Name:Islam,Rafiqul **Id:17-34438-1** **Sec:A**

(Unsupervised project without vitamin and rating)

Title: In this project we used a dataset of cereals along with their different percentage of calories, protein, fat etc.

Description: This dataset has different 77 cereals. After analyzing it helps to select the nutritional food for health requirements. The people who have different diseases like diabetes, high blood pressure, low blood pressure can choose food for their breakfast according to the elements they need and avoid those elements which contains less amount. This will help to choose all the necessary food elements that they need.

Literature Survey:

(Without 'vitamin' and 'rating' columns): 77 x 9 here:
http://www.cs.umd.edu/hcil/hce/examples/cereal/cereal.txt

The meaning of each column:

1. 1st column: Name of cereal
2. Calories: calories per serving
3. Protein: grams of protein
4. Fat: grams of fat
5. Sodium: milligrams of sodium
6. Fiber: grams of dietary fiber
7. Carbo: grams of complex carbohydrates
8. Sugars: grams of sugars
9. Potass: milligrams of potassium
10. Shelf: display shelf (1, 2, or 3, counting from the floor)

Methodology:

The dataset has been clustered by the hierarchical clustering technique. The cluster tree has been cut in several places. Then similarities between instances of individual clusters and dissimilarities between instances of different clusters have been analyzed.

Hierarchical cluster tree with cutting point:

=== Run information ===

Scheme: weka.clusterers.HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.EuclideanDistance -R first-last"

Relation: cerealWithoutVitamin

Instances: 77

Attributes: 10

- cerealName
- calories
- protein
- fat
- sodium
- dietaryfiber
- complexcarbohydrates
- suger
- displayshelf
- potassium

Test mode: evaluate on training data

=== Clustering model (full training set) ===

Cluster 0

((((1:0.44226,3:0.44226):0.1363,4:0.57856):0.08163,(((2:0.56333,((((((((((5:0.33764,(((((((8:0.20516,50:0.20516):0.06193,52:0.26708):0.00672,40:0.2738):0.01039,(((14:0.18836,60:0.18836):0.0719,20:0.26026):0.02185,(((33:0.2381,57:0.2381):0.00319,72:0.24129):0.00859,(34:0.19672,51:0.19672):0.05316):0.03223):0.00208):0.01722,((22:0.22189,(70:0.17355,73:0.17355):0.04835):0.04479,(24:0.24792,39:0.24792):0.01877):0.03473):0.01636,23:0.31778):0.01982,28:0.33761):0.00004):0.02293,35:0.36057):0.02023,54:0.3808):0.01647,((29:0.3668,(53:0.30072,71:0.30072):0.06607):0.01968,((45:0.17188,46:0.17188):0.15453,47:0.3264):0.06007):0.0108):0.00187,10:0.39914):0.10546,(((((((7:0.20954,25:0.20954):0.00491,(((15:0.03021,19:0.03021):0.14149,(30:0.01562,74:0.01562):0.15608):0.03784,43:0.20954):0.00491):0.01581,67:0.23027):0.01888,49:0.24915):0.00601,18:0.25516):0.01214,((11:0.0996,36:0.0996):0.15227,13:0.25187):0.01542):0.07879,32:0.34608):0.15117,41:0.49725):0.00735):0.0185,((((6:0.33799,((9:0.20919,(48:0.207,77:0.207):0.00219):0.09722,(75:0.09496,76:0.09496):0.21145):0.03158):0.01512,37:0.35311):0.01197,(26:0.0996,38:0.0996):0.26547):0.10996,(((16:0.04347,63:0.04347):0.09441,17:0.13787):0.11011,62:0.24798):0.22706):0.04806):0.01632,59:0.53943):0.00315,(((27:0.26217,69:0.26217):0.18008,44:0.44226):0.07239,61:0.51464):0.02793):0.00624,42:0.54881):0.00309,31:0.5519):0.00512,21:0.55703):0.0063):0.03183,(55:0.26816,56:0.26816):0.32701):0.00491,(64:0.28591,(65:0.10242,66:0.10242):0.1835):0.31415):0.06013):0.02044,(12:0.49034,68:0.49034):0.19029)

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 76 (99%)

1 1 (1%)

