



GoodVitamins

An Interactive Vitamin Review
Summarization App From iHerb.



Skills:

Web Scraping, Data Cleaning, Text Preprocessing, Visualization,
Machine Learning, NLP.



Tools:

Selenium, Universal Sentence Encoder, K-means clustering,
TextBlob.



linkedin.com/in/andyphamto/



github.com/andy-pham-72/

Introduction:

- Vitamin supplements are a **multibillion-dollar industry**. Many people are taking supplements daily to **compensate** for the **sparse diet**.
- Supplements are regulated by the FDA; However, the framework is different from drugs. The **FDA does not review dietary supplement products for safety and effectiveness** before they are marketed.
- Question: Are all the supplements effective?



About iHerb

- iHerb is an **online global supplements retailer** supplying a wide variety of products to **35 countries**.
- Offers **many options** for vegans, vegetarians, paleo, dairy-free, gluten-free.
- An **active community** where users keep posting and **updating detailed review contents** about the products.



The Work Progress



Raw Dataset

item_image_link	item_brand	item_name	item_description	item_price	total_rate	review_contents	individual_rate	product_helpful	product_not_helpful
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	I bought it because of good caustic vitamin C, and came back to try to buy it, but	5	3109	63
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	With such preparation as ascorbic acid or vitamin C, we are familiar with detstva.E	5	2247	125
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	I bought it for a long time without any doubt. Most of the people who buy at least	1	1584	50
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	I bought this vitamin at a promotional price, I personally really like it, but for such a	5	702	3
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	very cool product, I recommend it to	5	595	0
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	It's a very cool product. Recommend and satisfy everyone for purchase	4	588	3
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	Very cool product, me	5	578	3
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	I took it for the action and did not regret it, excellent packaging, 60 pieces will be e	5	430	1
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	cool product	5	403	2
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	cool very cool	5	392	2
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	Hello everyone, I decided to write a review on the California Gold Nutrition "C 100	5	407	7
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	I drink this vitamin C so as not to get sick. As soon as I feel that my throat is starti	5	484	12
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	We use vitamin C almost all year round with the whole family. Optimal dosages are	4	324	1
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	I ordered 60 ounces and ordered 240 ounces, but the packaging was different. I al	1	403	14
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	After getting up, I drank hot water and took this vitamin C. I am hungry. After 15 to	1	339	36
https://s3.images-iherb.com	California Gold Nutrition	Gold C, Vitamin C, 1,000 mg, 60 Veggie Capsules	California Gold Nutrition Gold C™, 1000 mg, Featuring USP Grade	CA\$6.52	4.7	I think few people will be interested in reading praise about vitamin "C", I will not t	5	131	3

- **275 products.**
- **22,000 review contents (80 reviews per products).**

Data Cleaning

item_description	item_price	total_rate	review_contents	individual_rate	product_helpful	product_not_helpful	product_id	brand_id
California Gold Nutrition Gold C™, 1000 mg, Fe...	6.52	4.7	I bought it because of good caustic vitamin C,...	5	3109	63	78	5
California Gold Nutrition Gold C™, 1000 mg, Fe...	6.52	4.7	With such preparation as ascorbic acid or vita...	5	2247	125	78	5
California Gold Nutrition Gold C™, 1000 mg, Fe...	6.52	4.7	I bought it for a long time without any doubt....	1	1584	50	78	5
California Gold Nutrition Gold C™, 1000 mg, Fe...	6.52	4.7	I bought this vitamin at a promotional price, ...	5	702	3	78	5
California Gold Nutrition Gold C™, 1000 mg, Fe...	6.52	4.7	very cool product, I recommend it to	5	595	0	78	5

