

Kolaboratívne odporúčanie

Matej Končál

Vedúci: Ing. Peter Gašpar

Podobnosť v správaní používateľov

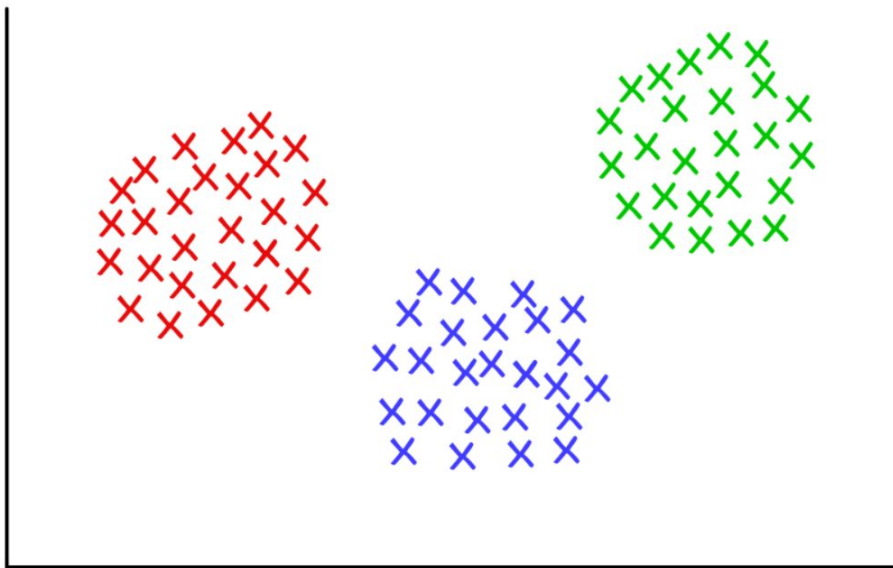
Používateľom s **podobným správaním** sa páčia **podobné produkty**