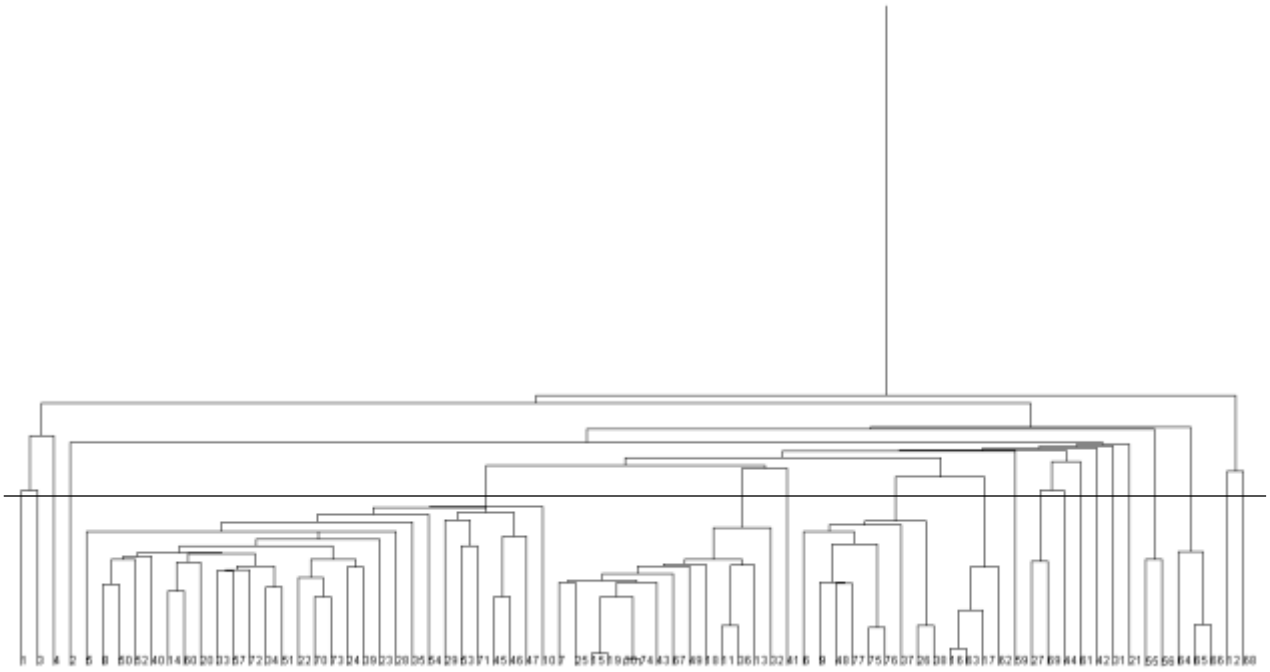

Fig 1: Hierarchical cluster tree with cutting point

Cluster Analysis:

For Cluster-1(1 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg
100%_Bran	70	4	1	130	10	5	6	3	280

Findings: Calories(medium),potass (medium), fat(low),sodium(medium), fiber(High) , carbo(low),sugar(low),potass(high)

For Cluster-2(1 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)

All-Bran	70	4	1	260	9	7	5	3	320

Findings: Calories(medium),protein(medium),potass(high), fat(low), fiber(high) , sugar(low), carbo(low)

For Cluster-3(1 instances) :

cereal name	calories	protein (g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
All-Bran_with_Extra_Fiber	50	4	0	140	14	8	0	3	330

Findings: Calories(low),sugars(0), fat(0), potass(high),fiber(high),carbo(low),shelf(high), sodium(low)

For Cluster-4(1 instances):

cereal name	calories	protein (g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars (g)	display shelf	potassium(mg)
100%_Natural_Bran	120	3	5	15	2	8	8	3	135

Findings:

Calories(high),protein(medium),fiber(low),carbo(low),potass(low),sugars(medium), sodium(low), fat(high)