- **211 products.**
- **16,800 review contents (80 reviews per products).**

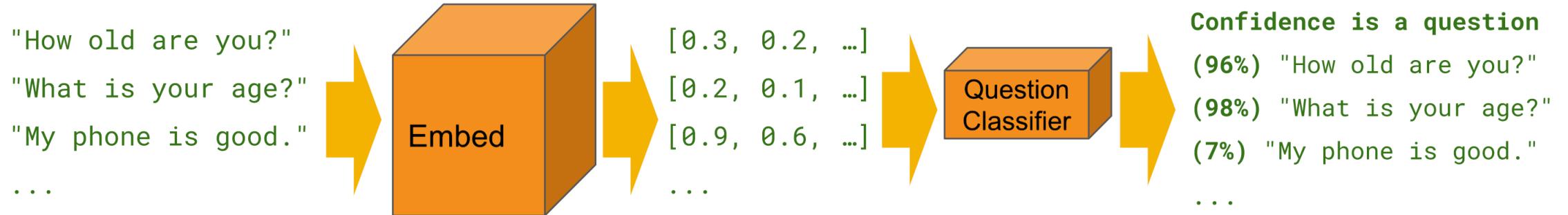
Text Data Preprocessing

- **Emoji Removal.**
- **Spelling Correction.** (The review contents were translated from different languages to English)

=> **Transforms** the text into **a form** that is **predictable** and **analyzable** so that **Machine Learning algorithms** can **perform better**.



Exploratory data analysis



First Approach:

- **Universal Sentence Encoder (USE)** from **TensorFlow Hub**.

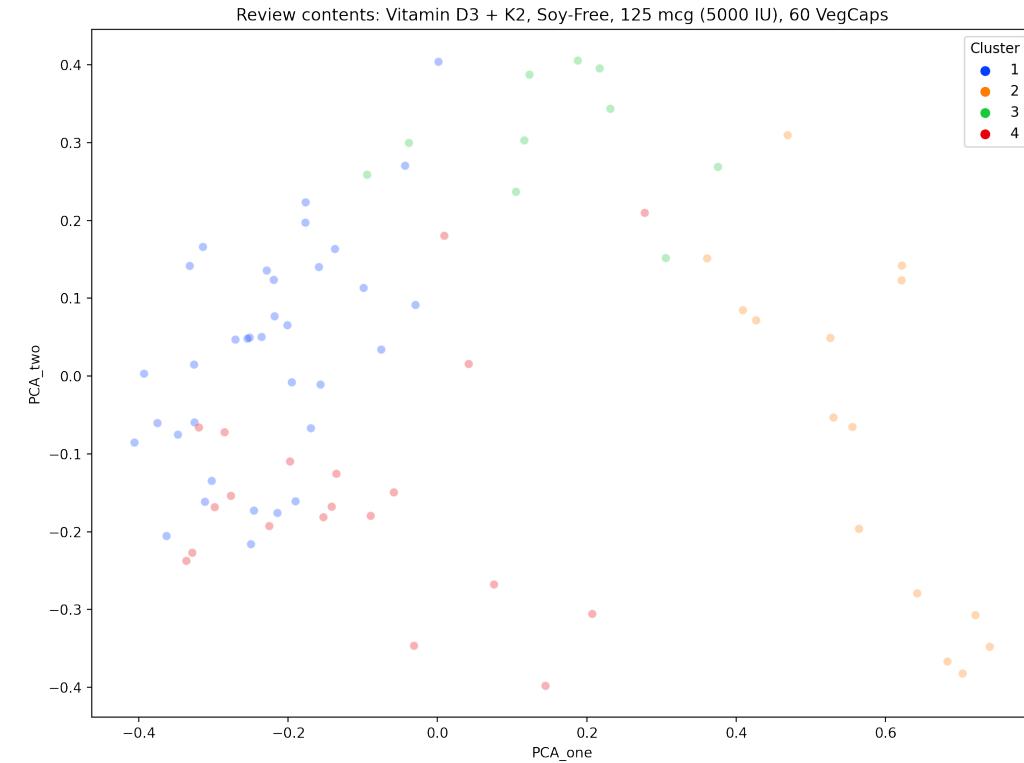
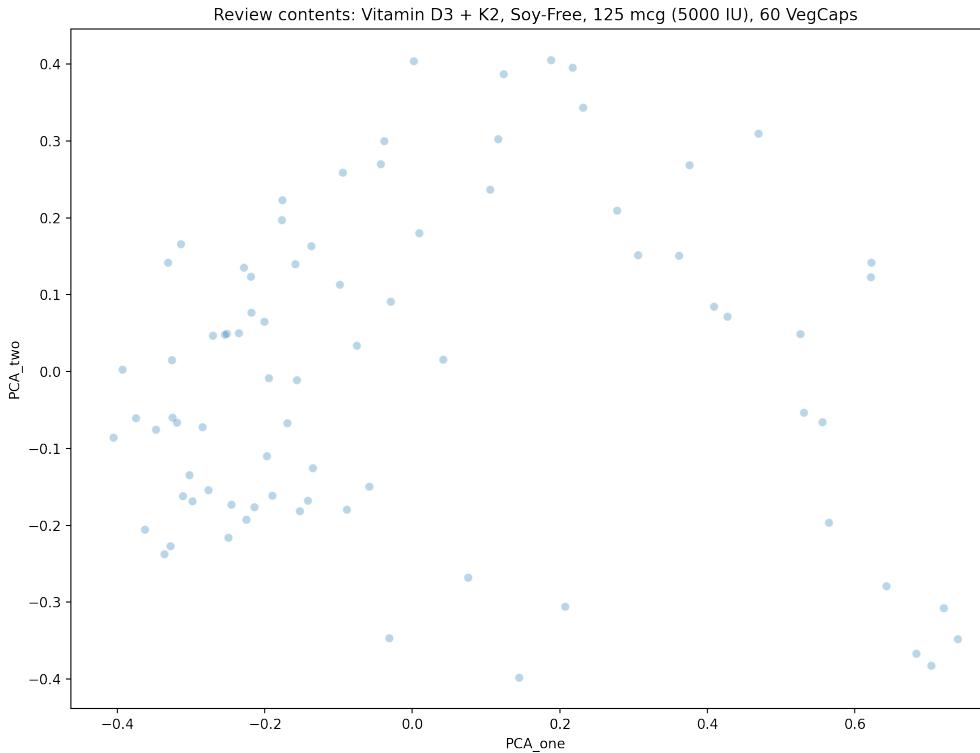
USE encodes text into high dimensional vectors that can be used for text classification, semantic similarity and clustering.

