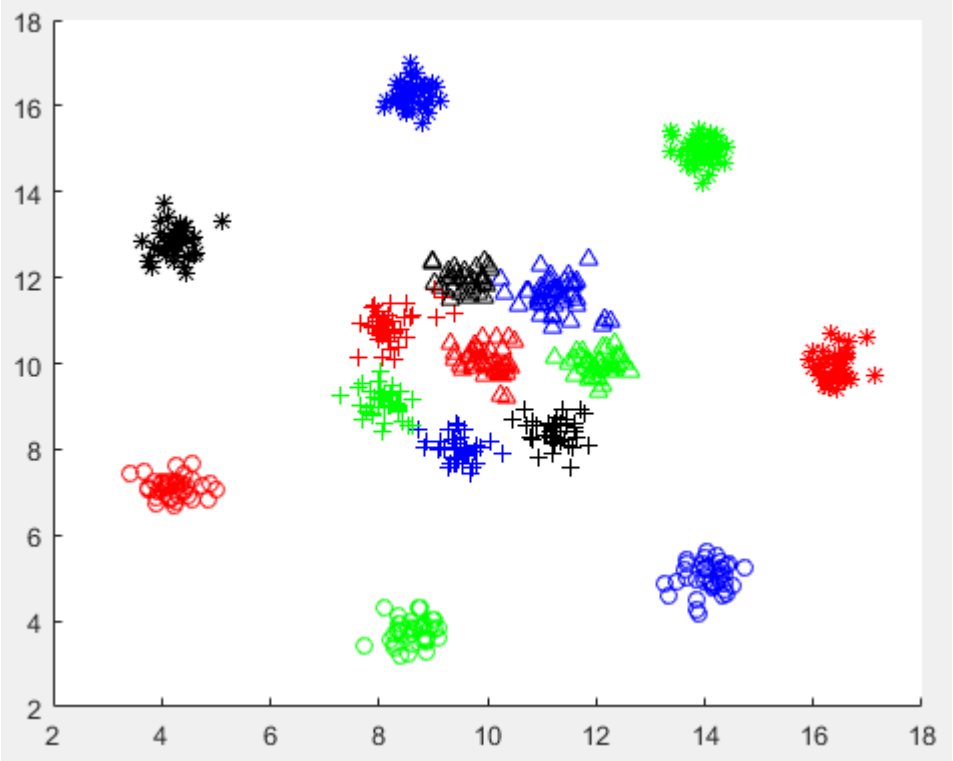
```
function [ outputClass ] = naiveBayes( dataTrain , meanOfAtribute , variance , prior )
likelihood     = [];
posterior      = [];
[likelihood]    = getLikeliHood( dataTrain, meanOfAtribute , variance );
[posterior]     = getPosterior( likelihood , prior );
[outputClass]   = getClass(posterior);
end

function [ outputPosterior ] = getPosterior( likeLiHood , prior )
a = []
for i=1 : 15
            a = [a ; likeLiHood(i,1) * likeLiHood(i,2) * prior(i)];
end
outputPosterior = a;
end

function [ outputClass ] = getClass( porsterior )
   a = [];
   for i=1 : 15
      a = [ a ; porsterior(i),i];
   end
   a = sortrows(a,1);
   outputClass = a(15,2);
end
```

III. CLASSIFY EACH DATA POINT USING THE TRAINED NAIVE BAYES. PLOT THE RESULTS USING SCATTER PLOT. FROM YOUR SCATTER PLOT, ONE COULD EASILY DISTINGUISH EACH CLASS. (HINT: USE DIFFERENT COLOR AND/OR DIFFERENT SYMBOL).[5 POINTS]

```
dataTest = []
for i=1 : 600
    likelihood      = [];
    posterior       = [];
    [likelihood]    = getLikeliHood( dataTrain(i,:), meanOfAtribute , variance );
    [posterior]     = getPosterior( likelihood , prior(:,2) );
    [outputClass]   = getClass(posterior);
    dataTest        = [dataTest ; dataTrain(i,1),dataTrain(i,2), outputClass];
end
```



Gambar 2 data Test

IV. BY VISUALLY COMPARING FIGURES CREATED FROM POINT 2(A) AND 2(B)III, WHAT DO YOU THINK OF THE CLASSIFICATION RESULTS USING NAIVE BAYES. [5 POINTS]

Hasil clasifikasi menggunakan naïve bayes masih sangat bagus, Karena hanya beberapa object saja yang gagal klasifikasi

(C) USE THE FUNCTION OF EXERCISE 1 TO EVALUATE PERFORMANCE OF NAIVE BAYES CLASSIFIER. WHAT IS THE F1-MICRO AVERAGE OF NAIVE BAYES CLASSIFIER ON THE DATA SET. [5 POINTS]

F1 micro average = 0,989898989898990

(D) PLOT THE DECISION BOUNDARY RESULTED FROM NAIVE BAYES CLASSIFIER ON THE FIGURE THAT HAS BEEN CREATED BY POINT 2(A). (HINTS: GENERATE DATA POINTS USING RANGE OF MINIMUM AND MAXIMUM VALUE OF EACH ATTRIBUTE, THEN CLASSIFY EACH GENERATED DATA POINTS USING TRAINED NAIVE BAYES CLASSIFIER. USE ATTRIBUTE 1 AND ATTRIBUTE 2 AS BOTH X -AXIS AND Y