

Solution Guide

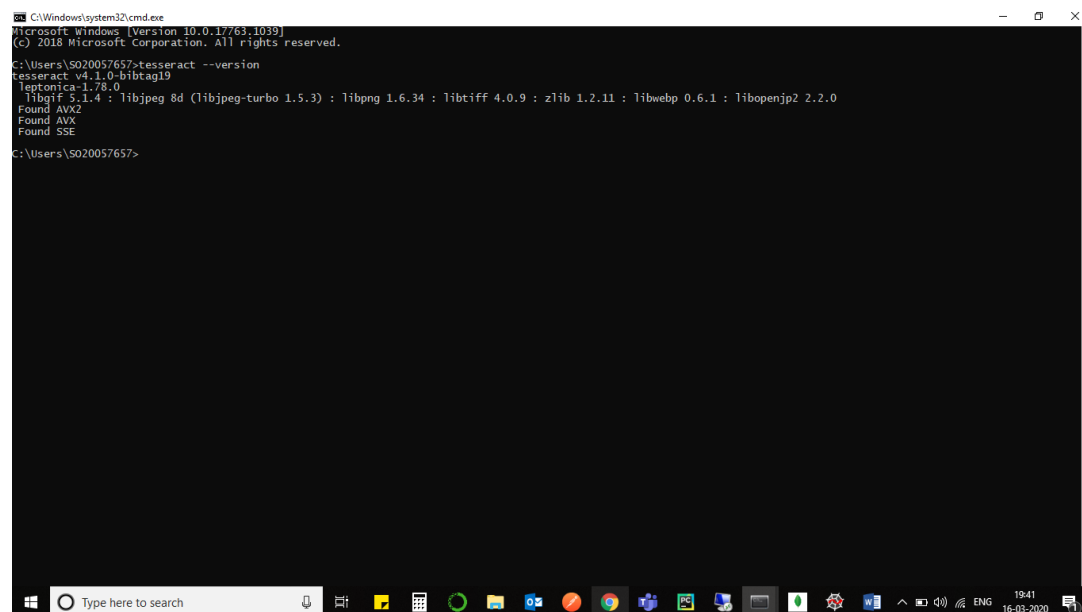
I believe you **have already deployed the solution by going through the steps present in readme – deployment doc**. If not please go through it first make the solution ready and then use this document to run and use the solution in different modes.

Once done please go through the readme-solution intuition file to understand the solution and its components so that you can much better understand the different modes this solution can work on.

I have attached screenshots of the solution working so that one can refer and see the same things when running in their local.

Before diving make sure you also have a tesseract version >4 install in your system and present in your path.

Check this by typing tesseract –version in your command prompt and see the below results.



```

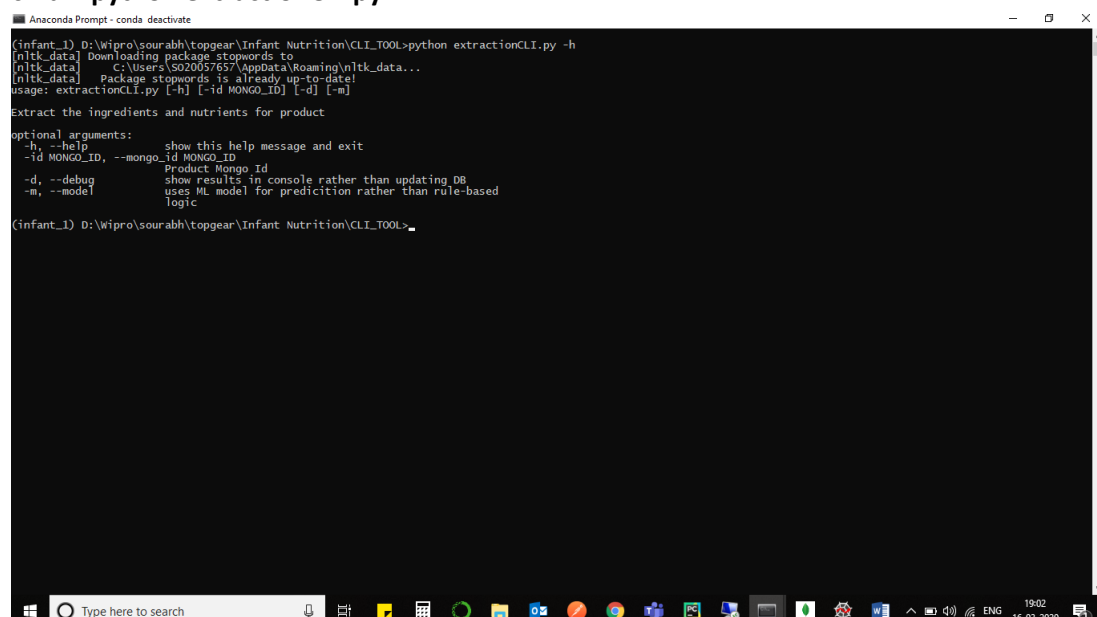
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 10.0.17763.1039]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\S020057657>tesseract --version
tesseract v4.1.0-bibtag19
  leptonica-1.78.0
  libgif 5.1.4 : libjpeg 8d (libjpeg-turbo 1.5.3) : libpng 1.6.34 : libtiff 4.0.9 : zlib 1.2.11 : libwebp 0.6.1 : libopenjp2 2.2.0
Found AVX2
Found AVX
Found SSE

C:\Users\S020057657>
```

