**Data Warehousing & Data Mining**

**Without 'vitamin' and 'rating' columns**

**Description:**

In this project we used a dataset of cereals along with their different percentage of calories, protein, fat etc. Besides, we did hierarchical clustering for finding out who needs which type of diet or food.

**Methods:**

* Hierarchical Clustering
* Data Visualization
* Correlation

**Hierarchical Clustering :**

**=== Run information ===**

Scheme: weka.clusterers.HierarchicalClusterer -N 2 -L SINGLE -P –A

"weka.core.EuclideanDistance -R first-last"

Relation: Cereal-weka.filters.unsupervised.attribute.Remove-R1

Instances: 77

Attributes: 9

calories

protein(g)

fat(g)

sodium(mg)

dietary fiber(g)

complex carbohydrates(g)

sugars(g)

display shelf

potassium(mg)

Test mode: evaluate on training data

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

Cluster 0

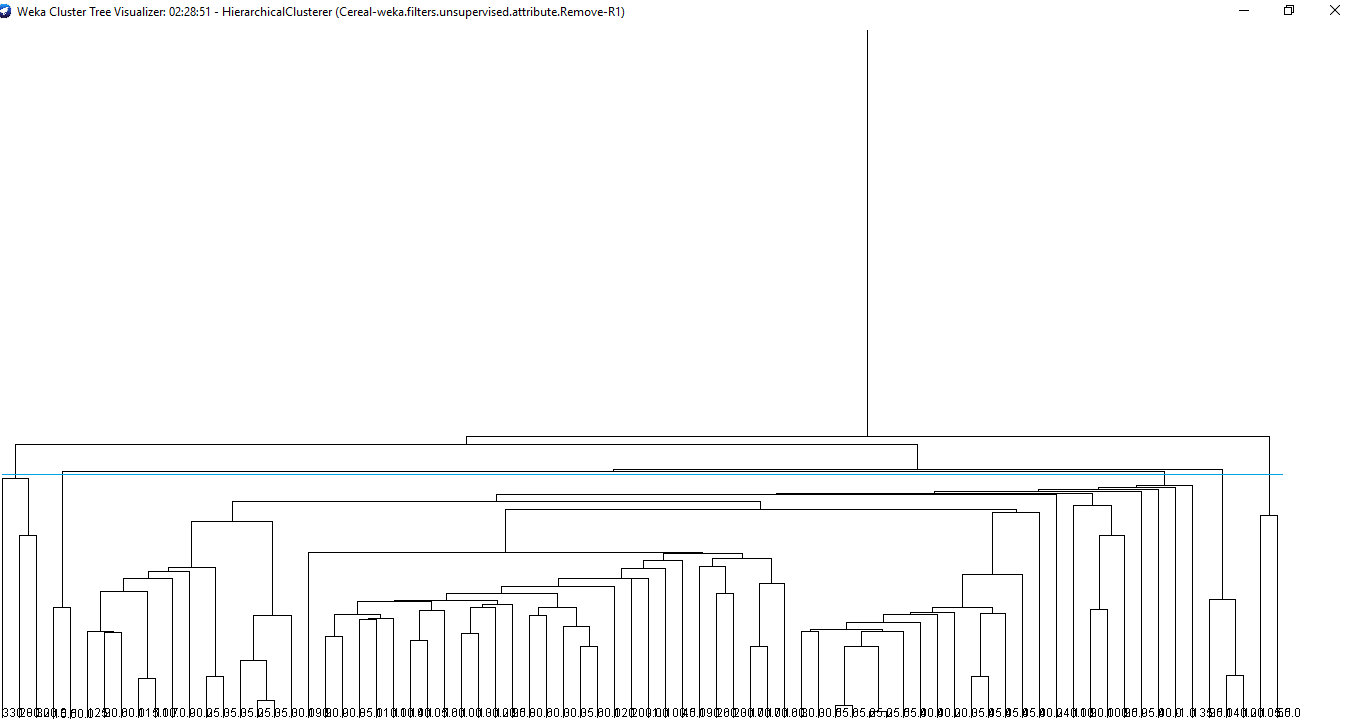
(((330.0:0.57856,(280.0:0.44226,320.0:0.44226):0.1363):0.08163,(((15.0:0.26816,50.0:0.26816):0.32701,(((((((((((((125.0:0.20919,(90.0:0.207,60.0:0.207):0.00219):0.09722,(115.0:0.09496,110.0:0.09496):0.21145):0.03158,70.0:0.33799):0.01512,90.0:0.35311):0.01197,(25.0:0.0996,35.0:0.0996):0.26547):0.10996,((35.0:0.13787,(25.0:0.04347,35.0:0.04347):0.09441):0.11011,30.0:0.24798):0.22706):0.04806,((190.0:0.39914