**UNSUPERVISED LEARNING**

**Project Definition:** This dataset contains 77 cereal item where each cereal contains different amount of nutrition attributes .It is possible to suggest any kind of nutrition combination for a specific person according to age or health condition .There are some attributes which can be used to determine that which food can be taken and which food needs to avoid.

**Literature Survey:** The total number of calories a person needs each day varies depending on a number of factors, including the person’s age, sex, height, weight, and level of physical activity. In addition, a need to lose, maintain, or gain weight and other factors affect how many calories should be consumed. For children and adolescents, reference height and weight vary. For adults, the reference man is 5 feet 10 inches tall and weighs 154 pounds. The reference woman is 5 feet 4 inches tall and weighs 126 pounds. Estimates range from 1,600 to 2,400 calories per day for adult women and 2,000 to 3,000 calories per day for adult men. Within each age and sex category, the low end of the range is for sedentary individuals; the high end of the range is for active individuals. So it is very important to know which food is to take to keep in good health.

**Methods:** (a)HierarchicalCluster

(b) Correlation

**Without “Vitamin” and “Ratings” Column**

**Information of Hierarchical cluster tree:**

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

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

Relation: nutrionWithoutVitamin

Instances: 77

Attributes: 10

id

calories

protein(g)

fat(g)

sodium(mg)

dietaryfiber(g)

complexcarbohydrates(g)

sugars(g)

displayshelf

potassium(mg)

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

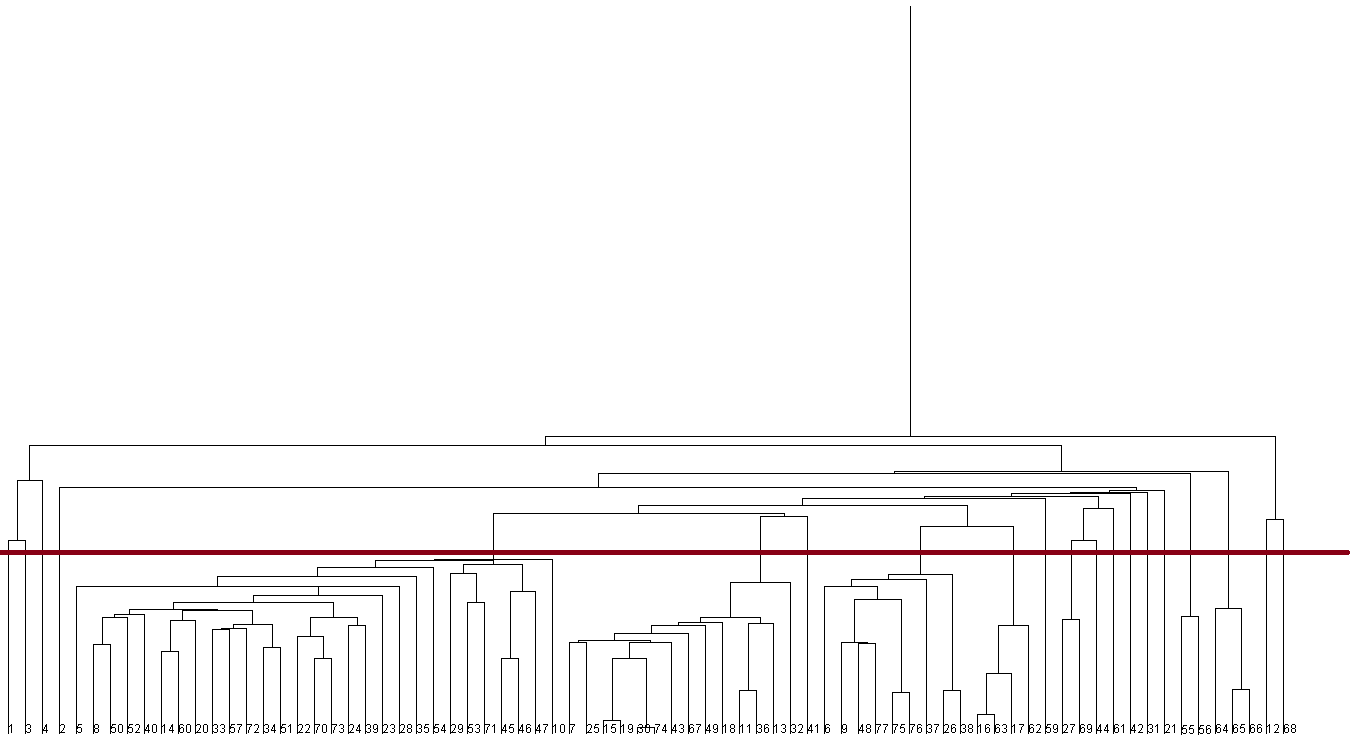
=== Model and evaluation on training set ===

