

CSX 4207 sec 541

NEWS & MEDIA

Group 4

AHAD RAHMAN
6411247

Thit Lwin Win
Thanh
6540122

Kaung Khant Lin
6540131

Thust
Thongsricharoen
6714508

CSX4207 sec541

NEWS & MEDIA

By Group 4



Contents

1

- INTRODUCTION

2

- OBJECTIVES

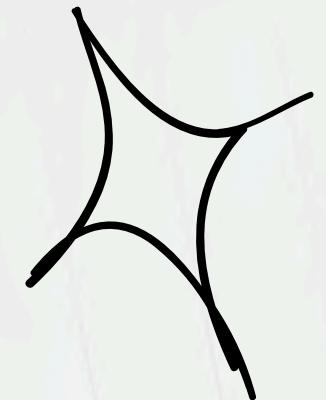
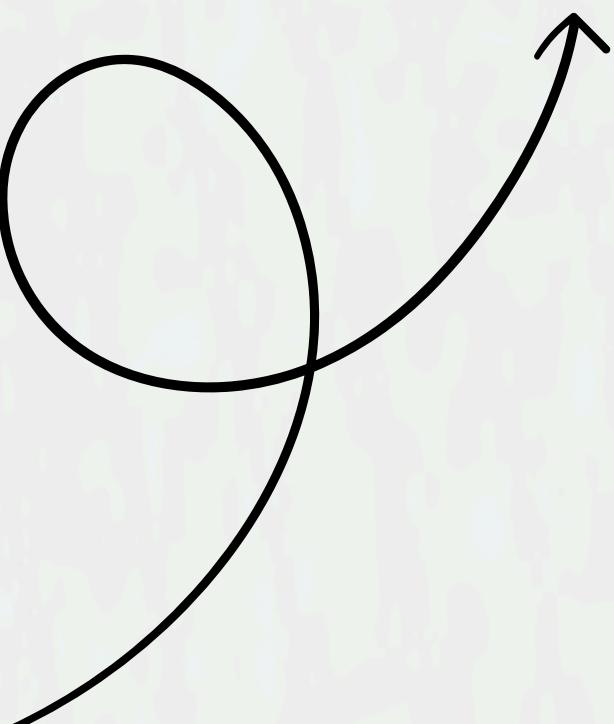
3