Dlhodobé

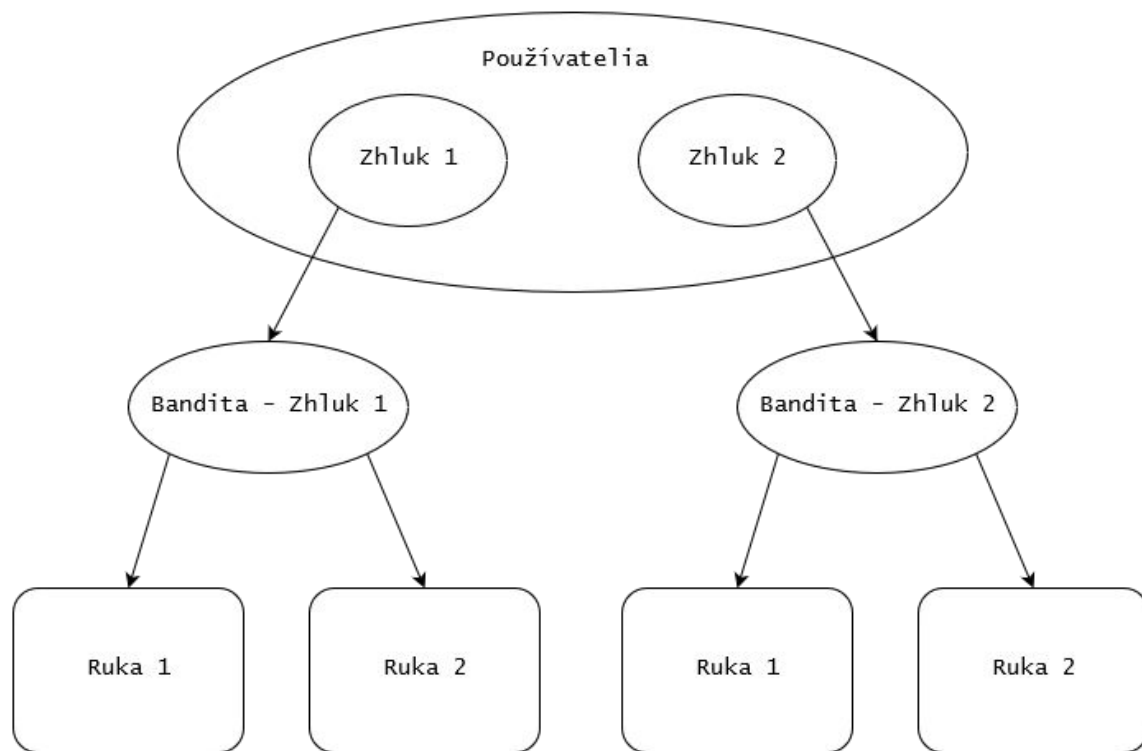
- mesiac
- rok

Krátkodobé

- návšteva (sedenie)
- deň



Kolaboratívny bandit



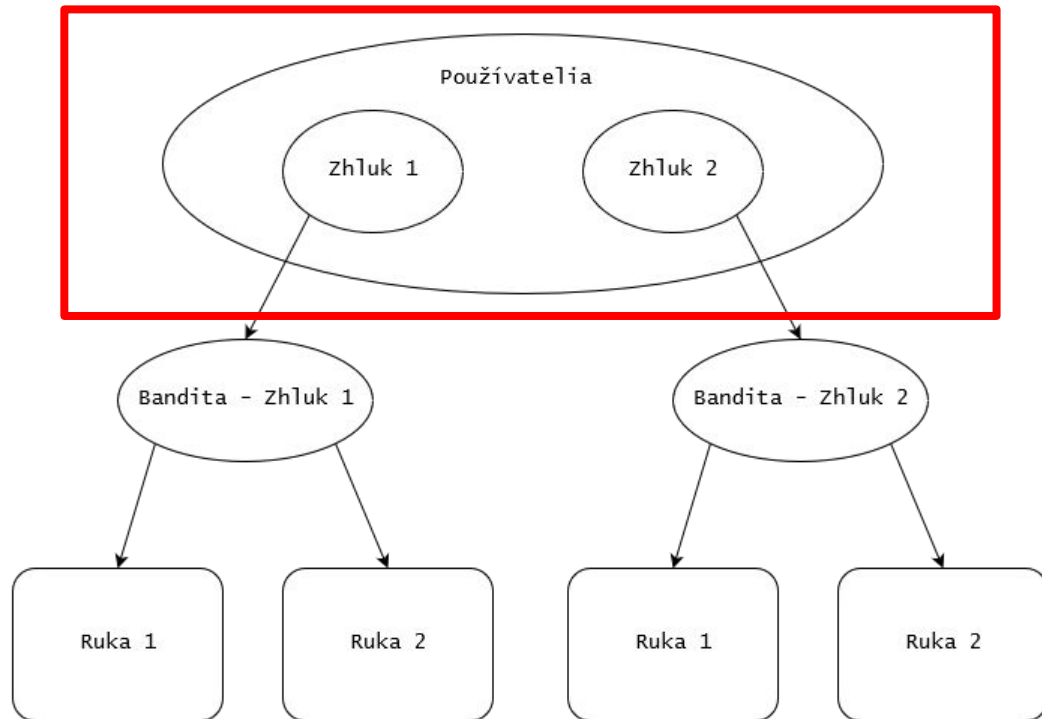
Zhlukovanie používateľov

1. Krok

- rozdelenie používateľov do skupín podľa správania

Črty používateľov [1]

- latentné
- maticová faktorizácia



[1] <https://www-users.cs.umn.edu/~chri3275/sdm-inter-rec.pdf>

Dátová vzorka

Dlhodobé / krátkodobé správanie

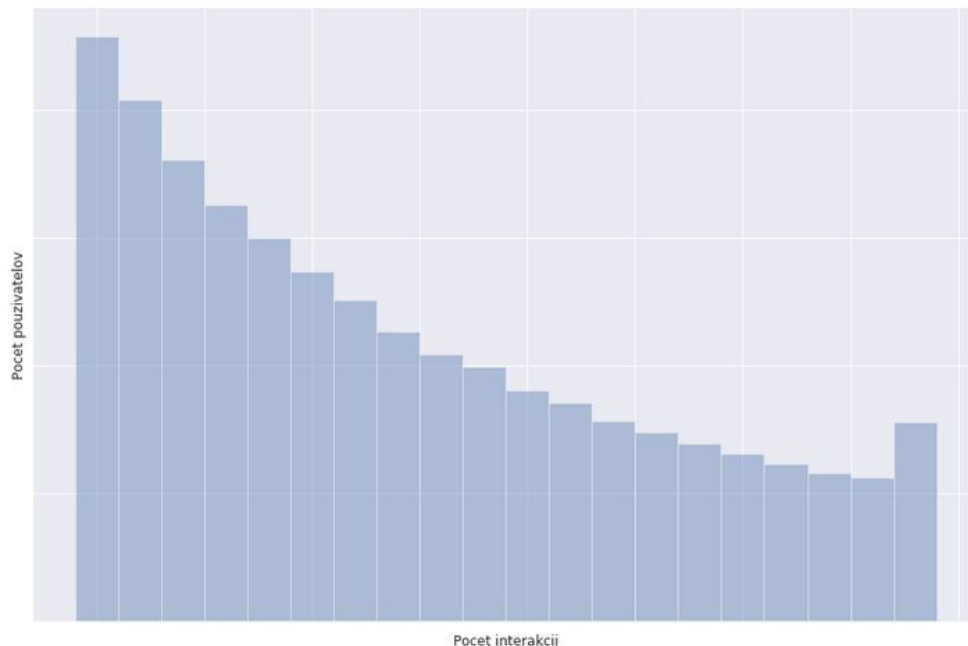
Min. interakcií používateľa: Q90

Max. interakcií používateľa: Q95

Počet používateľov: 200 000

Počet produktov: 85 000

Histogram počtu interakcií



Maticová faktORIZÁCIA

	Item			
	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

Rating Matrix

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

User Matrix

X

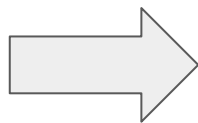
	W	X	Y	Z
	1.5	1.2	1.0	0.8
	1.7	0.6	1.1	0.4

Item Matrix

Vizualizácia MF pomocou t-SNE (očakávanie)

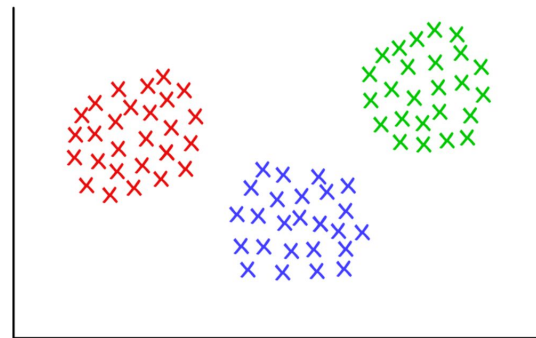
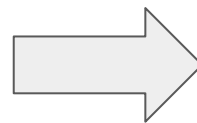
A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