1. Running the solution and seeing the arguments the CLI can take.

cmd – python extractionCLI.py –h



```

Anaconda Prompt - conda deactivate

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -h
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\S020057657\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
usage: extractionCLI.py [-h] [-id MONGO_ID] [-d] [-m]

Extract the ingredients and nutrients for product

optional arguments:
  -h, --help            show this help message and exit
  -id MONGO_ID, --mongo_id MONGO_ID
                        Product Mongo Id
  -d, --debug           show results in console rather than updating DB
  -m, --model           uses ML model for prediction rather than rule-based
                        logic

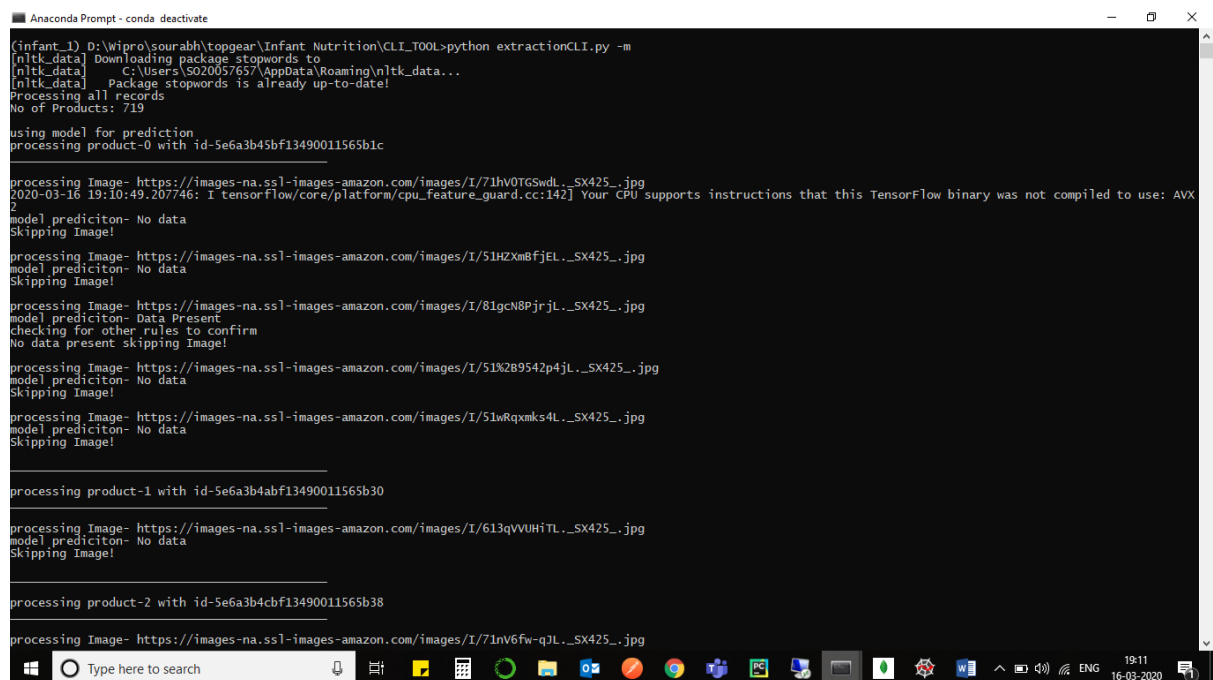
(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>
```