For Cluster-5(29 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carboh	sugars(g)	display shel	potassium(mg)

						ydrate s(g)		f	
Almond_Delight	110	2	2	200	1	14	8	3	-1
Basic_4	130	3	2	210	2	18	8	3	100
Nutri-Grain_Almond-Raisin	140	3	2	220	3	21	7	3	130
Oatmeal_Raisin_Crisp	130	3	2	170	1.5	13.5	10	3	120
Just_Right_Fruit_&_Nut	140	3	1	170	2	20	9	3	95
Clusters	110	3	2	140	2	13	7	3	105
Cracklin'_Oat_Bran	110	3	3	140	4	10	7	3	160
Grape_Nuts_Flakes	100	3	1	140	3	15	5	3	85
Raisin_Nut_Bran	100	3	2	140	2.5	10.5	8	3	140
Quaker_Oat_Squares	100	4	1	135	2	14	6	3	110
Total_Whole_Grain	100	3	1	200	3	16	3	3	110
Grape-Nuts	110	3	0	170	3	17	3	3	90
Nutri-grain_Wheat	90	3	0	170	3	18	2	3	90
Crispix	110	2	0	220	1	21	3	3	30
Double_Chex	100	2	0	190	1	18	5	3	80
Total_Corn_Flakes	110	2	1	200	0	21	3	3	35
Triples	110	2	1	250	0	21	3	3	60
Just_Right_Crunchy__Nuggets	110	2	1	170	1	17	6	3	60
Crispy_Wheat_&_Raisins	100	2	1	140	2	11	10	3	120

Fruit_&_Fibre_Dates,_Walnuts,_and_Oats	120	3	2	160	5	12	10	3	200
Great_Grains_Pecan	120	3	3	75	3	13	4	3	100
Fruitful_Bran	120	3	0	240	5	14	12	3	190
Post_Nat._Raisin_Bran	120	3	1	200	6	11	14	3	260
Product_19	100	3	0	320	1	20	3	3	45
Total_Raisin_Bran	140	3	1	190	4	15	14	3	230
Muesli_Raisins,_Dates,_&_Almonds	150	4	3	95	3	16	11	3	170
Muesli_Raisins,_Peaches,_&_Pecans	150	4	3	150	3	16	11	3	170
Mueslix_Crispy_Blend	160	3	2	150	3	17	13	3	160
Bran_Flakes	90	3	0	210	5	13	5	3	190

Findings: Calories(high), fat(low),protein(high),fiber(low),carbo(high),sugars(high), sodium(high),shelf(high),potass(high)

For Cluster-6(14 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Apple_Jacks	110	2	0	125	1	11	14	2	30
Cocoa_Puffs	110	1	1	180	0	12	13	2	55
Count_Chocula	110	1	1	180	0	12	13	2	65
Froot_Loops	110	2	1	125	1	11	13	2	30
Fruity_Pebbles	110	1	1	135	0	13	12	2	25

Lucky_Charms	110	2	1	180	0	12	12	2	55
Trix	110	1	1	140	0	13	12	2	25
Smacks	110	2	1	70	1	9	15	2	40
Cap'n'Crunch	120	1	2	220	0	12	12	2	35
Cinnamon_Toast_ Crunch	120	1	3	210	0	13	9	2	45
Corn_Pops	110	1	0	90	1	13	12	2	20
Honey_Graham_ Ohs	120	1	2	220	1	12	11	2	45
Cinnamon_Toast_ Crunch	120	1	3	210	0	13	9	2	45
Golden_Grahams	110	1	1	280	0	15	9	2	45

Findings: Calories(high), fat(low),protein(low),fiber(low),carbo(moderate),sugars(high), sodium(high) , shelf(moderate),potass(low)

For Cluster-7(1 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates (g)	sugars(g)	display shelf	potassium(mg)
Kix	110	2	1	260	0	21	3	2	40

Findings: Calori(high),protein(low),fat(Low), sodium(high),carbo(low),sugars(low),fiber(0) potass(low)

For Cluster-8(9 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Apple_Cinnamon_Cheerios	110	2	2	180	1.5	10.5	10	1	70
Bran_Chex	90	2	1	200	4	15	6	1	125
Multi-Grain_Cheerios	100	2	1	220	2	15	6	1	90
Wheaties_Honey_Gold	110	2	1	200	1	16	8	1	60
Frosted_Flakes	110	1	0	200	1	14	11	1	25
Honey_Nut_Cherios	110	3	1	250	1.5	11.5	10	1	90
Honey-comb	110	1	0	180	0	14	11	1	35
Wheat_Chex	100	3	1	230	3	17	3	1	115
Wheaties	100	3	1	200	3	17	3	1	110

Findings: Calories(high),protein(medium), Sodium(high) ,fiber(low),fat(low), sugars(medium),potass(low),carbo(medium)

For Cluster -9(4 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Corn_Chex	110	2	0	280	0	22	3	1	25