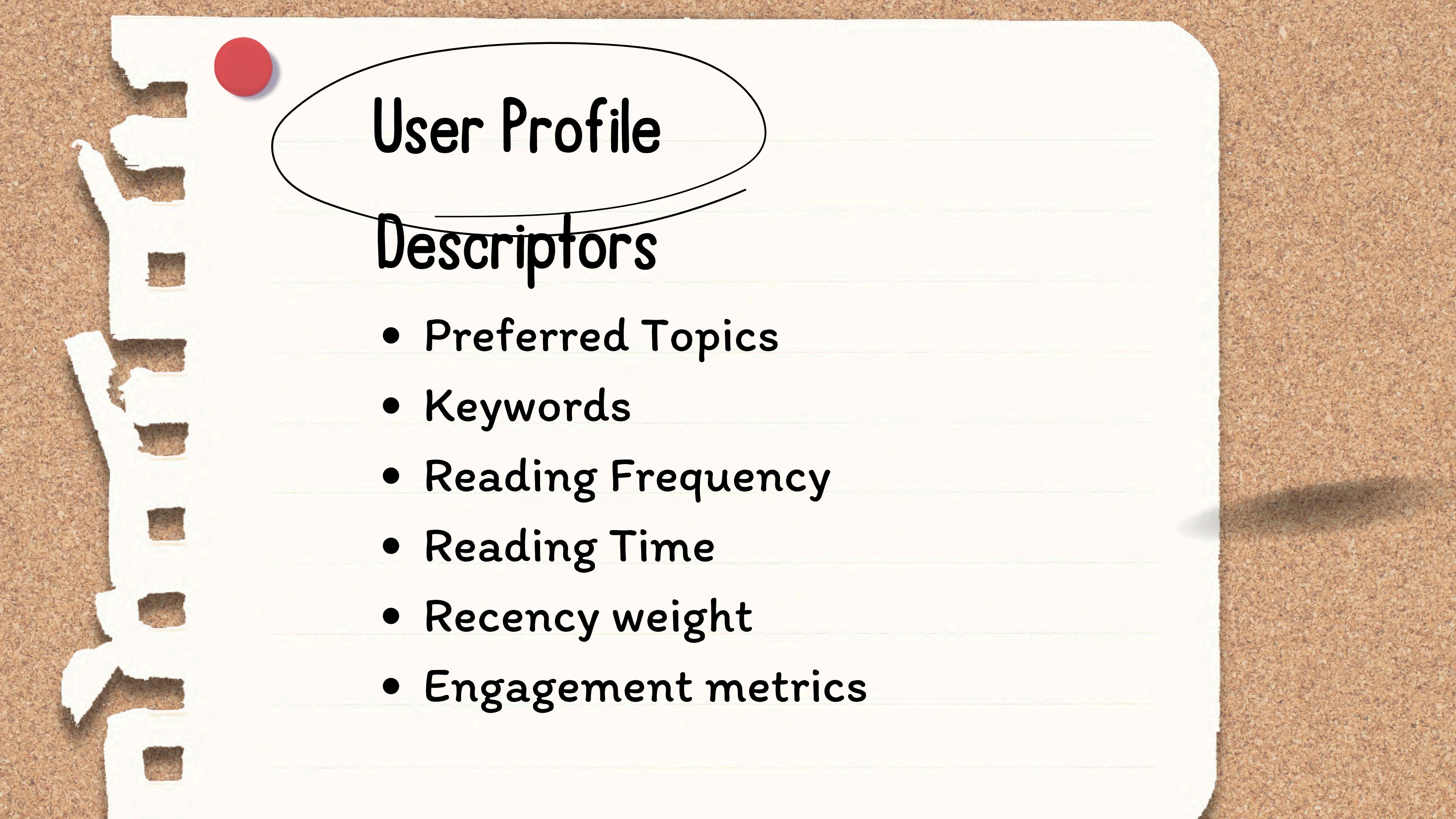
- REFERENCES



Overview

A news recommender system suggests articles to users based on their reading preferences, behaviors, and current trends. Examples include Google News, Apple News, and Flipboard.





User Profile

Descriptors

- Preferred Topics
- Keywords
- Reading Frequency
- Reading Time
- Recency weight
- Engagement metrics

User Preference Vector

- TF-IDF terms: ["Internet", "President", "tennis"]
- Avg Read Time (ART): in minutes, normalized (e.g., divided by 10)
- Recency Weight (RW): recent = close to 1, old = closer to 0



EXAMPLE :FEATURE VECTORS

NEWS 1: THE FUTURE OF SPORTS: HOW 5G WILL TRANSFORM THE SPORTS EXPERIENCE

MAR 30, 2022 | 4 MIN.

- ASSUME DATASET

	Tennis	Internet	AVG Read Time	Recency weight
N1	5	40	0.4	0.007

REF:[HTTPS://WWW.ERICSSON.COM/EN/BLOG/2022/3/FUTURE-OF-SPORTS](https://www.ericsson.com/en/blog/2022/3/future-of-sports)



EXAMPLE :FEATURE VECTORS

NEWS 2 : TRUMP ANNOUNCES A 'MAJOR TRADE DEAL' WITH THE UK
THURSDAY

THU MAY 8, 2025 | 8 MIN.