User
Matrix

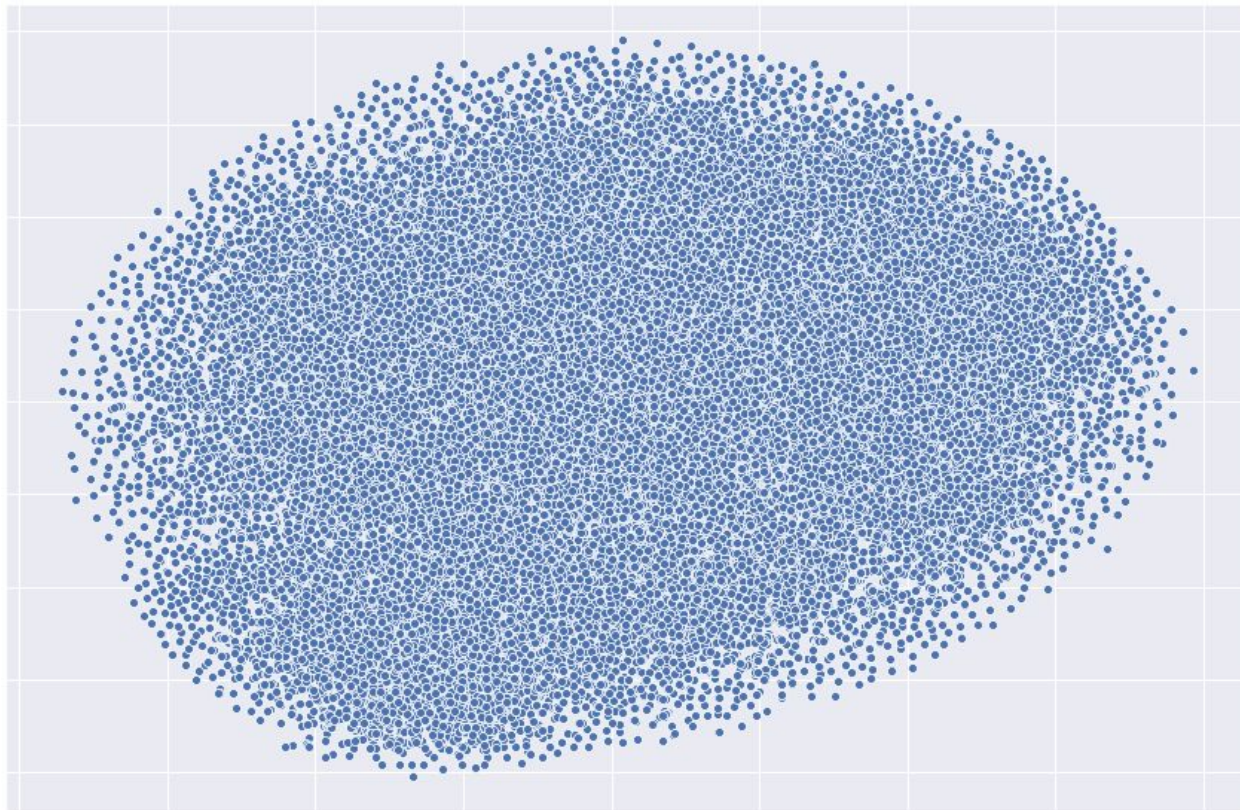


t-SNE

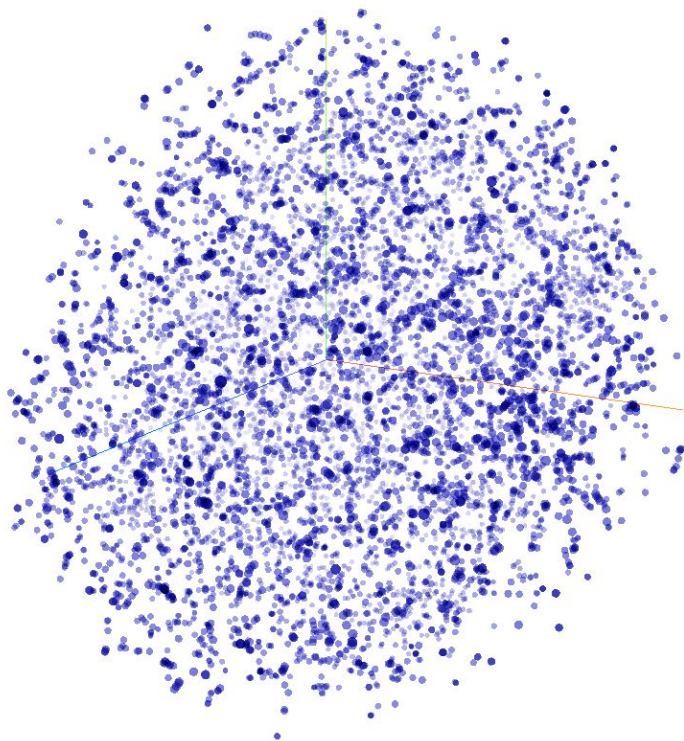
$x_D \rightarrow 2D$



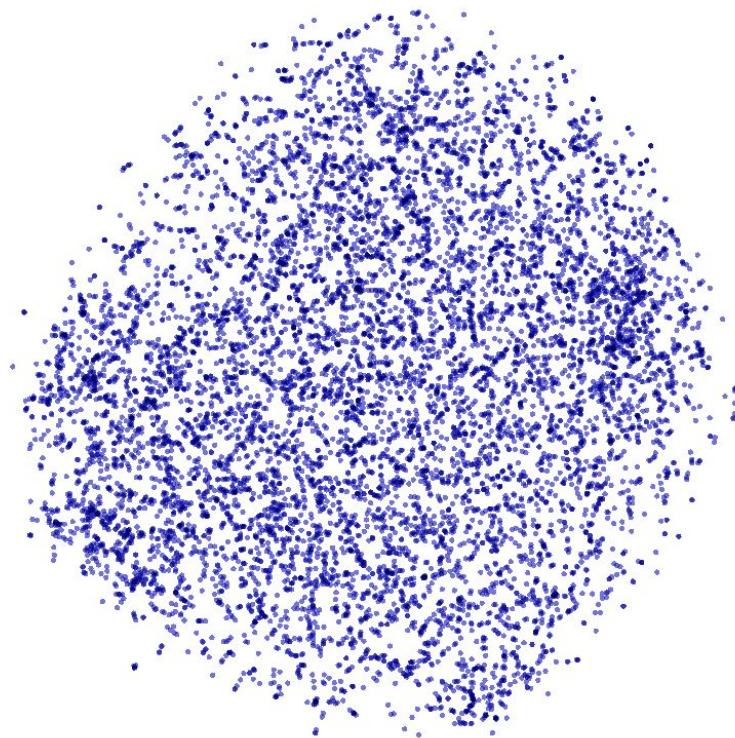
Vizualizácia MF pomocou t-SNE (realita)



