**CROP RECOMMENDATION SYSTEM**

**USING DATAMINING IN R**

Submitted in partial fulfillment

Of

Mini project in Bachelor of Technology

Submitted by

1.**A.LAVANYA**

2**.D.RAMYASREE**

3**.K.C.JYOTSNA**

Under the Supervision of

Mr.BKSP Kumar Raju Alluri



**NATIONAL INSTITUTE OF TECHNOLOGY,**

**Andhra Pradesh.**

**INTRODUCTION:**

Datamining is knowledge discovery from data (KDD).Its applications are used in various fields like business , government for future analysis.

The future analysis can be done by usingTechniques like association mining , classification ,Clustering etc,

We are trying to recommend the best crop For the farmer with respect to various factors by using Association mining, classification, clustering techniques in R and Weka.

**DATASET:**

The data set is prepared manually by us.The Dataset has details of crop,soil,rainfall etc,It consists of 17 attributes,301 tuples.

The attributes are:

1.Season:In which season the crop was grown Eg:Kharif,Rabi.

2.Crop:The crop which is grown Eg:Bajra,Rice etc,

3.N(kilograms):The content of Nitrogen used for growing the crop.

4.P(kilograms): The content of Phosporous used for growing the crop.

5.K(kilograms): The content of Potassium used for growing the crop.

6.Other fertilizers:Other than N,P,K the fertilizers which are used.Eg:FYM(Farm Yard Manure).

7.Brand:The brand of the fertilizers used. Eg:CFL(Coromondal Fertilizers Limited).

8.Area(in acres):The area in which the crop is grown.

9.Automation:Based on the equipments and tools used We will say the level of automation. Eg:M-Modern,T-Traditional,A-Average,G-Good.

10.Irrigation type:The type of irrigation system used For growing that particular crop.

Eg:Canal irrigationetc;

11.Soiltype:The type of soil in which the crop is grown. Eg:black soil,laterite soil etc,

12.Rainfall(cm):The amount of rainfall received when that crop is grown.

13.Temperature(degrees Celsius):The amount of Average temperature when that crop is grown.

14.Production(in kilograms for acre):The amount of Outcome received from that particular crop.

15:Prediction of Profit or Loss: how much profit the farmer got.Eg:1-verylow,2-low,3-medium,4-high.

16:Distance:we are assuming particular place as

Center from that place we are calculating distance of

place where the crop is grown.

17:State:The state in which the crop is grown.

We are using additional two datasets(dataset1-which has the details of rawmaterials used for that particular

Crop and dataset2-which has the details of the production of rawmaterials in different states and their distance from the center which we assumed).

**PROBLEM STATEMENT:**

The farmers in our country are mostly illiterate and they don’t know properly where to grow, when to grow, what conditions needed to grow, where they have to import rawmaterials for their crop.

Also many farmers grow the same crop in a particular region which again leads to the fall in the prices of that crop and leads to lower profits.

So we are developing this system to Overcome all these problems and help the farmers to Gain maximum profits.

**SCHEMATIC VIEW OF PROJECT:**

**MODULE1:**

ASSOCIATION RULES

DATA

SET

APRIORI

ALGORITHM

**MODULE2:**

CLASSIFICATION

ALGORITHM

CLASSIFICATION

RULES

PREDICTION

NEW DATA

DATA SET

**MODULE3:**

DATASET

K-MEANS CLUSTERING

CLUSTERS

PREDICT

**MODULE4:**

DATA SET

CLASSIFICATION

RULES

CLASSIFICATION

ALGORITHM

COMPARING

DISTANCE

PREDICTION

NEW DATA

**MODULE5:**

STATE WITH HIGHEST PRODUCTION

RAWMATERIAL

IMPORT MATERIAL FROM

IMPORT MATERIAL FROM

STATE WITH LOWEST DISTANCE

**DETAILED EXPLANATION OF MODULES:**

MODULE1:

This module generates association rules using apriori algorithm .We used R to generate association rules with template matching. In this module, we will suggest the farmer for a particular crop what are the required factors like rainfall, soil, temperature, N, P, K contents should be used for getting high profits.

