

08_recommender

February 27, 2026

1 08 — Recommender

We test 4 configurations side by side: - Word2Vec, no boost - Word2Vec, with recency boost - SBERT-MPNet, no boost - SBERT-MPNet, with recency boost

Note: W2V is retrained on the fly (~10s). SBERT embedding is loaded from file if available, otherwise recomputed (~20min on CPU).

Metrics: sector precision, diversity, mean similarity

```
[20]: import sys
sys.path.append('../src')

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import copy
import os

from user_profile import SYNTHETIC_USERS
from profile_builder import build_profile_vectors, save_profiles
from recommender import Recommender, evaluate_recommendations
from embeddings import train_word2vec, embed_corpus, load_sbert

df_pool      = pd.read_csv('../data/processed/article_pool.csv')
df_stocks    = pd.read_csv('../data/processed/psx_stocks_clean.csv')
df_cnhpsx    = pd.read_csv('../data/processed/cnhpsx_clean.csv')
df_news      = pd.read_csv('../data/processed/pakistan_news_clean.csv')
reading_log  = pd.read_csv('../data/processed/reading_log.csv')

print('Pool:', df_pool.shape)
print('Stocks:', df_stocks.shape)
```

Pool: (34770, 5)

Stocks: (813588, 10)

1.1 Step 1 — Word2Vec embeddings + profiles

W2V is fast to retrain (~10s) so we do it directly here to avoid numpy compatibility issues with saved files.

```
[21]: # Retrain W2V on the same corpus as notebook 05
w2v_corpus = list(df_cnhspx['headline_clean'].dropna()) +
↳ list(df_news['text_clean'].dropna())
print(f'Training corpus: {len(w2v_corpus)} texts')

w2v_model = train_word2vec(
    texts=w2v_corpus,
    vector_size=100,
    window=5,
    min_count=2,
    epochs=10
)

emb_w2v = embed_corpus(
    texts=list(df_pool['text_for_embedding'].fillna('')),
    method='word2vec',
    model=w2v_model
)
print('W2V pool embeddings:', emb_w2v.shape)
```

```
Training corpus: 34770 texts
[Word2Vec] Tokenizing 34770 texts...
[Word2Vec] Training (vector_size=100, window=5, epochs=10)...
[Word2Vec] Vocabulary size: 15243 words
[embed_corpus] Embedding 34770 texts with word2vec...

Word2Vec:
100%|          |
34770/34770 [00:06<00:00, 5436.26it/s]

[embed_corpus] Done. Shape: (34770, 100)
W2V pool embeddings: (34770, 100)
```

```
[22]: # Build W2V user profiles
users_w2v = copy.deepcopy(SYNTHETIC_USERS)
for user in users_w2v:
    user.clicked_articles = reading_log[
        reading_log['user_id'] == user.user_id
    ]['article_id'].tolist()

users_w2v = build_profile_vectors(
    users=users_w2v,
    embeddings=emb_w2v,
    df_tagged=df_pool
)
print('W2V profiles built')
```

```
[Builder] User_1: vector from 10 articles, dim=100
[Builder] User_2: vector from 10 articles, dim=100
```

```
[Builder] User_3: vector from 10 articles, dim=100
[Builder] User_4: vector from 10 articles, dim=100
W2V profiles built
```

1.2 Step 2 — SBERT embeddings + profiles

SBERT takes ~20min on CPU so we try to load the saved file first. If it fails (numpy version mismatch), we recompute and save properly. I screw up on the previous notebooks to save the embeddings

```
[23]: SBERT_PATH = '../data/processed/embeddings_pool_sbert_mpnet.npy'

# Try loading - recompute if anything goes wrong
emb_sbert = None
if os.path.exists(SBERT_PATH):
    try:
        emb_sbert = np.load(SBERT_PATH)
        print('SBERT embeddings loaded from file:', emb_sbert.shape)
    except Exception as e:
        print(f'Could not load saved file ({e}) - recomputing...')

if emb_sbert is None:
    sbert_model = load_sbert('mpnet')
    emb_sbert = embed_corpus(
        texts=list(df_pool['headline'].fillna('')),
        method='sbert',
        model=sbert_model
    )
    # Save with current numpy version so it loads correctly next time
    np.save(SBERT_PATH, emb_sbert)
    print('SBERT embeddings recomputed and saved:', emb_sbert.shape)
```