2. Understanding the various arguments.

- **-id** : It takes a string in quotes which is the mongo DB “_id “ of the product you want to run alone.
- **-m** : This when given, uses the trained Binary classifier to decide whether an image contains nutrients and ingredients information or not. When not given the solution uses the rule based coding to decide whether to consider this image or not.
- **-d** : This I added for the reviewer so that when given the ingredients and nutrients extracted could be seen in the standard output itself rather than the DB. When not given the solution updates the extracted nutrients and ingredients in the DB itself. Use when you want to see the working of the solution and not to run and save results.

3. Running the solution on all records and updating the DB with using model.

cmd – **python extractionCLI.py -m**



```
(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -m
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\SQ20057657\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
Processing all records
No of Products: 719

using model for prediction
processing product-0 with id-5e6a3b45bf13490011565b1c

processing Image- https://images-na.ssl-images-amazon.com/images/I/71hV0TGSwDL._SX425_.jpg
2020-03-16 19:10:49.207746: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51HZXmBfjEL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/81gcN8PjrjL._SX425_.jpg
model prediction- Data Present
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51X2B9542p4jL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51wRqxmks4L._SX425_.jpg
model prediction- No data
Skipping Image!

processing product-1 with id-5e6a3b4abf13490011565b30

processing Image- https://images-na.ssl-images-amazon.com/images/I/613qVUHiTL._SX425_.jpg
model prediction- No data
Skipping Image!

processing product-2 with id-5e6a3b4cbf13490011565b38

processing Image- https://images-na.ssl-images-amazon.com/images/I/71nV6fw-qJL._SX425_.jpg
```

4. Running the solution on all records and using debug mode without using model.

cmd - **python extractionCLI.py -d**

```
Anaconda Prompt - conda deactivate

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -d
(nltk_data) Downloading package stopwords to
(nltk_data) C:\Users\S020057657\AppData\Roaming\nltk_data...
(nltk_data) Package stopwords is already up-to-date!
Processing all records
No of Products: 719

using rule-based logic for prediction
processing product-0 with id-5e6a3b45bf13490011565b1c

processing Image- https://images-na.ssl-images-amazon.com/images/I/71hVOTGSwdL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51HZXmBfjEL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/81gcN8PjrjL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51%2B9542p4jL._SX425_.jpg
only ingredients present
Extracting Ingredients!
Extraction Complete!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51wRqxmks4L._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

Extracted Items-
Ingredients
[{'name': 'crowing', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'baby', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'in', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'd', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'safe', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'natural', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'subscribe', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'ingredients', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'save', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'option', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'award', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'winning', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'money', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'back', 'amount': 0, 'unit': 0, 'referenceValue': 0},
{'name': 'guarantee', 'amount': 0, 'unit': 0, 'referenceValue': 0}]
```

5. Running the solution on all records and using debug mode without using model.

cmd - python extractionCLI.py -d -m

```
Anaconda Prompt - conda deactivate

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -d -m
(nltk_data) Downloading package stopwords to
(nltk_data) C:\Users\S020057657\AppData\Roaming\nltk_data...
(nltk_data) Package stopwords is already up-to-date!
Processing all records
No of Products: 719

using model for prediction
processing product-0 with id-5e6a3b45bf13490011565b1c

processing Image- https://images-na.ssl-images-amazon.com/images/I/71hVOTGSwdL._SX425_.jpg
2020-03-16 19:15:11.312472: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX2
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51HZXmBfjEL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/81gcN8PjrjL._SX425_.jpg
model prediction- Data Present
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51%2B9542p4jL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51wRqxmks4L._SX425_.jpg
model prediction- No data
Skipping Image!

processing product-1 with id-5e6a3b4abf13490011565b30

processing Image- https://images-na.ssl-images-amazon.com/images/I/613qVVUHiTL._SX425_.jpg
model prediction- No data
Skipping Image!

processing product-2 with id-5e6a3b4cbf13490011565b38

processing Image- https://images-na.ssl-images-amazon.com/images/I/71nV6fw-qJL._SX425_.jpg
```