t-SNE - počet epoch



3D



2D

Zhlukovanie - DBSCAN

Davies-Bouldin skóre

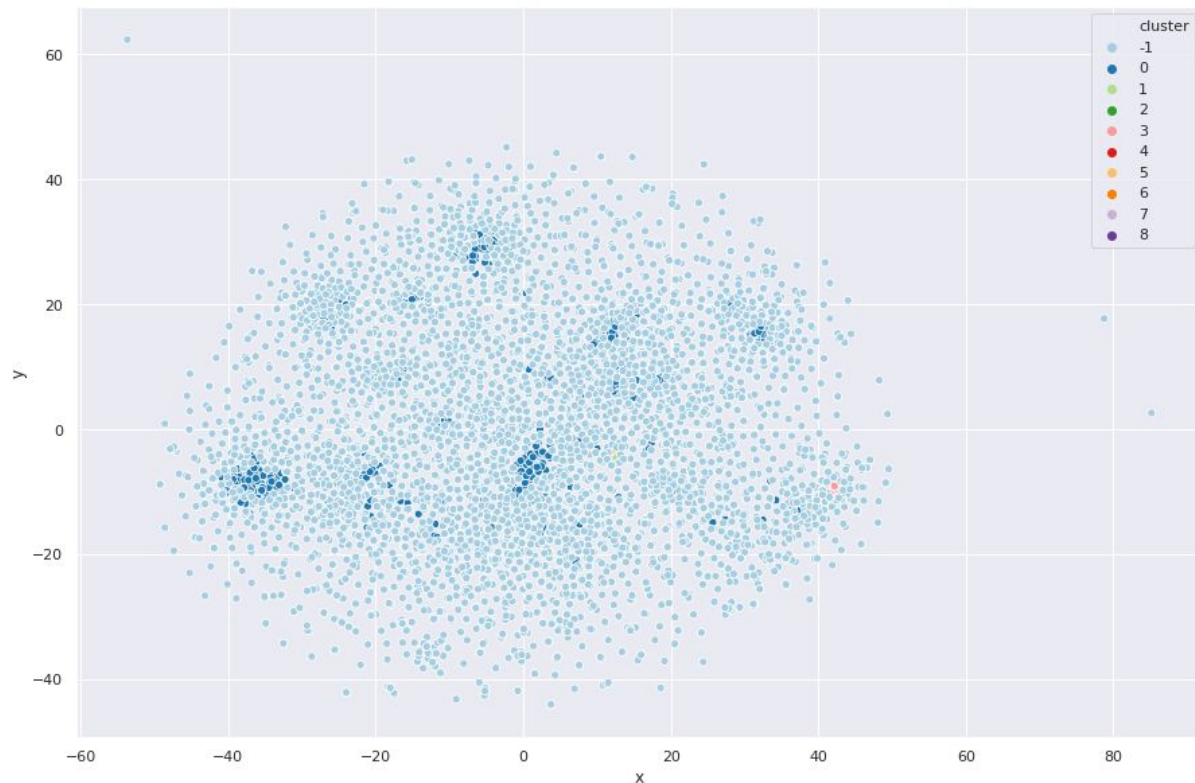
$[0, \infty)$

$= 4.287$

Silhouette koeficient

$[-1, 1]$

$= -0.312$



Maticová faktORIZÁCIA - výsledky

Rozdelenie podľa času - od 01.02.2019 testovacie dáta

Počet používateľov v teste: 43 000

LightFM

- loss = warp
- learning_schedule = adadelta
- no_components = 50
- epochs = 50

Precision@10 = 0.55%

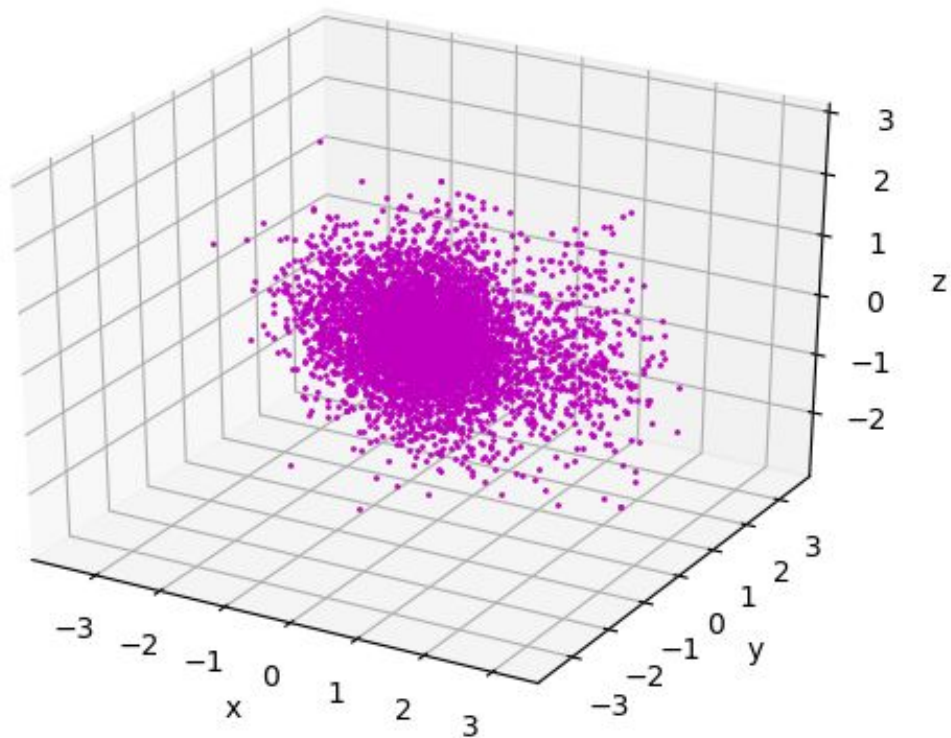
2000 používateľov != 0.0 = **HitRate@10 = 4.6%**

