



# Εισαγωγή στην Αναλυτική των Δεδομένων

*Άσκηση Πράξης*

**Παραγωγή Συστάσεων με UBCF και kNN**

Δ.Α. Δέρβος

<https://d-a-d.weebly.com>

Σίνδος, Δεκέμβριος 2022

# Πηγή - Αναφορά

*Bill Qualls, De Paul University:*

[http://www.billqualls.com/pa/Qualls\\_ECT584\\_RecommenderSystems.pdf](http://www.billqualls.com/pa/Qualls_ECT584_RecommenderSystems.pdf)



# Αναγνώστες Βαθμολογούν Βιβλία

	TRUE BELIEVER	THE DA VINCI CODE	THE WORLD IS FLAT	MY LIFE SO FAR	THE TAKING	THE KITE RUNNER	RUNNY BABBIT	HARRY POTTER
U1	1	5		3			3	5
U2	5	4			3	2	1	
U3	3		1	2	2			5
U4		3			4	1		3
U5	2	4	3			2	2	
U6	5			3	1		3	1
U7	1	4	5	5	2			4
U8	2	1			4	5	1	
U9			3	2	2			5
U10	3	5	1				4	4
U11			2	1		2		3
U12	4	4		2		1	1	4
U13			2		4		4	5
U14		5	3	3	2		1	1
U15		2			3	3		2
U16		3	2	1	1		4	4
U17	1	5	1	2		4		4
U18	5		4		3	3	4	5
U19		4		2		5	1	5
U20	2	5	1	1	5	3		4

# Αναγνώστες Βαθμολογούν Βιβλία

	TRUE BELIEVER	THE DA VINCI CODE	THE WORLD IS FLAT	MY LIFE SO FAR	THE TAKING	THE KITE RUNNER	RUNNY BABBIT	HARRY POTTER
U1	1	5		3			3	5
U2	5	4			3	2	1	
U3	3		1	2	2			5
U4		3			4	1		3
U5	2	4	3			2	2	
U6	5			3	1		3	1
U7	1	4	5	5	2			4
U8	2	1			4	5	1	
U9			3	2	2			5
U10	3	5	1				4	4
U11			2	1		2		3
U12	4	4		2		1	1	4
U13			2		4		4	5
U14		5	3	3	2		1	1
U15		2			3	3		2
U16		3	2	1	1		4	4
U17	1	5	1	2		4		4
U18	5		4		3	3	4	5
U19		4		2		5	1	5
U20	2	5	1	1	5	3		4

Πληροφορία  
Αναφοράς,  
“Μοντέλο”

NU1	3		5	4	2	3		5
NU2		5	2	2	4		1	3

Πρόβλεψη (σύσταση;) για αυτά που δεν έχουν διαβάσει οι NU<sub>1</sub> και NU<sub>2</sub>;



# Βαθμός Ομοιότητας (Pearson)

	TRUE BELIEVER	THE DA VINCI CODE	THE WORLD IS FLAT	MY LIFE SO FAR	THE TAKING	THE KITE RUNNER	RUNNY BABBIT	HARRY POTTER
U1	1	5		3			3	5
U2	5	4			3	2	1	
U3	3		1	2	2			5
U4		3			4	1		3
U5	2	4	3			2	2	
U6	5			3	1		3	1
U7	1	4	5	5	2			4
U8	2	1			4	5	1	
U9			3	2	2			5
U10	3	5	1				4	4
U11			2	1		2		3
U12	4	4		2		1	1	4
U13			2		4		4	5
U14		5	3	3	2		1	1
U15		2			3	3		2
U16		3	2	1	1		4	4
U17	1	5	1	2		4		4
U18	5		4		3	3	4	5
U19		4		2		5	1	5
U20	2	5	1	1	5	3		4

$\text{Sim}(U_6, NU_2)$

Google Sheets: CORREL()

NU1	3		5	4	2	3		5
NU2		5	2	2	4		1	3

# Αναγνώστης Βαθμολογεί Βιβλίο

B<sub>1</sub> B<sub>2</sub> B<sub>3</sub> B<sub>4</sub> B<sub>5</sub> B<sub>6</sub> B<sub>7</sub> B<sub>8</sub>

	TRUE BELIEVER	THE DA VINCI CODE	THE WORLD IS FLAT	MY LIFE SO FAR	THE TAKING	THE KITE RUNNER	RUNNY BABBIT	HARRY POTTER
U1	1	5		3			3	5
U2	5	4			3	2	1	
U3	3		1	2	2			5
U4		3			4	1		3
U5	2	4	3			2	2	
U6	5			3	1		3	1
U7	1	4	5	5	2			4
U8	2	1			4	5	1	
U9			3	2	2			5
U10	3	5	1				4	4
U11			2	1		2		3
U12	4	4		2		1	1	4
U13			2		4		4	5
U14		5	3	3	2		1	1
U15		2			3	3		2
U16		3	2	1	1		4	4
U17	1	5	1	2		4		4
U18	5		4		3	3	4	5
U19		4		2		5	1	5
U20	2	5	1	1	5	3		4

Αναγνώστης: U<sub>8</sub>

Βιβλίο: THE KITE RUNNER (B<sub>6</sub>):

$$r(U_8, B_6) = 5$$

Ζητούμενα:

$r(U_1, B_2)$ ,  $r(U_1, B_7)$

$r(U_2, B_1)$ ,  $r(U_2, B_6)$

NU1	3		5	4	2	3		5
NU2		5	2	2	4		1	3

# Ο Υπολογισμός

Έστω ότι οι  $U_1$ ,  $U_5$  και  $U_7$  είναι αυτοί που μοιάζουν περισσότερο στον  $NU_1$ , τότε:

$$r(NU_1, B_2) = \frac{r(U_1, B_2) * \text{sim}(U_1, NU_1) + r(U_5, B_2) * \text{sim}(U_5, NU_1) + r(U_7, B_2) * \text{sim}(U_7, NU_1)}{\text{sim}(U_1, NU_1) + \text{sim}(U_5, NU_1) + \text{sim}(U_7, NU_1)}$$

$$r(NU_1, B_7) = \frac{r(U_1, B_7) * \text{sim}(U_1, NU_1) + r(U_5, B_7) * \text{sim}(U_5, NU_1) + r(U_7, B_7) * \text{sim}(U_7, NU_1)}{\text{sim}(U_1, NU_1) + \text{sim}(U_5, NU_1) + \text{sim}(U_7, NU_1)}$$

Ανάλογα:  $r(NU_2, B_1)$  και  $r(NU_2, B_6)$

Μένει να βρεθούν ποιοί τρεις των  $U_1, U_2, \dots, U_{20}$  έχουν μεγαλύτερη τιμή  $\text{sim}()/\text{CORREL}()$  για τον  $NU_1$  και ποιοί τρεις για τον  $NU_2$ ...



# Ερωτήσεις;