The concept is to classify the sentences with similar meanings and place them to similar locations.



k-means clustering from Scikit-learn.

Example: 'Solaray - Vitamin D3 + K2, Soy-Free, 125 mcg (5000 IU), 60 VegCaps'



- **Cluster #1 sentences** (density score: 0.55118):

Many people don't know that vitamin D must be taken along with K2, which was only discovered in 2010. It is K2 that, when taking Vitamin D3, directs calcium in the right direction! K2 comes in two forms - MK2 and MK7. The first one causes changes in the density of the blood, therefore, one must be careful with K2 in this form. This preparation contains K2 in the required and most correct form of MK7. Contains a good dosage of vitamin D3 - 5000 mg, and also calcium. I will buy more !!!

I really like this vitamin D3, it comes with k2 for better absorption. I drink on an ongoing basis, I feel cheerful, with ARVI I take 5-6 tablets a day along with vitamin C and zinc, the disease does not progress. Recommend at purchase. Please like it if you liked the review and was helpful It is K2 that, when taking Vitamin D3, directs calcium in the right direction! K2 comes in two forms - MK2 and MK7. The first one causes changes in the density of the blood, therefore, one must be careful with K2 in this form. This preparation contains K2 in the required and most correct form of MK7. Contains a good dosage of vitamin D3 - 5000 mg, and also calcium. I order on the IHerb website and recommend it to everyone!

- **Cluster #2 sentences** (density score: 0.40833):

Very, very cool product, I recommend

Great product! Recommend!

cool very cool product recommend it

- **Cluster #3 sentences** (density score: 0.29849):

the whole family raises vitamin D perfectly. I attach a photo of my son. We drank for 3 months, if it ends up they took another one on trial, but we pass an analysis for dynamics

I take only this company for myself and my husband. We drink from October to April. D3 for analyzes is excellent. Thanks for the class.

I ordered it for my husband and mother. After the course of admission, the level of D in the blood increased

- **Cluster #4 sentences** (density score: 0.53988):

I learned that vitamin D is better absorbed with vitamin K and decided to try these vitamins. In combination with Vit. C, it raised ferritin well within a month. I continue to drink.

I love that this supplement contains vitamin K-2 in addition to the vitamin D-3 we need. A good combination of price and quality.

I've heard a lot of good reviews about these vitamins. Vitamin D3 should be drunk in combination with K2 for better absorption. I accept it constantly.

Future work!

- Scraps more review contents for each product.
- Experiments **Long Short-Term Memory Networks** to summarize the **Cluster 1 sentences**.
- Tries Sentence Scoring based on Word Frequency.
- Creates an APP.

THANK YOU
FOR YOUR ATTENTION