,(((((((((((90.0:0.19672,90.0:0.19672):0.05316,((85.0:0.2381,110.0:0.2381):0.00319,110.0:0.24129):0.00859):0.03223,((140.0:0.18836,105.0:0.18836):0.0719,160.0:0.26026):0.02185):0.00208,(((100.0:0.20516,130.0:0.20516):0.06193,120.0:0.26708):0.00672,95.0:0.2738):0.01039):0.01722,((80.0:0.24792,60.0:0.24792):0.01877,(30.0:0.22189,(35.0:0.17355,60.0:0.17355):0.04835):0.04479):0.03473):0.01636,120.0:0.31778):0.01982,200.0:0.33761):0.00004,1.0:0.33764):0.02293,100.0:0.36057):0.02023,45.0:0.3808):0.01647,((190.0:0.3668,(260.0:0.30072,230.0:0.30072):0.06607):0.01968,((170.0:0.17188,170.0:0.17188):0.15453,160.0:0.3264):0.06007):0.0108):0.00187):0.10546,((((((((30.0:0.20954,30.0:0.20954):0.00491,(((55.0:0.03021,65.0:0.03021):0.14149,(25.0:0.01562,25.0:0.01562):0.15608):0.03784,55.0:0.20954):0.00491):0.01581,40.0:0.23027):0.01888,40.0:0.24915):0.00601,20.0:0.25516):0.01214,((35.0:0.0996,45.0:0.0996):0.15227,45.0:0.25187):0.01542):0.07879,45.0:0.34608):0.15117,40.0:0.49725):0.00735):0.0185):0.01632,240.0:0.53943):0.00315,(110.0:0.51464,((90.0:0.26217,100.0:0.26217):0.18008,95.0:0.44226):0.07239):0.02793):0.00624,95.0:0.54881):0.00309,40.0:0.5519):0.00512,1.0:0.55703):0.0063,135.0:0.56333):0.03183):0.00491,(95.0:0.28591,(140.0:0.10242,120.0:0.10242):0.1835):0.31415):0.06013):0.02044,(105.0:0.49034,55.0:0.49034):0.19029)