SBERT embeddings loaded from file: (34770, 768)

```
[24]: # Build SBERT user profiles
users_sbert = copy.deepcopy(SYNTHETIC_USERS)
for user in users_sbert:
    user.clicked_articles = reading_log[
        reading_log['user_id'] == user.user_id
    ]['article_id'].tolist()

users_sbert = build_profile_vectors(
    users=users_sbert,
    embeddings=emb_sbert,
    df_tagged=df_pool
)
print('SBERT profiles built')
```

```
[Builder] User_1: vector from 10 articles, dim=768
```

```
[Builder] User_2: vector from 10 articles, dim=768
[Builder] User_3: vector from 10 articles, dim=768
[Builder] User_4: vector from 10 articles, dim=768
SBERT profiles built
```

1.3 Step 3 — Set up recommenders

```
[25]: rec_w2v = Recommender(emb_w2v, df_pool, df_stocks)
rec_sbert = Recommender(emb_sbert, df_pool, df_stocks)

configs = [
    ('W2V - no boost', rec_w2v, users_w2v, False),
    ('W2V - recency boost', rec_w2v, users_w2v, True),
    ('SBERT - no boost', rec_sbert, users_sbert, False),
    ('SBERT - recency boost', rec_sbert, users_sbert, True),
]

print('Recommenders ready')
```

Recommenders ready

1.4 Step 4 — Run all 4 configurations

```
[26]: print(df_stocks.columns.tolist())
print(df_stocks.head(2))
```

```
['date', 'symbol', 'ldcp', 'open', 'high', 'low', 'close', 'change',
'change_pct', 'volume']
```

	date	symbol	ldcp	open	high	low	close	change	change_pct	\
0	2018-09-04	786	2.999	3.99	3.99	3.99	3.99	0.991	33.044348	
1	2018-09-05	786	3.990	4.99	4.99	4.99	4.99	1.000	25.062657	


```
volume
0    21000
1   150500
```

```
[27]: import importlib, scorer, recommender
importlib.reload(scorer)
importlib.reload(recommender)
from recommender import Recommender, evaluate_recommendations

# Recreate recommenders with reloaded module
rec_w2v = Recommender(emb_w2v, df_pool, df_stocks)
rec_sbert = Recommender(emb_sbert, df_pool, df_stocks)
print('done')
```

done

```
[28]: TOP_K = 10
all_recs = {}

for config_name, rec, users, boost in configs:
    all_recs[config_name] = {}
    for user in users:
        recs = rec.recommend(user, top_k=TOP_K, use_recency_boost=boost)
        all_recs[config_name][user.user_id] = recs
    print(f'{config_name} - done')
```

W2V - no boost - done
W2V - recency boost - done
SBERT - no boost - done
SBERT - recency boost - done

1.5 Step 5 — Inspect recommendations per user

```
[29]: for config_name, user_recs in all_recs.items():
    print(f'\n{"="*60}')
    print(f'CONFIG: {config_name}')
    print(f'{"="*60}')
    for user_id, recs in user_recs.items():
        user = next(u for u in SYNTHETIC_USERS if u.user_id == user_id)
        print(f'\n {user_id} ({user.sector} / {user.sub_focus}):')
        print(
            recs[['rank', 'headline', 'primary_tag', 'similarity', 'final_score']].
            to_string(index=False))
```

```
=====
CONFIG: W2V - no boost
=====

User_1 (Construction / cement-heavy):
rank
headline primary_tag similarity final_score
1 ['PSO's poor assets management causes drop in market share: audit'] Energy 0.957947 0.957947
2 ['Silkbank's Continues Growth Ascend - 2017 Half Yearly Performance', 'Silkbank concludes Rs2bln TFC issue'] Other 0.954764 0.954764
3 ['OGDC may go for buying foreign assets worth over $500 million', 'PSO improves safety steps'] Energy 0.954127 0.954127
4 ['Shareholders laud PSO's performance'] Energy 0.953219 0.953219
5 ['MCB ATMs to remain operational during Eid holidays'] Other 0.953093 0.953093
6 ['Silkbank Continues Ascending Towards Excellence'] Other 0.952747 0.952747
```

7	['LSE delists K-Electric's Sukuk-1']	Other	0.950889	0.950889
8	Insurance expresses interest in Alico']	Other	0.948716	0.948716
9	foreign assets worth over \$500 million']	Energy	0.948476	0.948476
10	acquiring 40pc stake in Pakistan stocks']	Other	0.947369	0.947369

User_2 (Construction / infrastructure-heavy):

rank	headline	primary_tag	similarity	final_score
1	Sindh CM-governor meeting seen as positive sign for city's development projects	Other	0.971471	0.971471
2	Newspaper	Other	0.971330	0.971330
3	teachers	Other	0.964911	0.964911
4	Pakistan	Other	0.964346	0.964346
5	Newspaper	Other	0.963550	0.963550
6	Pakistan	Other	0.963189	0.963189
7	'Night-time is particularly harsh for people living under an open sky'	Other	0.962649	0.962649
8	Pakistan	Other	0.962222	0.962222
9	border	Other	0.962110	0.962110
10	projects	Other	0.961632	0.961632

