

Opdracht 3

Collaborative Filtering, most_viewed_products

We hebben een verzameling aan producten die worden gevormd door alle profielen. Al deze profielen zullen 1 of meerdere product id's mee geven. De verzameling aan producten is q, 1 product is p en een klant is K, R is recommended.

$$((K(p) > K(q)) \rightarrow R(p))$$

Content filtering, similar_product

Voordat we de logica gaan uitwerken, kunnen we eerst beter de wiskunde achter count vectorization begrijpen.

COUNT VECTORIZATION

$$A \cdot B = (a_1 \times b_1) + (a_2 \times b_2) + \dots + (a_n \times b_n)$$

$$\cos(a, b) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n (a_i)^2} \sqrt{\sum_{i=1}^n (b_i)^2}}$$

$$\|A\| = \sqrt{(a_1)^2 + (a_2)^2 + \dots + (a_n)^2} \quad \times \quad \|A\| = \sqrt{(a_1)^2 + (a_2)^2 + \dots + (a_n)^2}$$

Welk antwoord willen we?

Hoe Dichter bij de 1
Hoe meer vergelijkbaar

Hoe Dichter bij de -1
Hoe minder vergelijkbaar

COUNT VECTORIZATION

$$A \cdot B = (a_1 \times b_1) + (a_2 \times b_2) + \dots + (a_n \times b_n)$$

$$\cos(a, b) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n (a_i)^2} \sqrt{\sum_{i=1}^n (b_i)^2}}$$

$$\|A\| = \sqrt{(a_1)^2 + (a_2)^2 + \dots + (a_n)^2} \quad \times \quad \|A\| = \sqrt{(a_1)^2 + (a_2)^2 + \dots + (a_n)^2}$$

Bijvoorbeeld je wilt recommendations voor product

Input data:

29099 Huishouden Toilet en keuken Toiletblokken Kantoor Spray Katoen
1 1 1 1 1 1 1 1

Vergelijk je met:

31625 Huishouden Toilet en keuken Luchtverfrissers Kantoor Accessoires Vanille
0 1 1 1 0 1 0 0

$$A \cdot B = (a_1 \times b_1) + (a_2 \times b_2) + \dots + (a_n \times b_n)$$

$$(1 \times 0) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 0) + (1 \times 1) + (1 \times 0) + (1 \times 0) = 4$$

$$\|A\| = \sqrt{(a_1)^2 + (a_2)^2 + \dots + (a_n)^2}$$

$$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 \times 0 + 1 + 1 + 1 + 0 + 1 + 0 + 0$$

Wortel

$$2,82.... \times 2 = 5,65$$

UITEINDELIJK DELEN

$$4 / 5,65 = 0,71$$

0,71

|

Wanneer we er van uitgaan dat alle producten attributen hebben die overeen moeten komen met het product waarmee je uiteindelijk `similar_products` wilt genereren, kunnen we wel wat verzinnen. Hierin is `p` het product dat je uit kiest, en `q` het product waarmee je vergelijkt, `a` is een attribuut. `T` werkt als teller die de overeenkomsten bij houdt. `R` is een recommended product.

$$(p(a) = q(a) \rightarrow T(q) + 1)$$

$$((T(q) > T(q) \rightarrow R(q))$$