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

Clustered Instances

0 76 ( 99%)

1 1 ( 1%)



**Cluster 1:** (3 Instances)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cereal name | Calories | Protein | fat | sodium | fiber | Carbo | sugars | Shelf | potass |
| Bran | 70 | 4 | 1 | 130 | 10 | 5 | 6 | 3 | 280 |
| All-Bran | 70 | 4 | 1 | 260 | 9 | 7 | 5 | 3 | 320 |
| All-Bran\_with\_Extra\_Fiber | 50 | 4 | 0 | 140 | 14 | 8 | 0 | 3 | 330 |

**Findings:** High fiber and potassium.

**Cluster 2:** (67 Instances)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cereal name | calories | protein | fat | sodium | fiber | carbo | sugars | shelf | potass |
| Natural\_Bran | 120 | 3 | 5 | 15 | 2 | 8 | 8 | 3 | 135 |
| Almond\_Delight | 110 | 2 | 2 | 200 | 1 | 14 | 8 | 1 | 25 |
| Apple\_Cinnamon\_ | 110 | 2 | 2 | 180 | 1.5 | 10.5 | 10 | 1 | 70 |
| Apple\_Jacks | 110 | 2 | 0 | 125 | 1 | 11 | 14 | 2 | 30 |
| Basic\_4 | 130 | 3 | 2 | 210 | 2 | 18 | 8 | 3 | 100 |
| Bran\_Chex | 90 | 2 | 1 | 200 | 4 | 15 | 6 | 1 | 125 |
| Bran\_Flakes | 90 | 3 | 0 | 210 | 5 | 13 | 5 | 3 | 190 |
| CapCrunch | 120 | 1 | 2 | 220 | 0 | 12 | 12 | 2 | 35 |
| Cinnamon\_Toast\_Crunch | 120 | 1 | 3 | 210 | 0 | 13 | 9 | 2 | 45 |
| Clusters | 110 | 3 | 2 | 140 | 2 | 13 | 7 | 3 | 105 |
| Cocoa\_Puffs | 110 | 1 | 1 | 180 | 0 | 12 | 13 | 2 | 55 |
| Corn\_Chex | 110 | 2 | 0 | 280 | 0 | 22 | 3 | 1 | 25 |
| Corn\_Flakes | 100 | 2 | 0 | 290 | 1 | 21 | 2 | 1 | 35 |
| Corn\_Pops | 110 | 1 | 0 | 90 | 1 | 13 | 12 | 2 | 20 |
| Count\_Chocula | 110 | 1 | 1 | 180 | 0 | 12 | 13 | 2 | 65 |
| Cracklin\_Oat\_Bran | 110 | 3 | 3 | 140 | 4 | 10 | 7 | 3 | 160 |
| Cream\_of\_Wheat\_(Quick) | 100 | 3 | 0 | 80 | 1 | 21 | 0 | 2 | 0 |
| Crispix | 110 | 2 | 0 | 220 | 1 | 21 | 3 | 3 | 30 |
| Crispy\_Wheat\_&\_Raisins | 100 | 2 | 1 | 140 | 2 | 11 | 10 | 3 | 120 |
| Double\_Chex | 100 | 2 | 0 | 190 | 1 | 18 | 5 | 3 | 80 |
| Froot\_Loops | 110 | 2 | 1 | 125 | 1 | 11 | 13 | 2 | 30 |
| Frosted\_Flakes | 110 | 1 | 0 | 200 | 1 | 14 | 11 | 1 | 25 |
| Frosted\_Mini-Wheats | 100 | 3 | 0 | 0 | 3 | 14 | 7 | 2 | 100 |
| Fruit\_&\_Fibre\_Dates, | 120 | 3 | 2 | 160 | 5 | 12 | 10 | 3 | 200 |
| Fruitful\_Bran | 120 | 3 | 0 | 240 | 5 | 14 | 12 | 3 | 190 |
| Fruity\_Pebbles | 110 | 1 | 1 | 135 | 0 | 13 | 12 | 2 | 25 |
| Golden\_Crisp | 100 | 2 | 0 | 45 | 0 | 11 | 15 | 1 | 40 |
| Golden\_Grahams | 110 | 1 | 1 | 280 | 0 | 15 | 9 | 2 | 45 |
| Grape\_Nuts\_Flakes | 100 | 3 | 1 | 140 | 3 | 15 | 5 | 3 | 85 |