Corn_Flakes	100	2	0	290	1	21	2	1	35
Rice_Chex	110	1	0	240	0	23	2	1	30
Rice_Krispies	110	2	0	290	0	22	3	1	35

Findings: Calories(high), protein(low),fat(0), potassium(low) , sugars(low),sodium(high), fiber(low),carbo(high)

For Cluster-10(1 instances):

cereal name	calories	protein (g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Raisin_Bran	120	3	1	210	5	14	12	2	240

Findings: Calories(high), fat(low),sodium(medium),fiber(high),carbo(medium),sugars(high), potass(high)

For Cluster-11(2 instances) :

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber (g)	complex carbohydrates(g)	sugars (g)	display shelf	potassium (mg)
Frosted_Mini-Wheats	100	3	0	0	3	14	7	2	100
Strawberry_Fruit_Wheats	90	2	0	15	3	15	5	2	90

Findings: Calories(high), potass(low),fat(low),protein(medium),fiber(low),carbo(medium), sugars(low),potass(low)

For Cluster-12(1 instances):

cereal name	calories	protein (g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates (g)	sugars(g)	display shelf	potassium(mg)
Maypo	100	4	1	0	0	16	3	2	95

Findings: Calories(high),protein(medium), potass(low),fat(low),sodium(0),fiber(0), carbo(medium),sugars(low)

For Cluster-13(1 instances):

cereal name	calories	protein (g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars (g)	display shelf	potassium(mg)
Raisin_Squares	90	2	0	0	2	15	6	3	110

Findings: Sodium(0), calories(low), potass (low) , fat (0),carbo(medium),sugars(low), fiber(low)

For Cluster-14(1 instances) :

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates (g)	sugars(g)	display shelf	potassium(mg)

Life	100	4	2	150	2	12	6	2	95

Findings: Calories(high), fat(low), potass(low), sugar(low),protein(medium),fiber(low),sodium(medium)

For Cluster-15(1 instances):

cereal name	calories	protein (g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Golden_Crisp	100	2	0	45	0	11	15	1	40

Findings: Potass(low), fat(0), calories(high), sugar(high),protein(low),sodium(low), carbo(medium)

For Cluster-16(1 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber (g)	complex carbohydrates(g)	sugars (g)	display shelf	potassium (mg)
Cream_of_Wheat_(Quick)	100	3	0	80	1	21	0	2	-1

Findings: calories(high), fat(0), sodium(low), carbon(high),sugars(0),potassium(low), protein(low)

For Cluster-17(2 instances):

cereal name	calories	protein (g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Puffed_Rice	50	1	0	0	0	13	0	3	15
Puffed_Wheat	50	2	0	0	1	10	0	3	50

Findings: Calories(low),fat(0), protein(low),sodium(0),fiber(low),carbon(medium),sugar(0), potass(low)

For Cluster-18(3 instances) :

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Shredded_Wheat	80	2	0	0	3	16	0	1	95
Shredded_Wheat_'n'Bran	90	3	0	0	4	19	0	1	140
Shredded_Wheat_spoon_size	90	3	0	0	3	20	0	1	120

Findings: Calories(medium), fat(0),sodium(0), sugar(0), protein(low),carbo(high) potass(low),fiber(low)

For Cluster-19(1 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Cheerios	110	6	2	290	2	17	1	1	105

Findings : Calories(high) , protin(high),fat (low) , sodium(high) , carbon(high), potass(low),sugars(low)

For Cluster-20(1 instances):

cereal name	calories	protein(g)	fat(g)	sodium(mg)	dietary fiber(g)	complex carbohydrates(g)	sugars(g)	display shelf	potassium(mg)
Special_K	110	6	0	230	1	16	3	1	55

Findings: Calories(high), sodium(high), fat(low), fiber(low),carbo(medium),sugars(low), potass(low),protein(high)

Questions & Answers

1. Is a strong correlation between dietary fiber and potassium?

Ans:

There is correlation between dietary fiber and potassium .They are linearly proportional. When the value of dietary fiber rises then the value of potassium also rises. The visualized graph is presented below :