MODULE2:

We used multiple regression technique in order to predict the profit/loss(class label) when user Gives his required inputs. Not only we can predict class labels from this we can also predict production also from this and also we can say which attribute is mostly influencing the class label or production.

MODULE3:

Here initially we applied k-means clustering technique and then classified our entire dataset into four clusters.Then we ask the user to input a new tuple then for this new tuple we will predict into which cluster this new tuple belongs to.We also give the graphical representation of the clusters.

We also used a 3-D scatter plot to obtain clusters with 3 dimensions.We will get a 3D representation of the clusters here.

MODULE4:

Even if we predict a farmer that according to his inputs he may get higher profits,if more farmers already in that place are growing that particular crop.Then there will be fall in price for that crop in market and all those farmers will get loss who are growing the same crop.Hence we will take a threshold distance and threshold count.

For that particular new crop which is given as input from the farmer we will check already existing datasets and we will see how many farmers are already growing the same crop and calculate distances.If the distance is less than threshold distance then we will check the count.If the count is more than the Threshold count then we will decrease his classlabel(means even he gets profit of class label 4 we will predict it as 3).

MODULE5:

There are certain rawmaterials for all crops which are not available in that particular place/state.

Hence they should import those rawmaterials from other places.If in that other place also if the production is low,they will sell that rawmaterial for a higher cost.And also if that other state from which we need to import is far away then again the cost is increased for transport.Hence for each crop we will say what are the rawmaterials they need to import and from which state they have to import.Based on distance and production we will recommend them the states from which they can import their rawmaterials.So that there won’t be wastage of money for the farmers.

EXPERIMENTS AND EVALUATION:

1. Initially we went with a project proposal related to supermarket. But we don’t have much clarity and also whatever we thought to do already exist. So our proposal is rejected.

2. We went to sir for second time with a new idea about crop recommendation system .Then our idea is accepted.

3. Sir instructed us to study base paper and explain to him. We took a base paper to him and explained .But they used MATLAB and applied artificial neural networks in their paper. But it’s too complicated.

4. We decided to do everything using R only and tried to implement the system as simple as possible so that even someone sees our code they can understand it easily.

5. We searched for datasets in many sites like UCI repository, data world etc , but we failed to find the required dataset.So we started to prepare our own dataset.

6. Then we took our dataset and started applying apriori in weka but we didn’t find any useful rules.We also tried classification and clustering in weka.We thought weka is a limitation to our project.So we shifted to R.The highest accuracy for our dataset is with SVM:61%.

7. We wrote apriori code in R but those rules are also not useful. At that time sir told to us to do rough sets Technique on our dataset , we referred roughsets also But we can’t understand their implementation as it is too complex.We again created another dataset.

8. Again we used template matching and generated new required rules on the new dataset and we succeeded.

9. We then did clustering where initially the dataset is divided into four clusters by k-means clustering technique, we also did scatter plot on our dataset.

10. We predicted the classlabel according to user requirements. Like this we completed our existing module.

11. Then we started novelity .The first novelity is whenever the crop the farmer thought if growing, even though it will get high profits we will check the distances of the same crops which are grown nearer to that area , if there are more crops like that then we will decrease the profit prediction or else predict the same classlabel.

12. The second novelity is to request the farmer from which state he has to import the rawmaterials for his crop by checking the production and distance so that the cost won’t be increased.

13. These are the various modules of our project.

CONCLUSION:

Hence we addressed the problem statement and provided solutions to the problems. Now from this system the farmer knows the required conditions for growing that particular crop to get highest profits. Also if we give certain inputs he will get to know what profit /production he will get. And from where he should import his rawmaterials. We also addressed the problem of over production of a particular crop in a particular place.

REFERENCES:

1. IEEE Base paper on crop recommendation system using MATLAB.

2. UDEMY courses for learning R language.

3. Youtube, Google which really helped us in succeeding.

FUTURE WORK:

Actually we are considering distances the distance of that place from a fixed point and we are taking the radius of that new place as distance.

But it is not useful when we do in real time applications.The farmer may or may not know their Distance from the center.So,if we use Google maps

And ask them to switch on their gps we should get to know the latitude and longitude of their places and it should calculate distances automatically.

And also the entire system may be in the form of a website or a mobile app which can be used by all.