6. Running the solution on a single record without using the model.

cmd - python extractionCLI.py -id "5e6a3b6bbf13490011565be8"

```
Anaconda Prompt - conda deactivate

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -id "5e6a3b6bbf13490011565be8"
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\S020057657\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Processing 1 record with Id-5e6a3b6bbf13490011565be8
No of Products: 1

using rule-based logic for prediction
processing product-0 with id-5e6a3b6bbf13490011565be8

processing Image- https://images-na.ssl-images-amazon.com/images/I/71ey2Rxhd-L._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71XMI8mC6FL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/81pkvIMJ-oL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71e0Tt4mzSL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71JjBoz9zdl._SX425_.jpg
both nutrients and ingredients present
Extracting Nutrients!
Extracting Ingredients!
Extraction Complete!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71MPQcYGj6L._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51r0EEJZoEL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

updating DB

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>
```

7. Running the solution on a single record using the model.

cmd - `python extractionCLI.py -id "5e6a3b6bbf13490011565be8" -m`

```
Anaconda Prompt - conda deactivate

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -id "5e6a3b6bbf13490011565be8" -m
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\S020057657\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Processing 1 record with Id-5e6a3b6bbf13490011565be8
No of Products: 1

using model for prediction
processing product-0 with id-5e6a3b6bbf13490011565be8

processing Image- https://images-na.ssl-images-amazon.com/images/I/71ey2Rxhd-L._SX425_.jpg
2020-03-16 19:23:42.040497: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX
model predicton- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71XMI8mC6FL._SX425_.jpg
model predicton- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/81pkvIMJ-oL._SX425_.jpg
model predicton- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71e0Tt4mzSL._SX425_.jpg
model predicton- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71JjBoz9zdl._SX425_.jpg
model predicton- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71MPQcYGj6L._SX425_.jpg
model predicton- Data Present
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51r0EEJZoEL._SX425_.jpg
model predicton- No data
Skipping Image!

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>
```

8. Running the solution on a single record using the model in debug mode.

cmd - `python extractionCLI.py -id "5e6a3b6bbf13490011565be8" -m -d`

```
Anaconda Prompt - conda deactivate

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -id "5e6a3b6bbf13490011565be8" -m -d
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\S020057657\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Processing 1 record with Id-5e6a3b6bbf13490011565be8
No of Products: 1

using model for prediction
processing product-0 with id-5e6a3b6bbf13490011565be8

processing Image- https://images-na.ssl-images-amazon.com/images/I/71ey2Rxhd-L._SX425_.jpg
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\S020057657\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Processing 1 record with Id-5e6a3b6bbf13490011565be8
No of Products: 1

using model for prediction
processing product-0 with id-5e6a3b6bbf13490011565be8

model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71X0I8mC6FL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/81pKvIMJ-oL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71e0Tt4mzSL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71JjBoz9zdL._SX425_.jpg
model prediction- No data
Skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71MPQcYgJ6L._SX425_.jpg
model prediction- Data Present
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51r0EEJZoeL._SX425_.jpg
model prediction- No data
Skipping Image!

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>
```