Clustered Instances

0 76 ( 99%)

1. 1 ( 1%)

**Cutting point of Hierarchical Cluster tree:**

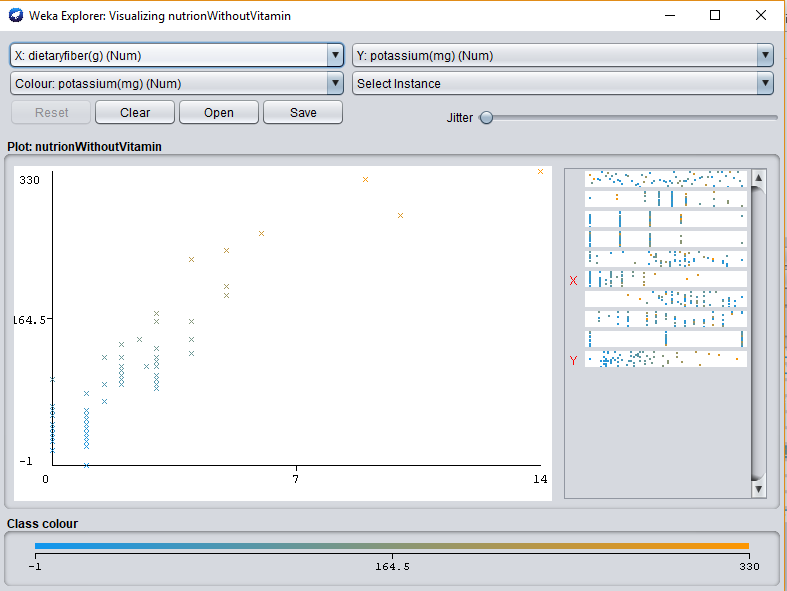


**Cluster Analysis:**

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster no | ID no | Total no of ID | Cluster analysis |
| 01 | 1 | 1 | High rate of potassium and low rate of fat. |
| 02 | 3 | 1 | High rate of potassium and low rate of fat. |
| 03 | 4 | 1 | Zero Fat and sugar |
| 04 | 2 | 29 | High rate of potassium and low rate of dietary fiber |
| 05 | 5,8,50,52,40,14,60,20,33  57,72,34,51,22,70,73,24,39  23,28,35,54,29,53,71,45,46  47,10 | 29 | Low rate of protein and dietary fiber |
| 06 | 7,25,15,19,30,74,43,67,49,  18,11,36,13,32 | 14 | Low rate of protein, fat, dietary fiber and display shelf |
| 07 | 41 | 1 | High rate of sodium |
| 08 | 6,9,48,77,75,76,37,26,38 | 9 | High rate of sodium, potassium |
| 09 | 16,63,17,62 | 4 | Low rate of fat and high rate of sodium |
| 10 | 59 | 1 | Low rate of fat and high rate of potassium |
| 11 | 27,69,44 | 3 | Zero fat |
| 12 | 61 | 1 | High rate of calorie and potassium |
| 13 | 42 | 1 | Low rate of fat, dietary fiber and sugar |
| 14 | 31 | 1 | Zero fat |
| 15 | 21 | 1 | Zero fat |
| 16 | 55,56 | 2 | Low rate of fat and high rate of sodium |
| 17 | 64,65,66 | 3 | Zero fat and sodium |
| 18 | 12 | 1 | High rate of sodium |
| 19 | 68 | 1 | Zero fat and high rate of sodium |

**Question answer:**

1. **Question:** Is there any strong correlation between dietary fiber and potassium?