Vizualizácia MF - 3D

Components = 3

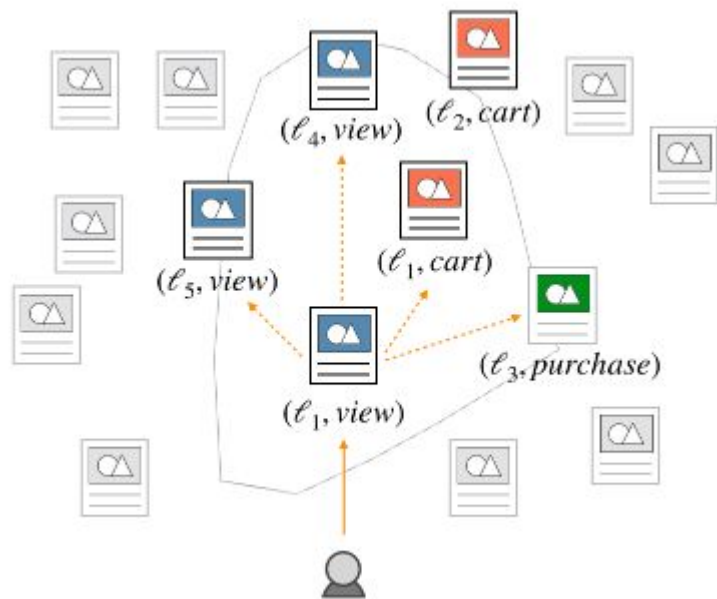
Epochs = 100, 250, ...

Precision@10 = 0.27%

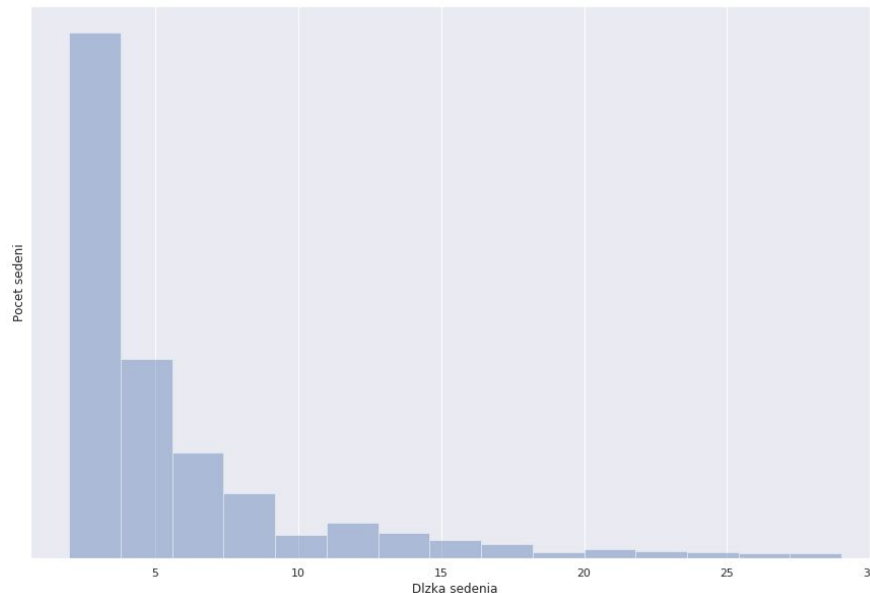


Modelovanie krátkodobého správania

Sedenie = [(A, pozretie), (B, pozretie), (A, pozretie), (A, nákup)]



Histogram dĺžky sedení

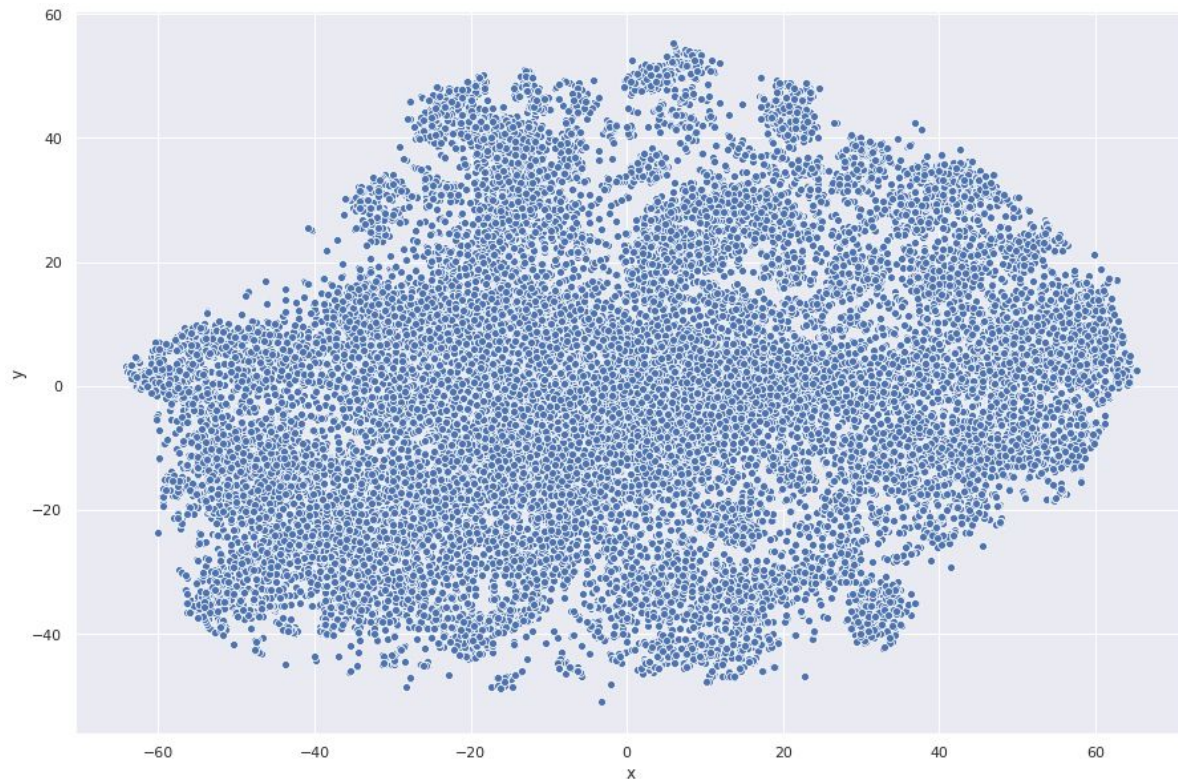


Krátkodobé správanie

Sedenie = [Produkt A, Produkt B, Produkt C, Produkt A]

