Yannick Lokpe

MIS 64060: Fundamentals of Machine Learning

Professor: Murali Shanker

Objectives:

In this study, our mission is to find the best five cereals among the dataset using the guidelines outlined in the US Governments Dietary Guidelines'Executive Summary:

* Follow a healthy eating pattern across the lifespan.
* Focus on variety, nutrient density, and amount.
* Limit calories from added sugars and saturated fats and reduce sodium intake.
* Shift to healthier food and beverage choices.
* Support healthy eating patterns for all.

Approach:

In this study, I used the normalization in Machine learning by Min-MaxScaling.

Subtract the minimum value from each column's highest value and divide by the range.

Each new column has a minimum value of 0 and a maximum of 1.

Results:

After normalizing the dataset by weight, we discovered that there doesn't seem to be a strong selection for our 'Good\_variables' (best cereals). , the values with the best protein do not have the highest vitamin percentage of potassium. The higher values of fiber do not have a strong presence of vitamins. The 'best' cereals do not make a standout.

Therefore, we considered using a different approach by focusing on the limits of the guidelines as 'Bad\_variables.' There were no five cereals that satisfied the guidelines.

So there are no significant, healthy cereals that stand above the others. Instead, let's select the best ones we can. We define a metricby giving a score for each sample calculated from its rank in each variable column. Therefore we use the metric to identify a higher score that is healthier. We were able to obtain the five best cereals based on the US Governments Dietary Guidelines'Executive.

Conclusion:

 We can see that our metric score works well because the bottom scores are sugary, clearly unhealthy cereals. And as the metric score decreases, the cereals generally become more unhealthy (increased sugars, fats, reduced protein, etc.).

Another benefit we get from using a metric score to rank our cereals is plotting the distribution. We can see that metric scores tend to increase towards lower 'bad' variables like sugar and higher 'good' variables like fiber and protein. So the metric score gives us an indicator of how healthy a cereal is concerning all variables.

Source:

DIETARY GUIDELINES 2015-2020 EIGHTH EDITION FOR AMERICANS:

<https://health.gov/sites/default/files/2019-10/DGA_Executive-Summary.pdf>

Cereals.csv:

https://kent.instructure.com/courses/10017/files/1631550?wrap=1

Good Variables

The values with the best protein do not have the highest vitamin percentage or potassium, and the higher values of fiber do not have a strong presence of vitamins. Therefore the best cereals do not stand out clearly.

Chart, scatter chart

Description automatically generated

Bad Variables

there are no significant, healthy cereals that stand above the others. Instead, let's select the best ones we can. We can define a metric by giving a score for each sample calculated from its rank in each variable column. To ensure good scoring, instead of using the position of each sample in a sorted column for a score value, we can bin the column value ranges and give each sample a value in that column based on which bin it falls into.

Chart, scatter chart

Description automatically generated

Best Cereals  
 name mfr type calories protein fat sodium fiber carbo sugars potass vitamins sugars\_criteria rank

3 All-Bran\_with\_Extra\_Fiber K C 50.0 4.0 0.0 140.0 14.0 8.0 0.0 330.0 5 1 47

0 100%\_Bran N C 70.0 4.0 9.0 130.0 10.0 5.0 24.0 280.0 5 0 40

2 All-Bran K C 70.0 4.0 9.0 260.0 9.0 7.0 20.0 320.0 5 0 39

64 Shredded\_Wheat\_'n'Bran N C 90.0 3.0 0.0 0.0 4.0 19.0 0.0 140.0 1 1 36

26 Frosted\_Mini-Wheats K C 100.0 3.0 0.0 0.0 3.0 14.0 28.0 100.0 5 0 35

43 Maypo A H 100.0 4.0 9.0 0.0 0.0 16.0 12.0 95.0 5 0 35

71 Total\_Whole\_Grain G C 100.0 3.0 9.0 200.0 3.0 16.0 12.0 110.0 8 0 35

55 Puffed\_Wheat Q C 100.0 4.0 0.0 0.0 2.0 20.0 0.0 100.0 1 1 35

Metric method

We are using a metric score to rank our cereals plots the distribution. We can see that metric scores tend to increase towards lower 'bad' variables like sugar and higher 'good' variables like fiber and protein. Therefore, the metric score gives us an indicator of how healthy a cereal is concerning all variables.

Chart, scatter chart

Description automatically generated