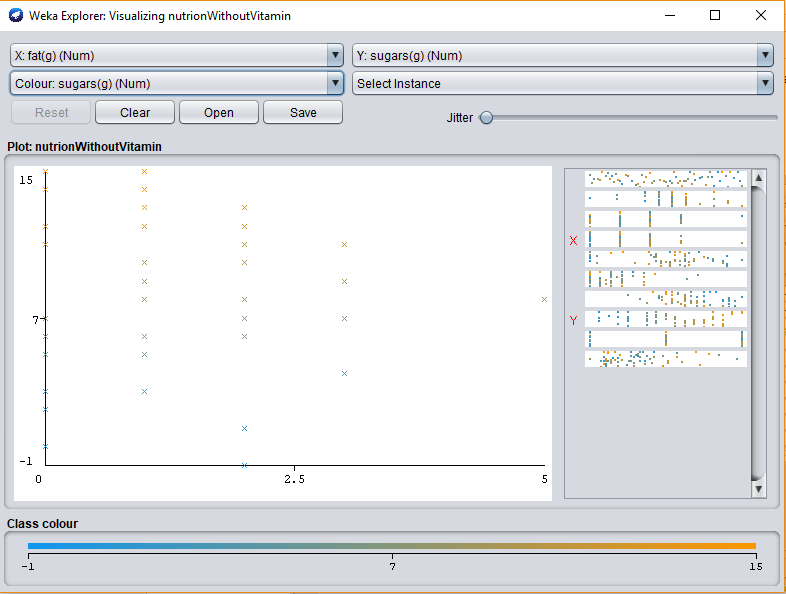
**Answer:** There is a strong linear correlation between dietary fiber and potassium. The relation is proportional.

**(B) Question:** Are groups of cereals from which we can choose according to our preferences?

**Answer:** Yes if we consider the clusters then we see we can choose cereals according to our preferences. For example if anyone wants zero fat cereals then he/she can choose any cereals from 3, 11, 14,15,17,19 cluster.

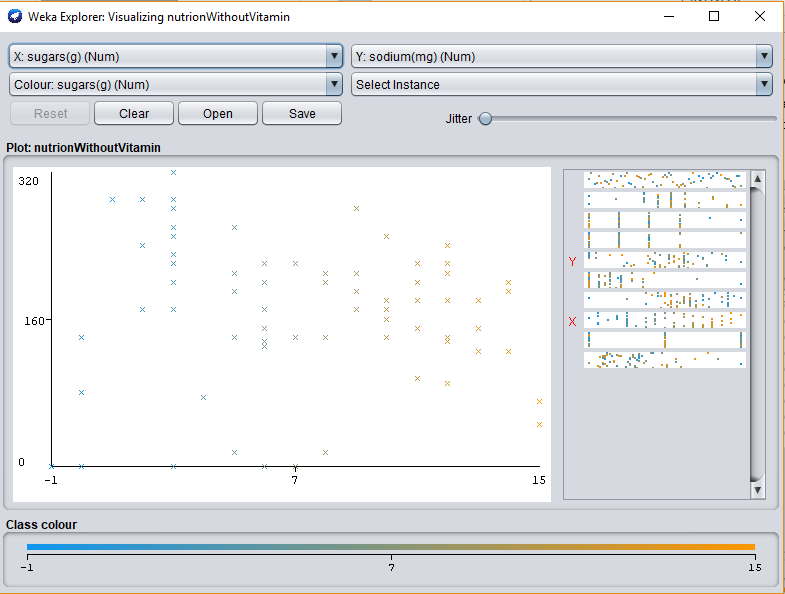
**Some other correlation:**

**(C) Fat VS Sugar:**



There is a strong correlation between fat and sugar. The relation is proportional.

**Sodium VS Sugar:**



The relation is inversely proportional. So if sodium increases sugar decreases and if sodium decreases sugar increases.

**(D) Other information finding:**(a) Cluster number 3, 11, 14,15,17,19 doesn’t contain any amount of fat.

(b) High rate of potassium and low rate of fat and sugar stays together and the relation is inversely proportional.