| Grape-Nuts | 110 | 3 | 0 | 170 | 3 | 17 | 3 | 3 | 90 |
| Great\_Grains\_Pecan | 120 | 3 | 3 | 75 | 3 | 13 | 4 | 3 | 100 |
| Honey\_Graham\_Ohs | 120 | 1 | 2 | 220 | 1 | 12 | 11 | 2 | 45 |
| Honey\_Nut\_Cheerios | 110 | 3 | 1 | 250 | 1.5 | 11.5 | 10 | 1 | 90 |
| Honey-comb | 110 | 1 | 0 | 180 | 0 | 14 | 11 | 1 | 35 |
| Just\_Right\_Crunchy | 110 | 2 | 1 | 170 | 1 | 17 | 6 | 3 | 60 |
| Just\_Right\_Fruit\_&\_Nut | 140 | 3 | 1 | 170 | 2 | 20 | 9 | 3 | 95 |
| Kix | 110 | 2 | 1 | 260 | 0 | 21 | 3 | 2 | 40 |
| Life | 100 | 4 | 2 | 150 | 2 | 12 | 6 | 2 | 95 |
| Lucky\_Charms | 110 | 2 | 1 | 180 | 0 | 12 | 12 | 2 | 55 |
| Maypo | 100 | 4 | 1 | 0 | 0 | 16 | 3 | 2 | 95 |
| Muesli\_Raisins,\_Dates | 150 | 4 | 3 | 95 | 3 | 16 | 11 | 3 | 170 |
| Muesli\_Raisins,\_Peaches | 150 | 4 | 3 | 150 | 3 | 16 | 11 | 3 | 170 |
| Mueslix\_Crispy\_Blend | 160 | 3 | 2 | 150 | 3 | 17 | 13 | 3 | 160 |
| Multi-Grain\_Cheerios | 100 | 2 | 1 | 220 | 2 | 15 | 6 | 1 | 90 |
| Nut&Honey\_Crunch | 120 | 2 | 1 | 190 | 0 | 15 | 9 | 2 | 40 |
| Nutri-Grain\_Almond-Raisin | 140 | 3 | 2 | 220 | 3 | 21 | 7 | 3 | 130 |
| Nutri-grain\_Wheat | 90 | 3 | 0 | 170 | 3 | 18 | 2 | 3 | 90 |
| Oatmeal\_Raisin\_Crisp | 130 | 3 | 2 | 170 | 1.5 | 13.5 | 10 | 3 | 120 |
| Post\_Nat.\_Raisin\_Bran | 120 | 3 | 1 | 200 | 6 | 11 | 14 | 3 | 260 |
| Product\_19 | 100 | 3 | 0 | 320 | 1 | 20 | 3 | 3 | 45 |
| Quaker\_Oat\_Squares | 100 | 4 | 1 | 135 | 2 | 14 | 6 | 3 | 110 |
| Quaker\_Oatmeal | 100 | 5 | 2 | 0 | 2.7 | -1 | -1 | 1 | 110 |
| Raisin\_Bran | 120 | 3 | 1 | 210 | 5 | 14 | 12 | 2 | 240 |
| Raisin\_Nut\_Bran | 100 | 3 | 2 | 140 | 2.5 | 10.5 | 8 | 3 | 140 |
| Raisin\_Squares | 90 | 2 | 0 | 0 | 2 | 15 | 6 | 3 | 110 |
| Rice\_Chex | 110 | 1 | 0 | 240 | 0 | 23 | 2 | 1 | 30 |
| Rice\_Krispies | 110 | 2 | 0 | 290 | 0 | 22 | 3 | 1 | 35 |
| Smacks | 110 | 2 | 1 | 70 | 1 | 9 | 15 | 2 | 40 |
| Strawberry\_Fruit\_Wheats | 90 | 2 | 0 | 15 | 3 | 15 | 5 | 2 | 90 |
| Total\_Corn\_Flakes | 110 | 2 | 1 | 200 | 0 | 21 | 3 | 3 | 35 |
| Total\_Raisin\_Bran | 140 | 3 | 1 | 190 | 4 | 15 | 14 | 3 | 230 |
| Total\_Whole\_Grain | 100 | 3 | 1 | 200 | 3 | 16 | 3 | 3 | 110 |
| Triples | 110 | 2 | 1 | 250 | 0 | 21 | 3 | 3 | 60 |
| Trix | 110 | 1 | 1 | 140 | 0 | 13 | 12 | 2 | 25 |
| Wheat\_Chex | 100 | 3 | 1 | 230 | 3 | 17 | 3 | 1 | 115 |
| Wheaties | 100 | 3 | 1 | 200 | 3 | 17 | 3 | 1 | 110 |
| Wheaties\_Honey\_Gold | 110 | 2 | 1 | 200 | 1 | 16 | 8 | 1 | 60 |

