UNSUPERVISED LEARNING

- DOMAIN
- DATA
- WHO
- PURPOSE
- HOW

EXAM

MAT	NE	S	DATE
123	DS	30	3/3/22
123	A	27	1/2/22
124	DS	25	3/3/22

① DATA REPRESENTATION

② SIMILARITY MEASURE

EXAM NAME

STUDENTS	DS	A	ML	DT	STAT
U_1 : 123	30	27	-	30	-
U_2 : 122	25	-	-	18	30
U_3 : 130	29	27		30	
U_5	-	30	30	-	24
U_6	-	25	25	-	26

SIMILARITY MATRIX

	U_1	U_2	U_3	...	U_N
U_1	1	0.2	0.6		
U_2		1			
⋮			1		
U_N					1

THRESHOLD

0.5

STUDENTS	EXAM NAM DS	A	ML	DT	STAT
U_1 : 123	30	27	-	30	- -
→ U_2 : 122	25	-	-	18	30
U_3 : 120	29	27		30	
U_5	-	30	30	-	24
U_6	-	25	25	-	26

THRESHOLD ON MARK DIFFERENCES

5

$SIM(U_1, U_2) = 0.2$

$SIM(U_1, U_3) = 0.6$

PLATFORM

U_1
U_2
U_3

IT_1
IT_2
IT_3
...

USER PROFILING

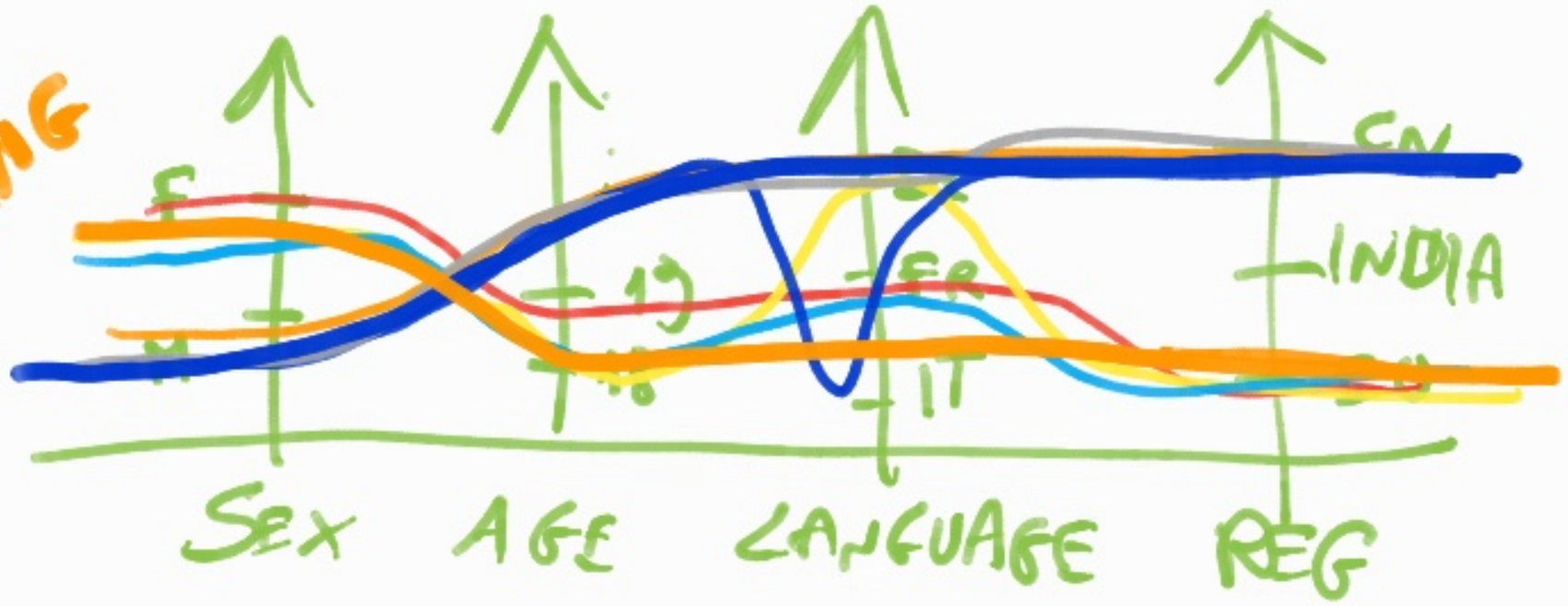
(U_1, IT)