- ASSUME DATASET

	President	Financial	AVG Read Time	Recency weight
N2	14	30	0.8	0.1

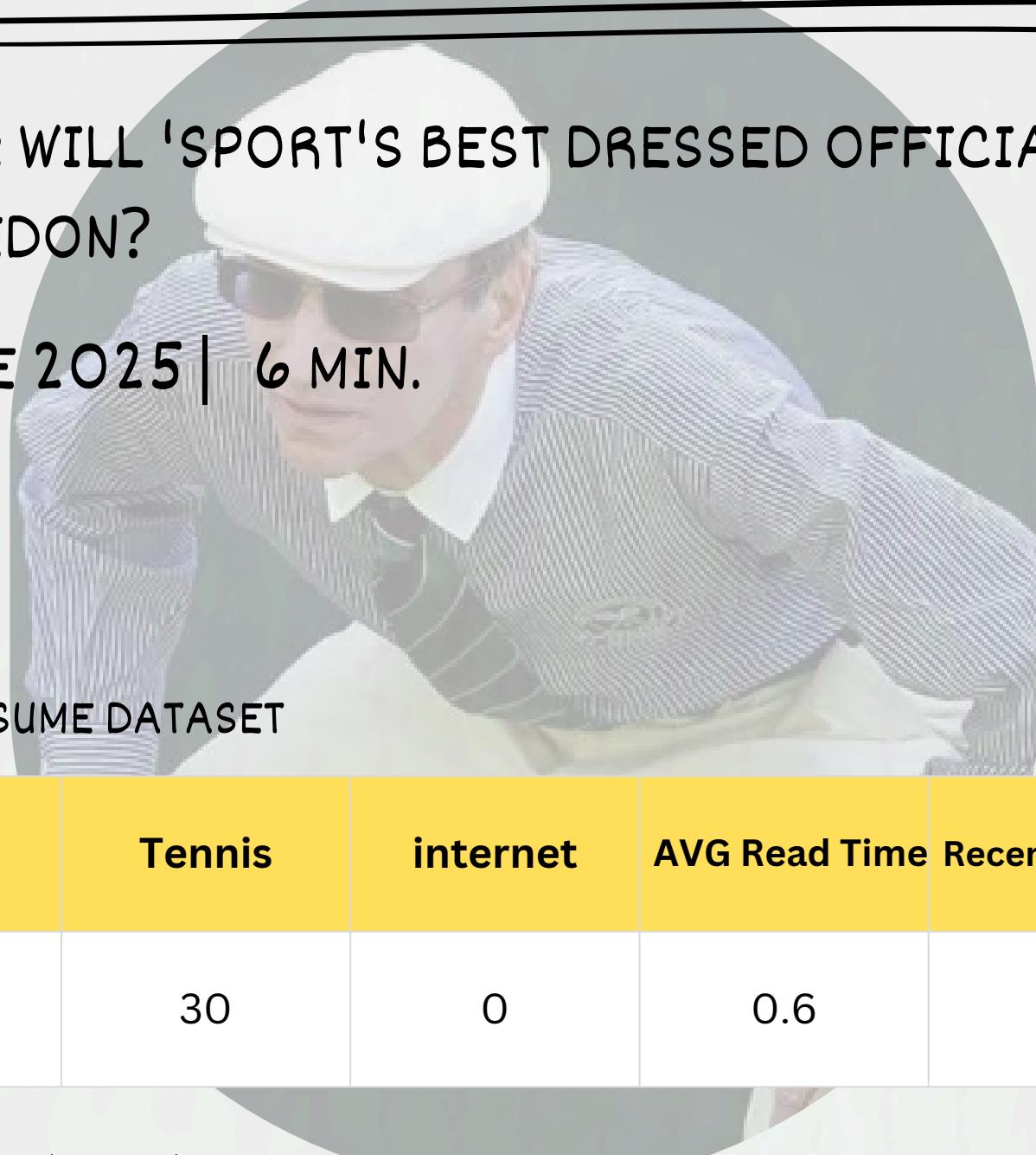
REF:[HTTPS://EDITION.CNN.COM/2025/05/07/BUSINESS/TRADE-DEAL-TRUMP](https://edition.cnn.com/2025/05/07/business/trade-deal-trump)



EXAMPLE :FEATURE VECTORS

NEWS 3: WILL 'SPORT'S BEST DRESSED OFFICIALS' BE MISSED AT WIMBLEDON?

25 JUNE 2025 | 6 MIN.



- ASSUME DATASET

	Tennis	internet	AVG Read Time	Recency weight
N3	30	0	0.6	0.9

REF:[HTTPS://WWW.BBC.COM/SPORT/TENNIS/ARTICLES/CVGNMD8V82MO](https://www.bbc.com/sport/tennis/articles/cvgnmd8v82mo)

REF: [HTTPS://MAKEHEADWAY.COM/BLOG/HOW-LONG-DOES-IT-TAKE-TO-READ-300-PAGE-BOOK/](https://makeheadway.com/blog/how-long-does-it-take-to-read-300-page-book/)



EXAMPLE :FEATURE VECTORS : TF

- ASSSUME(DATASET)