**With “Vitamin” and “Ratings” Column**

**Information of Hierarchical cluster tree:**

=== Run information ===

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

Relation: cerealWithVitamin

Instances: 75

Attributes: 12

cerealName

cal

protein

fat

sodium

fiber

carbo

suger

potassium

vitamin

shelf

rating

Test mode: evaluate on training data

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

Cluster 0

(((((1:0.45956,3:0.45956):0.21047,4:0.67003):0.00706,(((2:0.63675,(((((((((5:0.34946,((8:0.24314,(46:0.21537,75:0.21537):0.02777):0.07118,(73:0.09803,74:0.09803):0.2163):0.03513):0.01125,35:0.36071):0.0113,(24:0.10622,36:0.10622):0.26578):0.11218,(((15:0.04606,61:0.04606):0.10861,16:0.15467):0.09347,60:0.24814):0.23605):0.04578,((((((((6:0.20993,23:0.20993):0.01771,(((14:0.03206,18:0.03206):0.15431,(28:0.01603,72:0.01603):0.17034):0.02973,41:0.2161):0.01153):0.0032,65:0.23084):0.02228,47:0.25311):0.02191,17:0.27503):0.00006,((10:0.11214,34:0.11214):0.14118,12:0.25332):0.02177):0.07193,30:0.34702):0.17681,((((((((((((7:0.21278,48:0.21278):0.06881,50:0.28159):0.03189,((13:0.19164,58:0.19164):0.06953,19:0.26118):0.05231):0.00829,21:0.32177):0.00038,((31:0.24174,55:0.24174):0.0414,(32:0.21347,49:0.21347):0.06966):0.03902):0.00223,((20:0.27122,71:0.27122):0.002,22:0.27322):0.05116):0.01553,26:0.33991):0.02772,33:0.36763):0.03526,((43:0.17638,44:0.17638):0.15406,45:0.33044):0.07245):0.00072,(27:0.37831,51:0.37831):0.02529):0.05821,9:0.46182):0.04317,39:0.50499):0.01885):0.00614):0.01013,57:0.5401):0.00472,(((25:0.2627,67:0.2627):0.18199,42:0.44469):0.07307,59:0.51776):0.02706):0.00696,40:0.55179):0.0068,29:0.55859):0.07816):0.01932,(53:0.27195,54:0.27195):0.38412):0.0045,(62:0.29323,(63:0.10657,64:0.10657):0.18666):0.36734):0.01652):0.00506,(11:0.49375,66:0.49375):0.18839):0.08807,(((((37:0.28993,68:0.28993):0.10692,70:0.39685):0.03047,38:0.42733):0.0605,52:0.48783):0.11208,69:0.59991):0.1703)