**PROJECT CODE**

**-------START OF APRIORI CODE------**

p<-read.csv(file.choose())//Choose our dataset

df<-data.frame(p)//it is converted into dataframe

df$N<-cut(df$N,br=c(0,40,80,100,200,400),labels=c("low","bavg","avg","aavg","high"))//a

df$P<-cut(df$P,br=c(0,40,80,100,200,400),labels=c("low","bavg","avg","aavg","high"))//b

df$K<-cut(df$K,br=c(0,40,80,100,150,200),labels=c("low","bavg","avg","aavg","high"))//c

df$area<-cut(df$area,br=c(0,5,10),labels=c("less","more"))//d

df$rainfall<-cut(df$rainfall,br=c(0,50,100,150,200),labels=c("low","avg","aavg","high"))//e

df$temperature<-cut(df$temperature,br=c(0,20,25,30,35,50),labels=c("low","bavg","avg","aavg","high"))//f

df$production<-cut(df$production,br=c(0,200,400,600,800,1000),labels=c("low","bavg","avg","aavg","high"))//g

df$distance<-cut(df$distance,br=c(0,100,200,300,400,500),labels=c("low","bavg","avg","aavg","high"))//h

//in a,b,c,d,e,f,g,h we are discretizing all the numeric attributes.

Library(arules)

rules1<-apriori(df,parameter=list(minlen=2,maxlen=20,support=0.05,confidence=0.1),appearance=list(rhs=c("prediction.of.profit.loss=H"),default="lhs"))//we are using apriori function,here df is the datset,in appearance we are fixing the RHS so that we will get rules only having highest profits.

rules1

inspect(rules1)//for viewing the rules

inspect(subset(rules1,subset=lhs %pin% "crop="))//

from the rules obtained we are again taking into consideration which have crop attribute as a subset in them.

------END OF APRIORI CODE------

--------Start of clustering code-----------

data<-read.csv(file.choose())//choosing the dataset

df<-data[,c(3,4,5,8,12,13,14)]//consider only numerical attributes

result <- kmeans(df,4);//applying kmeans,df=dataset,4 clusters

result//viewing the row number with their cluster number

plot(df[c("area","production")],col=result$cluster)//graphical representation of the clusters

result$cluster

legend("topleft",inset=0.01,legend=c("cluster1","cluster2","cluster3","cluster4"),col=(1:4),lty=1,horiz=F)//keeping legend for the plot of the clusters obtained

library("ggpubr")

ggscatter(df, x = "area", y = "production",

add = "reg.line", conf.int=TRUE, cor.coef = TRUE, cor.method = "pearson",

xlab = "area(in hectares)", ylab = "production")//scatter plot for area and production using pearson correlation method

data1<-data//taking the dataset into data1

data1<-myfunction(data1,"Rabi","Rice",90,70,50,"NULL","CFL",15000,"T","canal irrigation","black",150,35,10000,"M")//taking inputs

myfunction<-function(da,a,b,c,d,e,f,g,h,i,j,k,m,n,q,s){

l<-nrow(da)//l has no of rows of dataset

da[l+1,1]<-a

da[l+1,2]<-b

da[l+1,3]<-c

da[l+1,4]<-d

da[l+1,5]<-e

da[l+1,6]<-f

da[l+1,7]<-g

da[l+1,8]<-h

da[l+1,9]<-i

da[l+1,10]<-j

da[l+1,11]<-k

da[l+1,12]<-m

da[l+1,13]<-n

da[l+1,14]<-q

da[l+1,15]<-s//inserting those attributes

return(da)//return the dataset with the added tuples

}

g1<-data1[,c(3,4,5,8,12,13,14)]//taking numeric values

g2<-kmeans(g1,4)//apply kmeans again

g2//check the cluster into which the new tuple has fallen into

plot(g1[c("area","production")],col=g2$cluster)//plot the cluster

g2$cluster

legend("topleft",inset=0.01,legend=c("cluster1","cluster2","cluster3","cluster4"),col=(1:4),lty=1horiz=F)//use legend the for the clusters again