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight
N1	5	0	40	0	0.4	0.007
N2	0	14	0	30	0.8	0.1
N3	30	0	0	0	0.6	0.9



EXAMPLE :FEATURE VECTORS : TF

- $TF(T) = (\text{NUMBER OF TIMES TERM } T \text{ APPEARS IN DOCUMENT D}) / (\text{TOTAL NUMBER OF TERMS IN DOCUMENT D})$

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight
N1	0.111	0	0.888	0	0.4	0.007
N2	0	0.318	0	0.681	0.8	0.1
N3	1	0	0	0	0.6	0.9



EXAMPLE :FEATURE VECTORS : TF

- $\text{IDF} = \log(\text{TOTAL NUMBER OF DOCUMENTS} / \text{NUMBER OF DOCUMENTS CONTAINING TERM T})$

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight
IDF	0.405	1.098	1.098	1.098	N/A	N/A



EXAMPLE :FEATURE VECTORS : TF

- $\text{TF.IDF}(T) = \text{TF}(T) * \text{IDF}(T)$

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight
N1	0.045	0	0.976	0	0.4	0.007
N2	0	0.349	0	0.749	0.8	0.1
N3	0.405	0	0	0	0.6	0.9



USER PREFERENCE VECTOR

ASSUME USER 1 HAS BEEN READ

	Tennis	President	Internet	Financial	Avg ReadTime	Recency weight
N_user1	98	0	18	0	0.5	0.9
N_user2	0	72	36	18	0.6	0.05

SIMPLE UNARY

N_user	98	72	54	18	1.1	0.95
--------	----	----	----	----	-----	------



USER PREFERENCE VECTOR

TF USER 1

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight
N_user	0.404	0.297	0.223	0.07	1.1	0.95

TF-IDF USER 1

N_user	0.164	0.326	0.245	0.08	1.1	0.95
--------	-------	-------	-------	------	-----	------

Top ' k ' Recommendations

- Distance-based similarity
- Cosine similarity

Top 'k' Recommendations

- Distance-Based similarity

Euclidean Distance(N1) : $\sqrt{(SD_Tennis + SD_President + SD_Internet + SD_Financial + SD_AVG\ ReadTime + SD_Recency\ weight)}$

square difference

Euclidean Distance

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight
N1	0.014	0.106	0.549	0.006	0.49	0.889
N2	0.026	0.0005	0.06	0.445	0.09	0.722
N3	0.058	0.106	0.06	0.006	0.25	0.0025

1.428
1.159
0.695

Top 'k' Recommendations

- Distance-Based Similarity

Similarity (N1) = $1 / (1 + \text{Euclidean distance}(N1))$

	Euclidean Distance	Similarity
N1	1.428	0.411
N2	1.159	0.462
N3	0.695	0.589

Rating

- N3 : Will 'sport's best dressed officials'
- N2 : Trump announces a 'major trade deal'
- N1 :The future of sports

Top 'k' Recommendations

cosine_sim(N_user, N1) -

- Cosine Similarity (numerator)

$N_{user} \cdot N1 = (N_{user1} \times N11) + (N_{user2} \times N12) + \dots (N_{usern} \times Nn)$

numerator

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight	Total
N1	0.007	0	0.2393	0	0.44	0.0066	0.693
N2	0	0.1142	0	0.061	0.88	0.095	1.15
N3	0.066	0	0	0	0.66	0.855	1.58

Top 'k' Recommendations

- Cosine Similarity

The norm of a vector: $\| N1 \| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$

	Tennis	President	Internet	Financial	AVG ReadTime	Recency weight	Total
N1	0.002	0	0.9536	0	0.16	0.00049	1.056
N2	0	0.1221	0	0.561	0.64	0.01	1.154
N3	0.1644	0	0	0	0.36	0.81	1.155

N_user	0.0033	0.011	0.00036	4.458	0.625	0.00000625	0.2845
--------	--------	-------	---------	-------	-------	------------	--------

Top 'k' Recommendations

$$\text{cosine_sim}(N_{\text{user}}, N1) = (N_{\text{user}} \cdot N1) / (||N_{\text{user}}|| \times ||N1||)$$

	Norm of a vector	Similarity
N1	1.056	2.307
N2	1.154	3.501
N3	1.551	4.811
N_user	0.284	N/A