USER

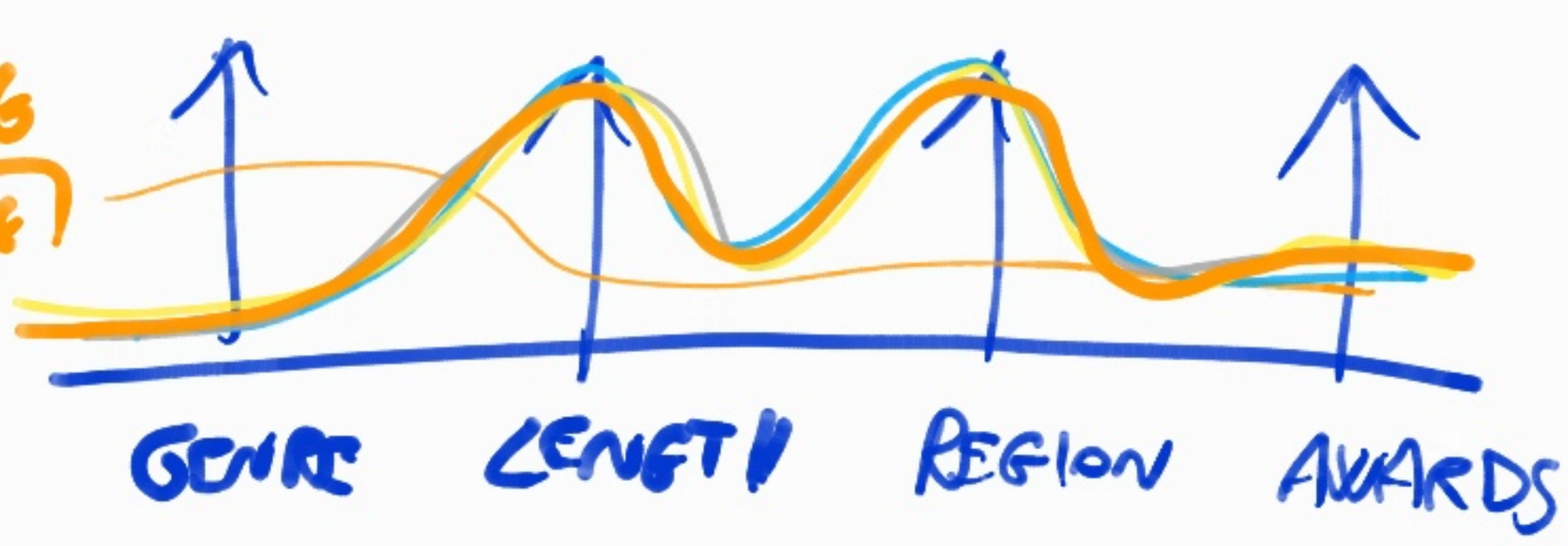
	CF	N	S	SEX	AGE	PLACE	LANGUAGE	..
U_1	AA	MARIO	..	M	20	NA	IT	...
U_2								
U_3								