---------END OF CLUSTER CODE--------

---------START OF NOVELITY1---------

d<-read.csv(file.choose())//Choosing dataset

data1<-d//storing it in data1

str(data1)

m<-lm(prediction.of.profit.loss~crop+area+N+P+K+rainfall+temperature,data=data1)//create a learned model

summary(m)//obtain which attribute mostly affects the attribute taken on LHS

s<-readline(prompt="enter season:")

c<-readline(prompt="enter crop:")

n<-readline(prompt="enter N:")

n<-as.integer(n)

p<-readline(prompt="enter P:")

p<-as.integer(p)

k<-readline(prompt="enter K:")

k<-as.integer(k)

f<-readline(prompt="enter other fertilizers:")

b<-readline(prompt="enter brand:")

a<-readline(prompt="enter area:")

a<-as.integer(a)

ma<-readline(prompt="enter automation:")

i<-readline(prompt="enter irri\_type:")

so<-readline(prompt="enter soiltype:")

r<-readline(prompt="enter rainfall:")

r<-as.integer(r)

t<-readline(prompt="enter temperature:")

t<-as.integer(t)

pr<-readline(prompt="enter production:")

pr<-as.integer(pr)

d1<-readline(prompt="enter distance:")//ask input from user

d1<-as.integer(d1)//input stored into d1

typeof(s)

x<-predict(m,data.frame("crop"=c,"N"=n,"P"=p,"K"=k,"rainfall"=r,"area"=a,"temperature"=t))//by taking inputs from user we will predict the attribute in lhs in learned model

x//class label predicted

if((0<x)&&(x<=1.5)){ //dicretising the class label means the class label must have only four values either 1 or 2 or 3 or 4

x<-1

}else if((1.5<x)&&(x<=2.5)){

x<-2

}else if((2.5<x)&&(x<=3.5)){

x<-3

}else

x<-4

x

tail(data1)

library(pracma)//to include strcmpi function

thres<-20//taking some threshold distance

count<-0//setting count to zero

near<-3//threshold count

n1<-nrow(data1)//taking number of rows of data1 to n1

for(i in 1:n1)//comparing the input crop with other crops in the dataset

{

result<-strcmpi(as.character(data1[i,2]),c)

if(result)

{

d2<-abs(data1[i,16]-d1)

if(d2<=thres)

count<-count+1

}

}

if(count>=near)//if more than threshold count

{

if(x==4)

x<-3 //decreasing the classlabel

else if(x==3)

x<-2

else if(x==2)

x<-1

else

x<-1

}

x//print the classlabel

**---------END OF NOVELITY1------------**

--------**START OF NOVELITY2**------------

t1<-read.csv(file.choose())//choosing dataset

t2<-read.csv(file.choose())//choosing dataset

install.packages("sqldf")

library(sqldf)

data<-sqldf("SELECT crop,t1.rawmaterial,AVG(punjab) as p,AVG(madhyapradesh) as m,AVG(himachalpradesh) as h,AVG(mizoram) as z,AVG(ABS(d-distance1)) as pd,AVG(ABS(d-distance2)) as md,AVG(ABS(d-distance3)) as hd,AVG(ABS(d-distance4)) as zd FROM t1,t2 WHERE t1.rawmaterial=t2.rawmaterial GROUP BY crop,t1.rawmaterial

")//storing crop with its rawmaterial,average production in different states,average distance from the state where the crop is grown to different states into data

Data//print data

d<-sqldf("SELECT crop,t1.rawmaterial,MAX(AVG(punjab),AVG(madhyapradesh),AVG(mizoram),AVG(himachalpradesh)) AS MAXPRO,MIN(AVG(ABS(d-distance1)),AVG(ABS(d-distance2)),AVG(ABS(d-distance3)),AVG(ABS(d-distance4))) AS MINDIS FROM t1,t2 WHERE t1.rawmaterial=t2.rawmaterial GROUP BY crop,t1.rawmaterial

")//storing crop with its rawmaterial,max production of rawmaterial in particular state,min distance from state where crop is grown to different state into d

d//printing d

c<-readline(prompt="enter crop")

Banana

x<-readline(prompt="enter rawmaterial1 for crop")

zinc

y<-readline(prompt="enter rawmaterial2 for crop")

sulphur

if(sqldf("SELECT MAXPRO from d where crop='Banana' and rawmaterial='zinc'")==sqldf("SELECT AVG(punjab) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='zinc'"))