**Findings:** High Calories.

**Cluster 3:** (2 Instances)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cereal name | calories | protein | Fat | sodium | fiber | Carbo | sugars | shelf | potass |
| Puffed\_Rice | 50 | 1 | 0 | 0 | 0 | 13 | 0 | 3 | 15 |
| Puffed\_Wheat | 50 | 2 | 0 | 0 | 1 | 10 | 0 | 3 | 50 |

**Findings:** Low calories, low protein, low fat, low sodium, low fiber, low sugars, low potassium.

**Cluster 4:** (3 Instances)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cereal name | calories | protein | fat | sodium | fiber | carbo | sugars | shelf | potass |
| Shredded\_Wh | 80 | 2 | 0 | 0 | 3 | 16 | 0 | 1 | 95 |
| Shredded\_Whe | 90 | 3 | 0 | 0 | 4 | 19 | 0 | 1 | 140 |
| Shredded\_Whe | 90 | 3 | 0 | 0 | 3 | 20 | 0 | 1 | 120 |

**Findings:** low fat, low sodium, low sugars.

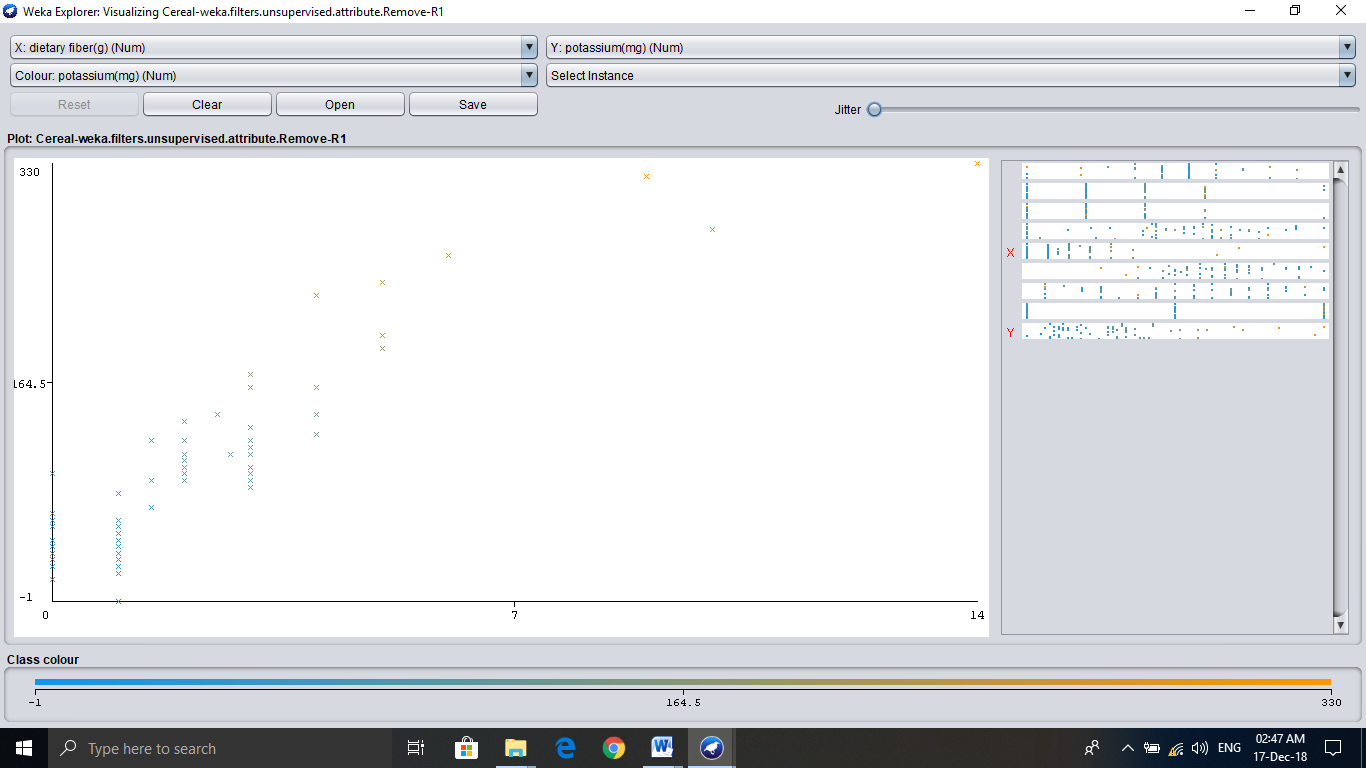
**Cluster 5:** (2 Instances)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cereal name | calories | protein | fat | sodium | fiber | carbo | sugars | Shelf | potass |
| Cheerios | 110 | 6 | 2 | 290 | 2 | 17 | 1 | 1 | 105 |
| Special\_K | 110 | 6 | 0 | 230 | 1 | 16 | 3 | 1 | 55 |

**Findings:** High Calories and high protein.

**Questions Answer:**

1. Yes there is correlation between dietary fiber and potassium and they are linearly proportional. If the value of dietary fiber rises then the value of potassium also rises.



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

* Those High Sugar then he/she can choose cluster **1**.
* Those High Sugar and High Presser then he/she can choose cluster **3**.
* Those who want high sodium and potassium and dietary fiber can pick a cereal from cluster **5**.
* Those who have low pressure can choose cereal from cluster **1, 2, 5**. Here three different clusters can be chosen

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