Time taken to build model (full training data) : 0.01 seconds

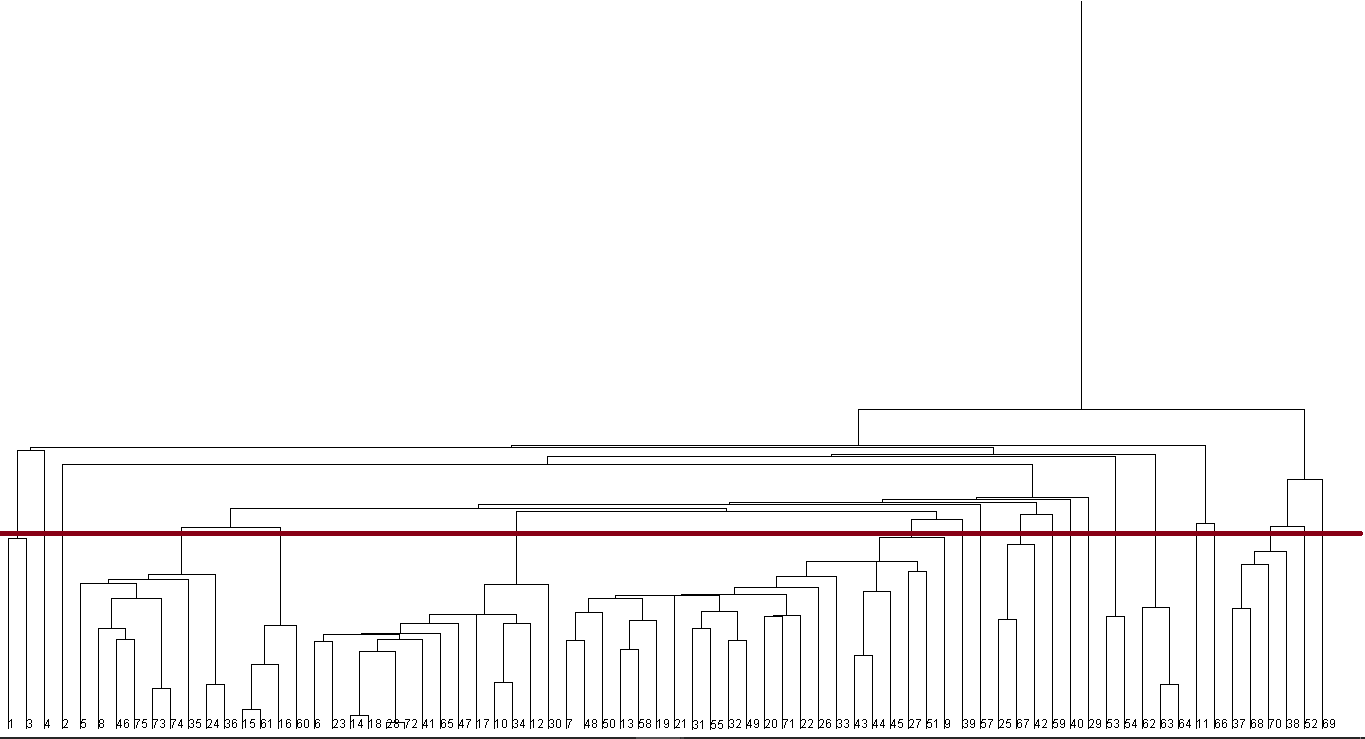
=== Model and evaluation on training set ===

Clustered Instances

0 74 ( 99%)

1. 1 ( 1%)

**Cutting point of Hierarchical cluster tree:**



|  |  |  |  |
| --- | --- | --- | --- |
| Cluster no | ID no | Total no of ID | Cluster analysis |
| 1 | 1,3 | 2 | Low rate of fat and shelf |
| 2 | 4 | 1 | Zero fat and zero sugar |
| 3 | 2 | 1 | Low rate of vitamin and high rate of calorie |
| 4 | 5,8,46,75,73,  74,35,24,36 | 9 | Low rate of shelf and high rate of sodium |
| 5 | 15,61,16,60 | 4 | High rate of calorie and low rate of fat and fiber |
| 6 | 6,23,14,18,28,72,41  65,47,17,10,34,12,30 | 14 | Low rate of fat and shelf |
| 7 | 7,48,50,13,58,19,21,  31,55,32,49,20,71,22,  26,33,43,44,45,27,51 | 21 | High rate of sodium |
| 8 | 9 | 1 | Zero fat and high sodium |
| 9 | 39 | 1 | Zero fiber and high calorie |
| 10 | 57 | 1 | High rate of sodium and potassium |
| 11 | 25,67,42 | 3 | Low rate of sodium |
| 12 | 59 | 1 | Zero fat |
| 13 | 40 | 1 | High rate of sodium and low fiber |
| 14 | 29 | 1 | Medium vitamin |
| 15 | 53,54 | 3 | Zero fat and sodium |
| 16 | 62,63,64 | 3 | Zero fat,protein,sugar and vitamin |
| 17 | 11 | 1 | High rate of sodium |
| 18 | 66 | 1 | Low rate of sugar |
| 19 | 37,68,70,38 | 4 | High rate of vitamin |
| 20 | 52 | 1 | Zero fat and high sodium |
| 21 | 69 | 1 | High rate of sodium and potassium |

**Question:** Are groups of cereals from which we can choose according to our preferences?

**Answer:** Yes if we consider the clusters then we see we can choose cereals according to our preferences. For example if anyone wants zero fat cereals then he/she can choose any cereals from 2, 8, 9,15,16,20 cluster.

**Other information finding:**

(a) The relation between fat, sugar, and shelf is proportional.

(b) Most of the clusters have medium rate of vitamin.