Rating

- N3 : Will 'sport's best dressed officials'
- N2 : Trump announces a 'major trade deal'
- N1 :The future of sports:

Summary

1

CONSTRUCTING THE USER
PROFILE

2

REPRESENTING CANDIDATE
ITEMS

3

CALCULATING
SIMILARITY

4

RANKING AND SELECTING
TOP-K



UPDATING THE USER PROFILE VECTOR

- 1. Collect New Interaction Data:
 - Record content topics, read time, and timestamp for user engagements.
- 2. Apply Time Decay Strategy:
 - Option A (Time Window):
 - Discard data outside defined window (e.g., last 24 hrs / 1000 interactions).
 - Re-calculate feature vector based only on remaining data.
 - Option B (Time Decay Function):
 - Calculate $f(t)$ for new data point based on timestamp.
 - Weight new data contribution with $f(t)$.
 - Existing data's influence diminishes as t increases.
 - Feature vector becomes a weighted sum/average of past interactions.
- 3. Recalculate Feature Vector Components:



Reference

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems:
<https://doi.org/10.1109/TKDE.2005.99>
- Content-based and collaborative filtering in news RS:
<https://dl.acm.org/doi/10.1145/3109859.3109873>



Reference

- Li, L., Wang, D., Li, T., Knox, D., & Padmanabhan, B. (2011). SCENE: A Scalable Two-Stage Personalized News Recommendation system:
<https://doi.org/10.1145/2009916.2010034>
- Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based Recommender systems: State of the Art and Trends. In Recommender Systems Handbook
https://doi.org/10.1007/978-0-387-85820-3_3



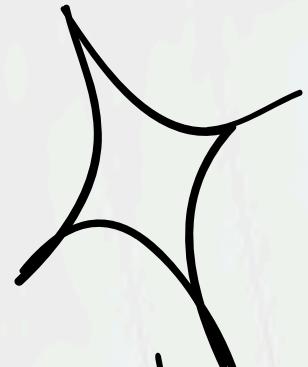
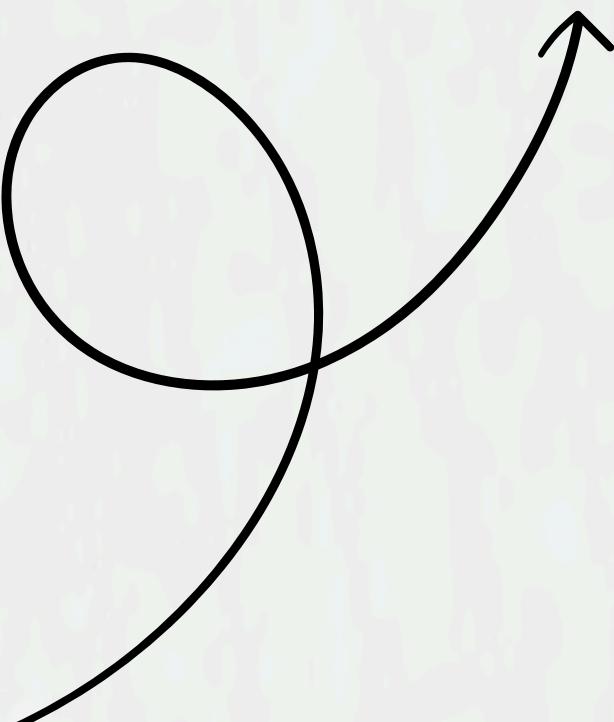
Reference

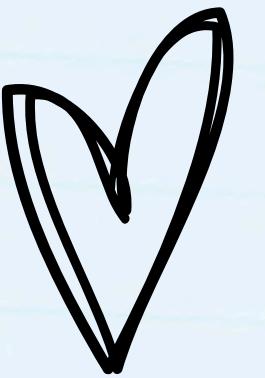
- Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The Adaptive Web* (pp. 325–341). Springer.
https://doi.org/10.1007/978-3-540-72079-9_10
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
<https://nlp.stanford.edu/IR-book/>



Reference

- Koren, Y. (2009). Collaborative filtering with temporal dynamics. Proceedings of the 15th ACM SIGKDD, 447–456.
<https://doi.org/10.1145/1557019.1557072>





Thanks!

for your attention

Presented by Group 4