{

print("for growing banana crop recommendable state for importing zinc rawmaterial is punjab w.r.t production")

}

else if(sqldf("SELECT MAXPRO from d where crop='Banana' and rawmaterial='zinc'")==sqldf("SELECT AVG(madhyapradesh) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='zinc'"))

{

print("for growing banana crop recommendable state for importing zinc rawmaterial is madhyapradesh w.r.t production")

}

else if(sqldf("SELECT MAXPRO from d where crop='Banana' and rawmaterial='zinc'")==sqldf("SELECT AVG(himachalpradesh) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='zinc'"))

{

print("for growing banana crop recommendable state for importing zinc rawmaterial is himachalpradesh w.r.t production")

}

else{

print("for growing banana crop recommendable state for importing zinc rawmaterial is mizoram w.r.t production")

}//printing recommendable state for particular crop for importing rawmaterial w.r.t max production

if(sqldf("SELECT MINDIS from d where crop='Banana' and rawmaterial='zinc'")==sqldf("SELECT

AVG(ABS(d-distance1)) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana'

and t1.rawmaterial='zinc'"))

{

print("for growing banana crop recommendable state for importing zinc rawmaterial is punjab w.r.t distance")

}

else if(sqldf("SELECT MINDIS from d where crop='Banana' and rawmaterial='zinc'")==sqldf("SELECT AVG(ABS(d-distance2)) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='zinc'"))

{

print("for growing banana crop recommendable state for importing zinc rawmaterial is madhyapradesh w.r.t production")

}

else if(sqldf("SELECT MINDIS from d where crop='Banana' and rawmaterial='zinc'")==sqldf("SELECT AVG(ABS(d-distance3)) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='zinc'"))

{

print("for growing banana crop recommendable state for importing zinc rawmaterial is himachalpradesh w.r.t distance")

}

else{

print("for growing banana crop recommendable state for importing zinc rawmaterial is mizoram w.r.t distance")

}//printing recommendable state for particular crop for importing rawmaterial w.r.t min distance

if(sqldf("SELECT MAXPRO from d where crop='Banana' and rawmaterial='sulphur'")==sqldf("SELECT AVG(punjab) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='sulphur'"))

{

print("for growing banana crop recommendable state for importing sulphur rawmaterial is punjab w.r.t production")

}

else if(sqldf("SELECT MAXPRO from d where crop='Banana' and rawmaterial='sulphur'")==sqldf("SELECT AVG(madhyapradesh) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='sulphur'"))

{

print("for growing banana crop recommendable state for importing sulphur rawmaterial is madhyapradesh w.r.t production")

}

else if(sqldf("SELECT MAXPRO from d where crop='Banana' and rawmaterial='sulphur'")==sqldf("SELECT AVG(himachalpradesh) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='sulphur'")){

print("for growing banana crop recommendable state for importing sulphur rawmaterial is himachalpradesh w.r.t production")

}

else{

print("for growing banana crop recommendable state for importing sulphur rawmaterial is mizoram w.r.t production")

}

if(sqldf("SELECT MINDIS from d where crop='Banana' and rawmaterial='sulphur'")==sqldf("SELECT AVG(ABS(d-distance1)) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='sulphur'"))

{

print("for growing banana crop recommendable state for importing sulphur rawmaterial is punjab w.r.t distance")

}

else if(sqldf("SELECT MINDIS from d where crop='Banana' and rawmaterial='sulphur'")==sqldf("SELECT AVG(ABS(d-distance2)) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='sulphur'"))

{

print("for growing banana crop recommendable state for importing sulphur rawmaterial is madhyapradesh w.r.t production")

}

else if(sqldf("SELECT MINDIS from d where crop='Banana' and rawmaterial='sulphur'")==sqldf("SELECT AVG(ABS(d-distance3)) FROM t1,t2 where t1.rawmaterial=t2.rawmaterial and crop='Banana' and t1.rawmaterial='sulphur'"))

{

print("for growing banana crop recommendable state for importing sulphur rawmaterial is himachalpradesh w.r.t distance")

}

else{

print("for growing banana crop recommendable state for importing sulphur rawmaterial is mizoram w.r.t distance")

}

------------END OF NOVELITY2-----------