CLUSTERING

USER PROFILING



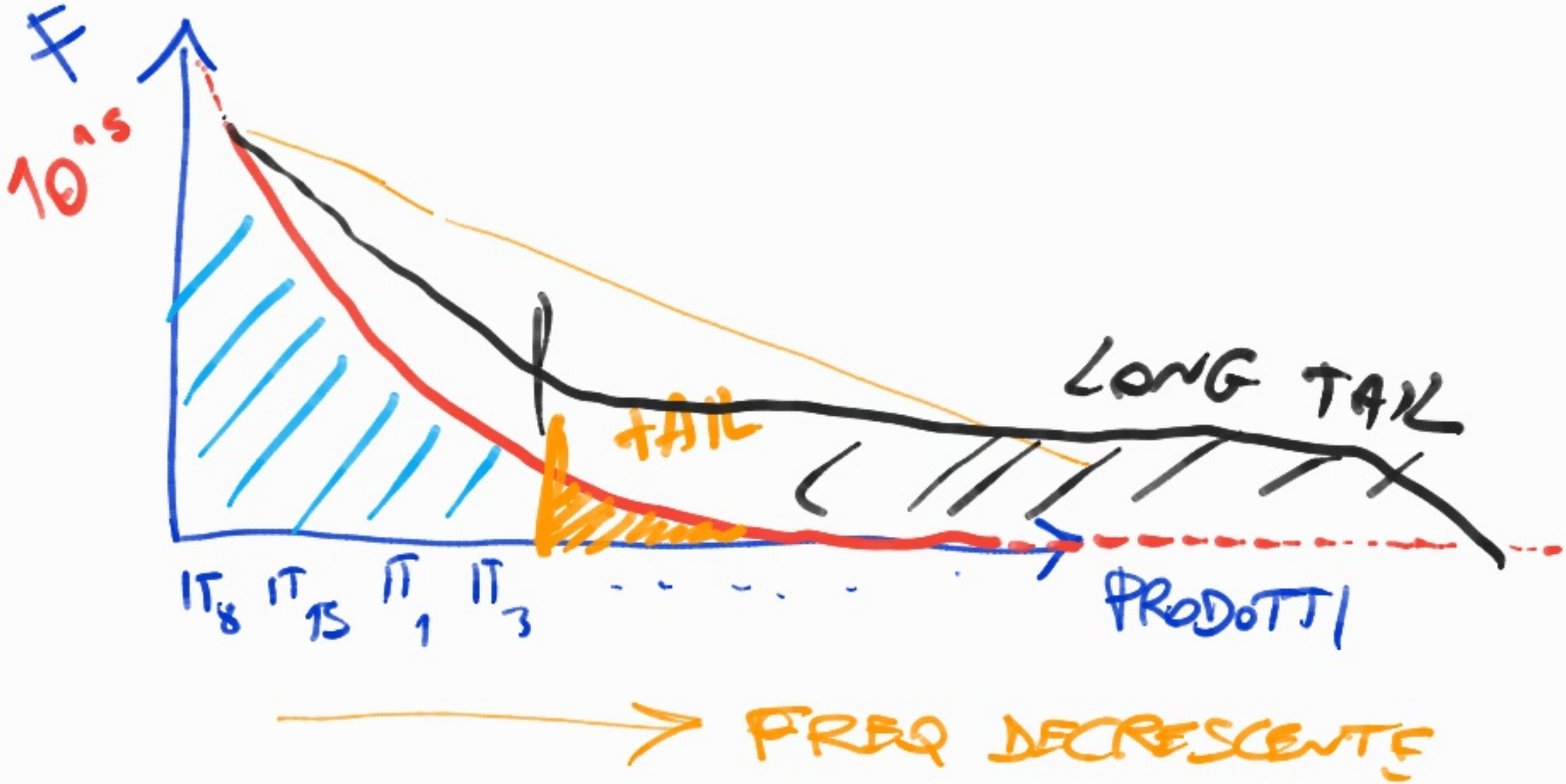
CONTENT PROFILING (CLOUDING)



	IT_1	IT_2	IT_3	...	IT_N
U_1	5	1	3		3
U_2	5	1	-		-
U_3					
⋮					
U_M					

PROBLEMI

1. COLD START
2. SPARSNESS
3. CONTEXT INDEPENDENCE
4. COHERENCE
5. DIVERSITY
6. LONG TAIL


€
10^5
LONG TAIL
TAIL
IT_8 IT_{15} IT_1 IT_3
PRODOTTI
FREQ DECRESCENTE

RECENSIONI

- SENTIMENT ANALISYS
- FAKE REVIEWS
- POLARIZZAZIONE
- BIAS

ASSOCIATION RULES

C_1 (IT_1 , IT_2 , IT_7 , ... , IT_{50})

C_2 (IT_2 , IT_7 , IT_3)

C_3 (IT_2 , IT_3)

⋮

SUPPORTO

CONFIDENZA

$SUPP(IT_2) = \frac{3}{3}$

$CONF = \frac{2}{3}$