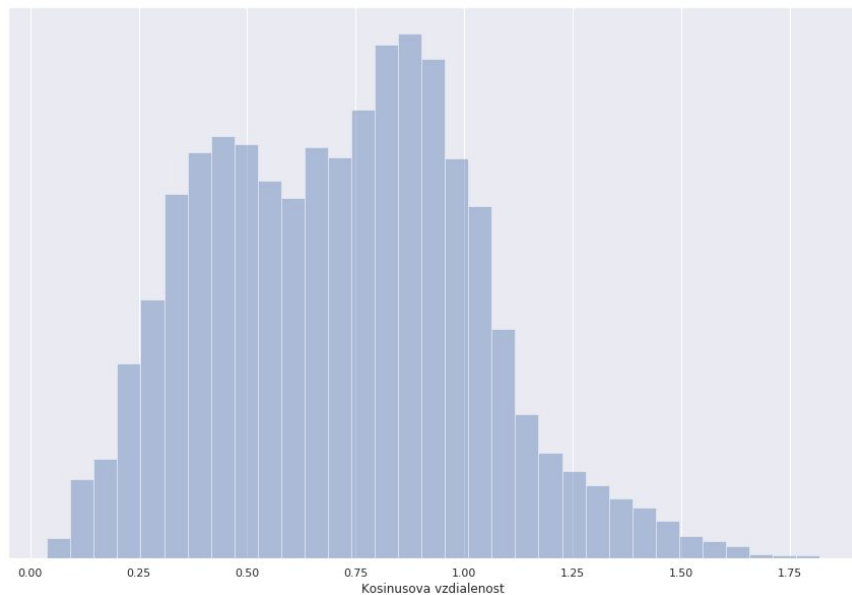


Vizualizácia sedení pomocou t-SNE

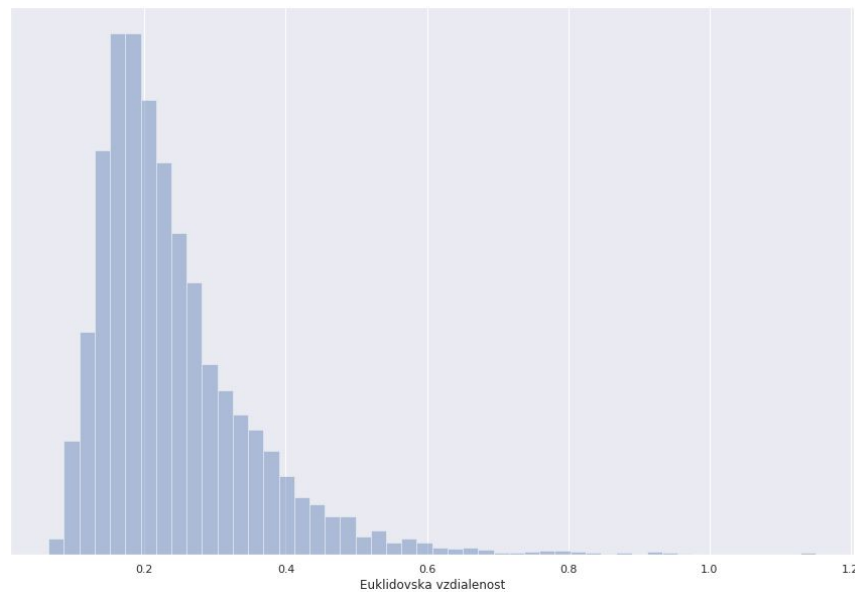


Vzdialenosť sedení pre používateľa

Kosínusová vzdialenosť



Euklidovská vzdialenosť

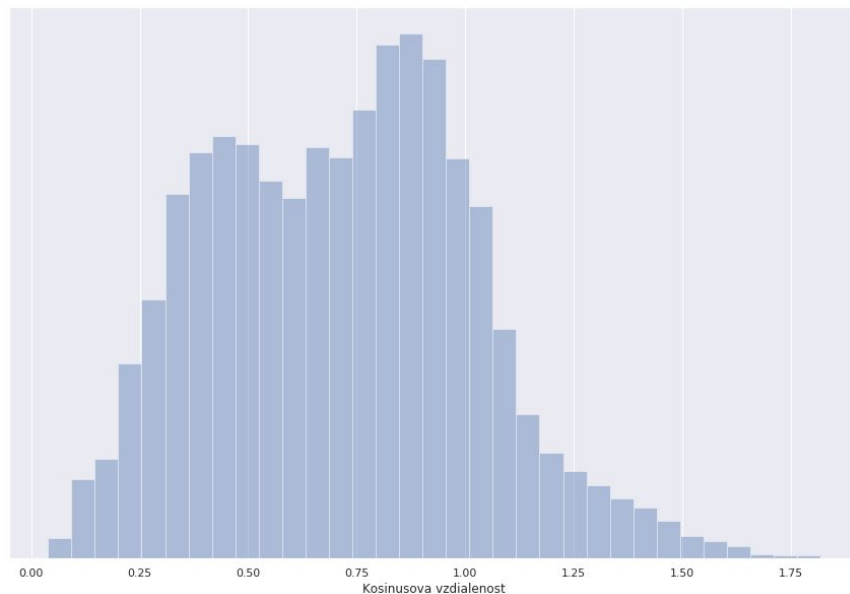


Správanie používateľov

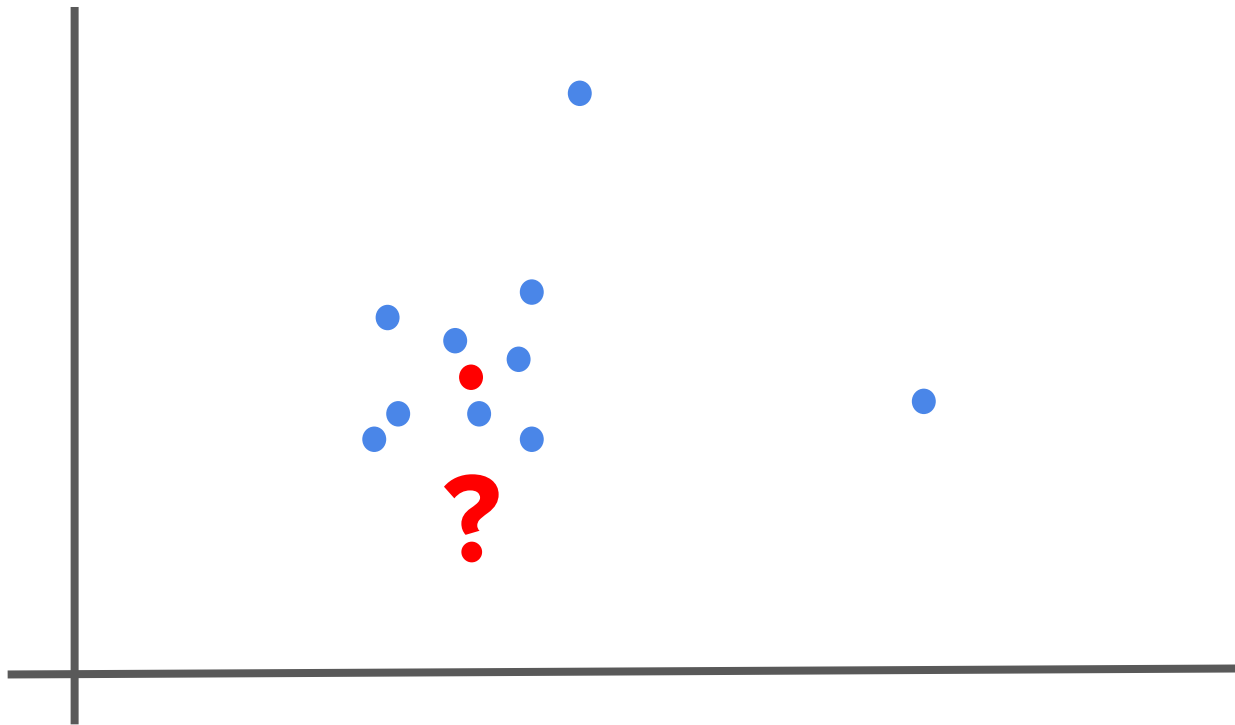
2 skupiny používateľov

- dlhodobé správanie
 - Sedenia sú relatívne blízko seba
- krátkodobé správanie
 - Sedenia sú ďaleko od seba