9. Running the solution on a single record without using the model in debug mode.

cmd - python extractionCLI.py -id "5e6a3b6bbf13490011565be8" -d

```
Anaconda Prompt - conda deactivate

(infant_1) D:\Wipro\sourabh\topgear\Infant Nutrition\CLI_TOOL>python extractionCLI.py -id "5e6a3b6bbf13490011565be8" -d
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\S020057657\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Processing 1 record with Id-5e6a3b6bbf13490011565be8
No of Products: 1

using rule-based logic for prediction
processing product-0 with id-5e6a3b6bbf13490011565be8

processing Image- https://images-na.ssl-images-amazon.com/images/I/71ey2Rxhd-L._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71X0I8mC6FL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/81pKvIMJ-oL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71e0Tt4mzSL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71JjBoz9zdL._SX425_.jpg
both nutrients and ingredients present
Extracting Nutrients!
Extracting Ingredients!
Extraction Complete!

processing Image- https://images-na.ssl-images-amazon.com/images/I/71MPQcYgJ6L._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

processing Image- https://images-na.ssl-images-amazon.com/images/I/51r0EEJZoeL._SX425_.jpg
checking for other rules to confirm
No data present skipping Image!

Extracted Items-
Ingredients
[{'name': 'calcium carbonate', 'amount': 0, 'unit': 0, 'referenceValue': 0}
{'name': 'ferrous sulfate', 'amount': 0, 'unit': 0, 'referenceValue': 0}
{'name': 'manganese sulfate', 'amount': 0, 'unit': 0, 'referenceValue': 0}
{'name': 'sodium selenate', 'amount': 0, 'unit': 0, 'referenceValue': 0}
{'name': 'potassium phosphate', 'amount': 0, 'unit': 0, 'referenceValue': 0}
{'name': 'sunflower', 'amount': 0, 'unit': 0, 'referenceValue': 0}]
```

```
Anaconda Prompt - conda deactivate

name: monoglycerides, amount: 0, unit: 0, referenceValue: 0}
name: calmi, amount: 0, unit: 0, referenceValue: 0}
name: chloride, amount: 0, unit: 0, referenceValue: 0}
name: calcium, amount: 0, unit: 0, referenceValue: 0}
name: magnesium, amount: 0, unit: 0, referenceValue: 0}
name: bitartrate, amount: 0, unit: 0, referenceValue: 0}
name: palmitate, amount: 0, unit: 0, referenceValue: 0}
name: roon, amount: 0, unit: 0, referenceValue: 0}
name: citrate, amount: 0, unit: 0, referenceValue: 0}
name: taurine, amount: 0, unit: 0, referenceValue: 0}
name: inositol, amount: 0, unit: 0, referenceValue: 0}
name: mixed, amount: 0, unit: 0, referenceValue: 0}
name: tocopherol, amount: 0, unit: 0, referenceValue: 0}
name: concentrate, amount: 0, unit: 0, referenceValue: 0}
name: vitamin, amount: 0, unit: 0, referenceValue: 0}
name: acetyl, amount: 0, unit: 0, referenceValue: 0}
name: palette, amount: 0, unit: 0, referenceValue: 0}
name: cup, amount: 0, unit: 0, referenceValue: 0}
name: thymine, amount: 0, unit: 0, referenceValue: 0}
name: folk, amount: 0, unit: 0, referenceValue: 0}
name: acid, amount: 0, unit: 0, referenceValue: 0}
name: woman, amount: 0, unit: 0, referenceValue: 0}
name: satin, amount: 0, unit: 0, referenceValue: 0}
name: cyanocobalamin, amount: 0, unit: 0, referenceValue: 0}
name: nucleotides, amount: 0, unit: 0, referenceValue: 0}
name: adenosine, amount: 0, unit: 0, referenceValue: 0}

Nutrients
name: meg, amount: 160, unit: 0, referenceValue: 0}
name: phosphor, amount: 82, unit: 0, referenceValue: 0}
name: fly, amount: 56, unit: 0, referenceValue: 0}
name: carbohydrate, amount: 107, unit: 0, referenceValue: 0}
name: vitamin-b, amount: 026, unit: 0, referenceValue: 0}
name: water, amount: 133, unit: 0, referenceValue: 0}
name: 133, amount: 100, unit: 0, referenceValue: 0}
name: long, amount: 19, unit: 0, referenceValue: 0}
name: unoleicaco, amount: 8, unit: 0, referenceValue: 0}
name: 838%, amount: 079, unit: 0, referenceValue: 0}
name: flaccid, amount: 16, unit: 0, referenceValue: 0}
name: pantothenic-acid, amount: 470, unit: 0, referenceValue: 0}
name: viamnak, amount: 300, unit: 0, referenceValue: 0}
name: vitamme, amount: 15, unit: 0, referenceValue: 0}
name: s-sodium, amount: 110, unit: 0, referenceValue: 0}
name: potassium, amount: 100, unit: 0, referenceValue: 0}
name: 49-chloride, amount: 68, unit: 0, referenceValue: 0}
name: comments, amount: 30, unit: 0, referenceValue: 0}
name: selenium, amount: 0, unit: 0, referenceValue: 0}
name: manufactured, amount: 0, unit: 0, referenceValue: 0}
name: coppery, amount: 0, unit: 0, referenceValue: 0}
name: questions, amount: 0, unit: 0, referenceValue: 0}
```

Type here to search

19:32
16-03-2020