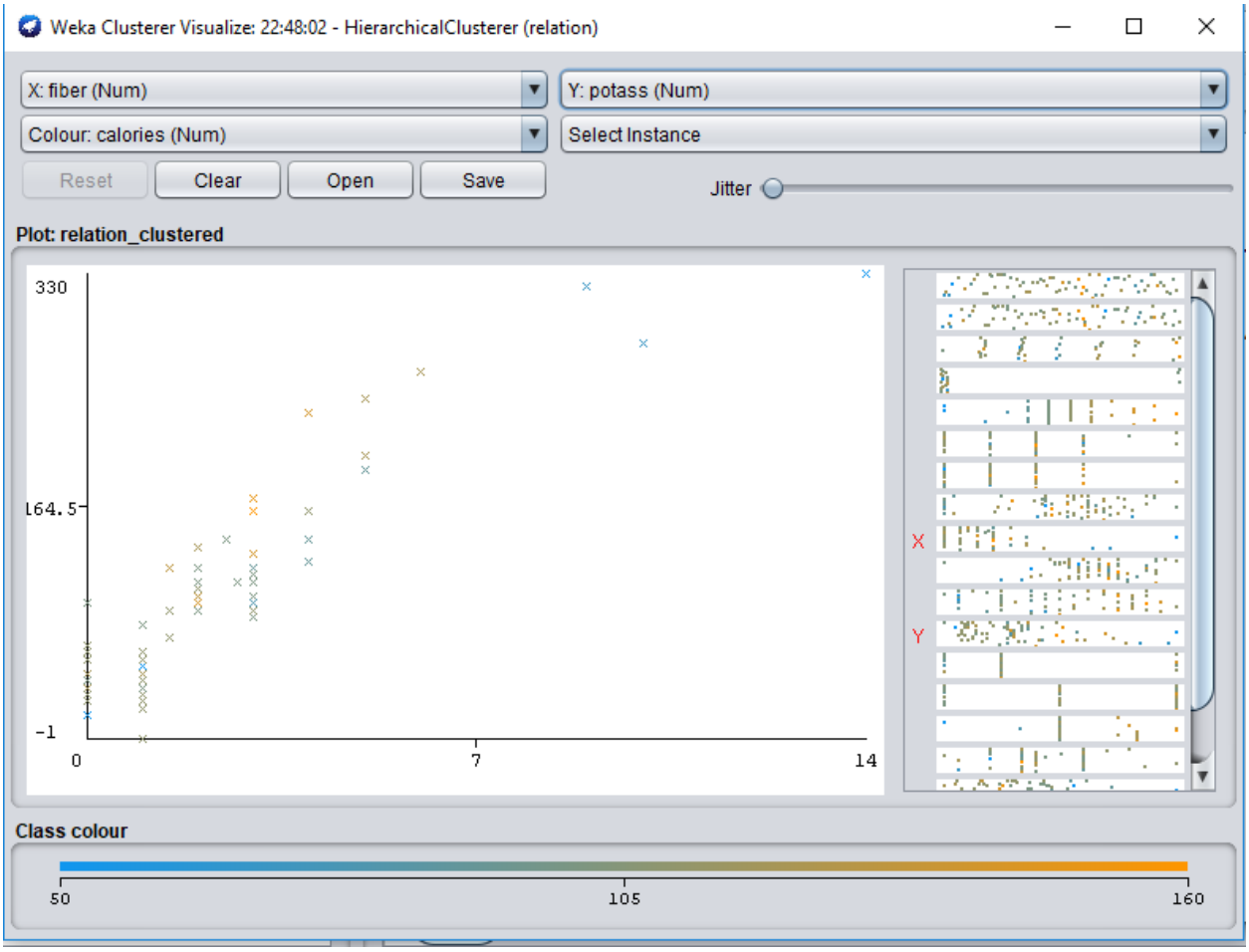

Figure1: Correlation between Dietary fiber(x) and potassium(y)

2. Are groups of cereals from which we can choose according to our preferences?

Yes, they are in groups so one can choose according to their own preferences.

- The diabetes patients can choose cereal from cluster -3 as it contains low fat and low sugar
- Anyone looking for all food values he/she can choose a cereal from cluster two.
- Those who want high sodium, high potassium can pick a cereal from cluster-5 and cluster 6.
- Those who have both blood pressure and diabetics can chose from cluster-14
- Anyone looking for high calories , high protin , high sodium , high carbon can choose cluster-19
- Anyone in need of high protein and potassium can choose from cluster

3. See other correlation between the data given in the files.

Ans:There is correlation between Calories and Carbohydrates. If the value of Calories goes higher then there is a possibility of higher carbohydrate value. This confirms low carbohydrate value for low calories.