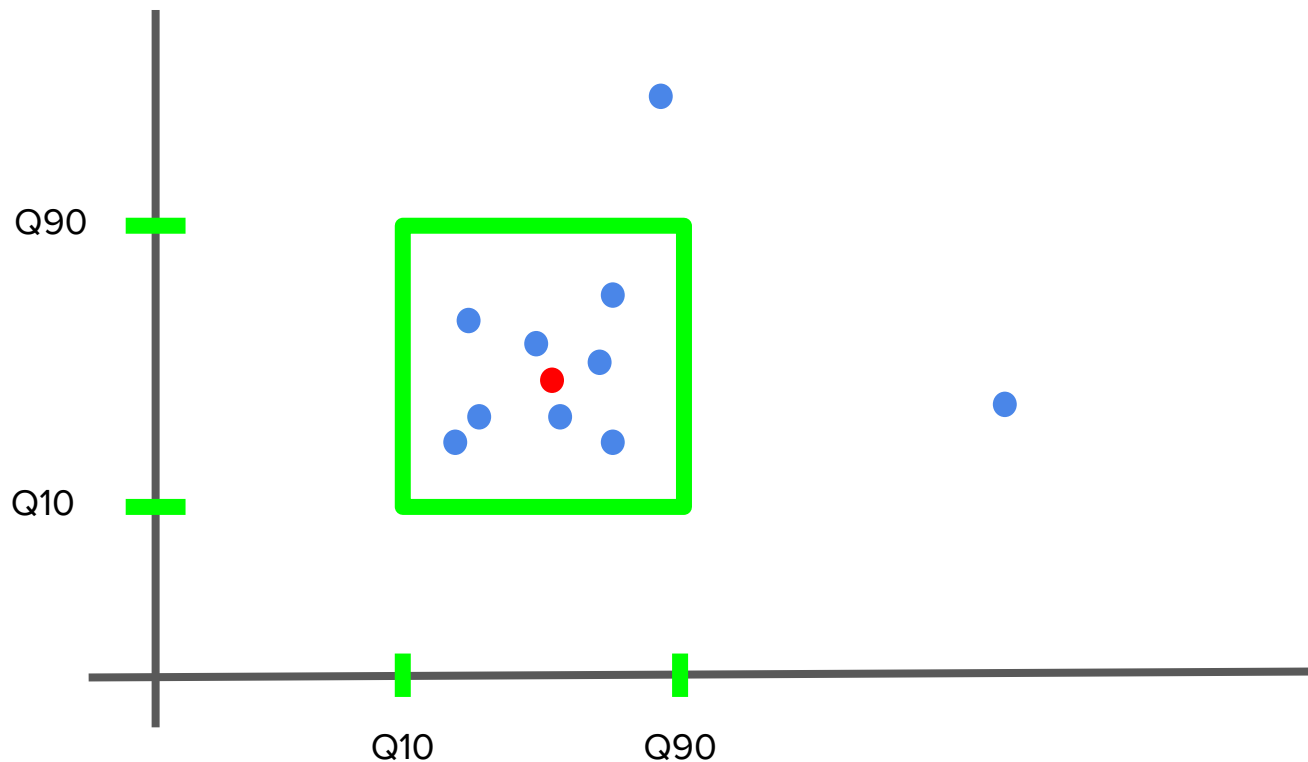
Kosínusová vzdialenosť



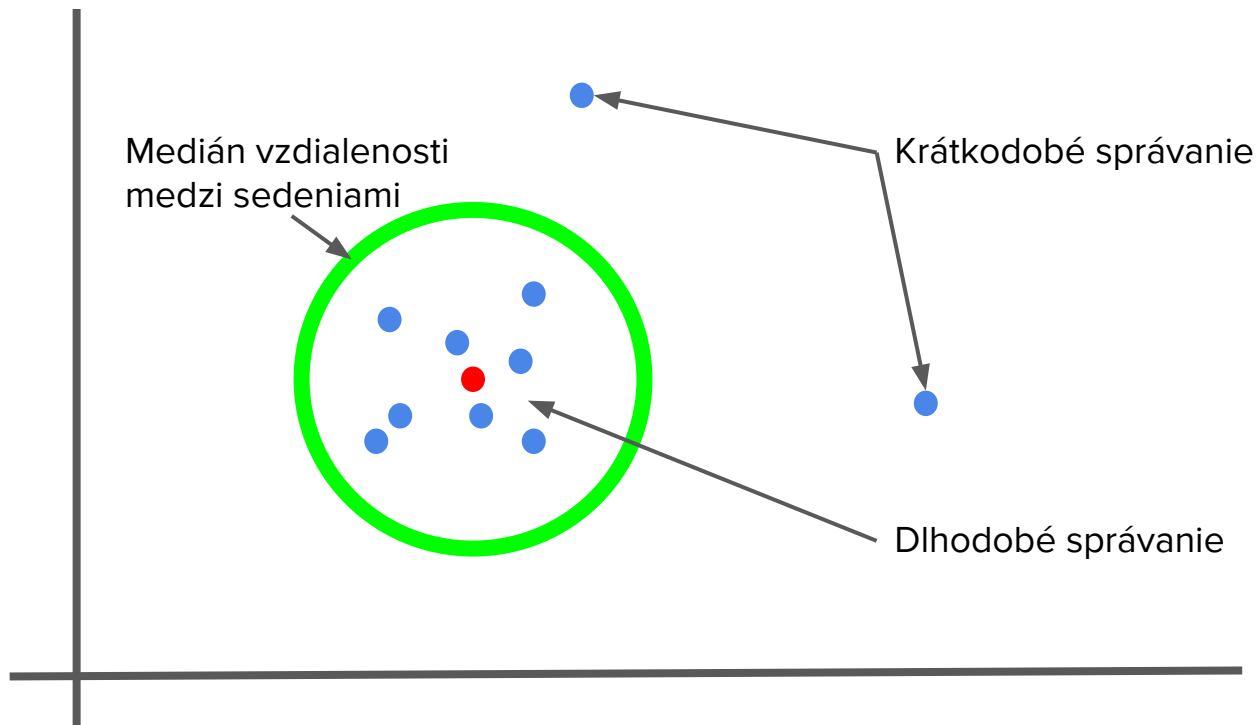
Modelovanie správania používateľa



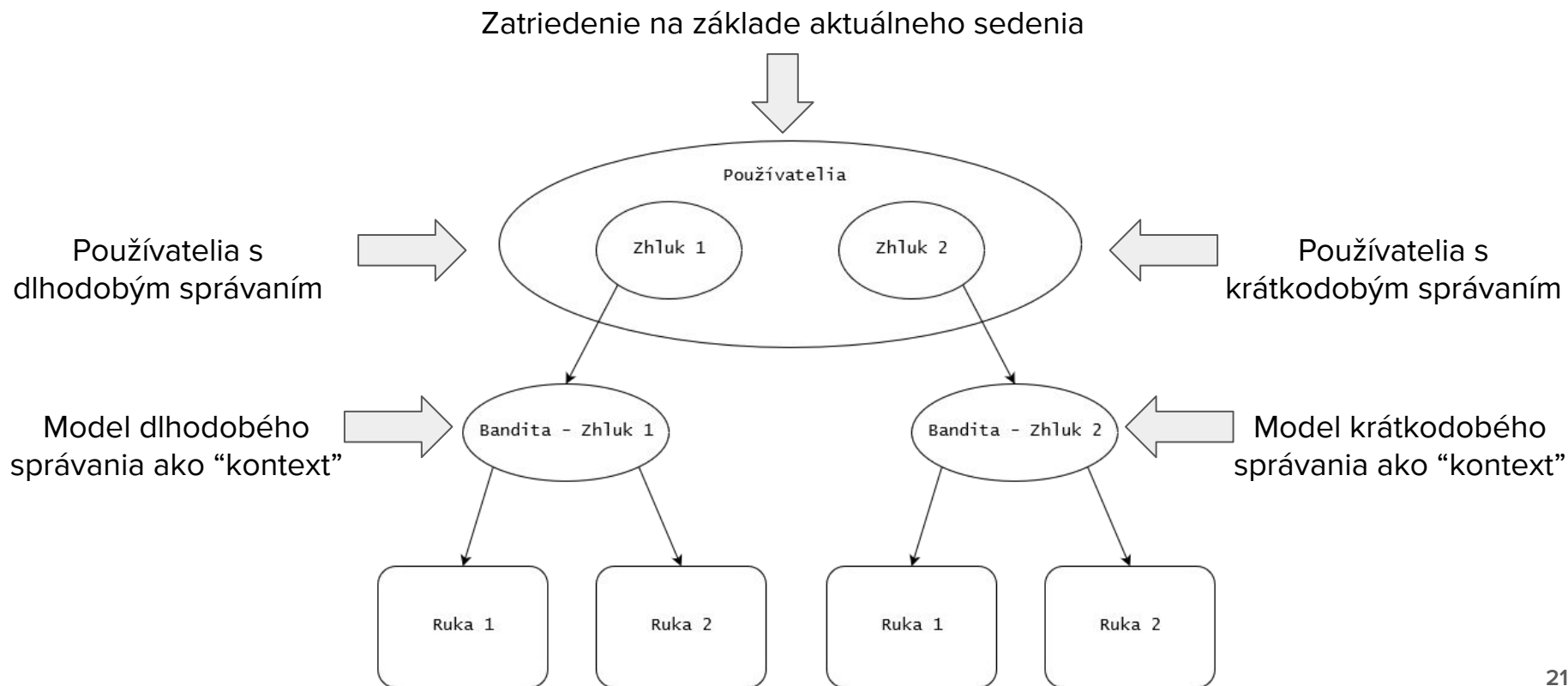
Modelovanie správania používateľa



Modelovanie správania používateľa



Návrh riešenia



Problém

Nedeterministická vektorizácia sekvencií



Iné prístupy vektorizácie sekvencií?

Histogram počtu sedení