User_3 (Banking / None):

rank	headline	primary_tag	similarity	final_score
1	introduces cashless payment']	Banking	0.993285	0.993285
2	accommodate retrenched employees']	Banking	0.991472	0.991472
3	per unit for Income Fund']	Banking	0.989064	0.989064
4	transactions by mid-July: Dar']	Banking	0.988590	0.988590
5	adopts Asset Manager Code']	Banking	0.987374	0.987374
6	to be appointed by 22nd']	Banking	0.987311	0.987311

7					['NBP
to reach flood victims']	Banking	0.986919	0.986919		
8					['HBL Income Fund declares
third interim dividend']	Banking	0.986790	0.986790		
9					['HBL,
MasterCard announce partnership']	Banking	0.986158	0.986158		
10					['HBL
grants Rs10m to Labard']	Banking	0.985919	0.985919		

User_4 (Energy / None):

rank	headline	primary_tag	similarity	final_score	
1	['Govt recovering oil subsidies: Shell', 'Shell declares loss in Q2']				
	Other	0.984806	0.984806		
2	520,000 tonnes of fuel oil', 'SSGC issues Rs86,808 bill to residential consumer']	Energy	0.984247	0.984247	['PSO buys
3	report', 'SSGC wins Rs4bn financing facility', 'PSO seeks 780,000T fuel oil']				['Oil prices climb ahead of US energy
	Energy	0.983029	0.983029		
4	['PSO keen to increase revenues from LPG retail, lubricants']	Energy			
	0.982115	0.982115			
5	['PSO plans to extend E-10 fuel in 100 outlets']	Energy		0.981897	
	0.981897				
6	['PSO rejects \$30m fuel oil cargo, returns ship']	Energy		0.980370	
	0.980370				
7	['FBR detects Rs1.7 billion tax evasion by oil refinery']	Energy			
	0.980081	0.980081			
8	['Obama defends energy policy amid high gas prices', 'CCoP approves sale of 2.5pc govt-owned PPL shares', 'Shell signs China's first shale gas deal']				
	Energy	0.979964	0.979964		
9	palm oil development', 'SSGC team in BD to learn about reducing gas losses']				['BoI seeks Malaysian investment for
	Other	0.979962	0.979962		
10	['PSO puts cap on fuel supply to PIA', 'SSGC recovers Rs194m']	Energy			
	0.979766	0.979766			

=====
 CONFIG: W2V - recency boost
 =====

User_1 (Construction / cement-heavy):

rank	headline	primary_tag	similarity	final_score	
1					['Mirza becomes PPL acting
MD']	Energy	0.896730	1.033209		
2					['KEL, Dadabhoy Institute
collaborate']	Other	0.923414	1.026612		
3					['Hascol reaches settlement with Mena
Energy']	Other	0.936395	1.020181		
4					['Local gross refinery margins likely to fall by 7
percent']	Energy	0.916925	1.020122		
5					['Hascol EPS at
Rs2.85']	Other	0.935733	1.019519		
6					['Deteriorating law & order situation breaks bullish spell on
KSE']	Other	0.910898	1.018685		
7					Number of tax filers rises by 40pc as deadline
looms	Other	0.897376	1.018196		
8					['LSE delists K-Electric's
Sukuk-1']	Other	0.950889	1.017027		
9					['PSO vying to become single largest shareholder in Pakistan Refinery
Ltd']	Energy	0.880032	1.016512		
10					['KSE falls as investors become cautious on political
situation']	Other	0.879492	1.015972		

User_2 (Construction / infrastructure-heavy):

rank	headline	primary_tag	similarity	final_score	
1					FBR extends deadline to file tax returns
till Feb 28	Other	0.946011	1.066830		
2					CNG association wants filling stations to stay
open 24/7	Other	0.942388	1.063207		
3					Murad orders hiring of aircraft to eradicate locust swarms in deserts
across Sindh	Other	0.924549	1.060987		
4					Croatia becomes latest country to move troops out of Iraq amid soaring
tensions	Other	0.922515	1.058995		
5					Power distribution companies start filing petitions for increase
in tariff	Other	0.932588	1.053407		
6					Dawn CEO Hameed Haroon sends legal notice to
filmmaker Jami	Other	0.922669	1.043488		
7					Pakistan to become self-sufficient in four antisera
by June	Other	0.906024	1.042504		
8					Schools in KP, Punjab to remain shut as country braces for another
winter spell	Other	0.934330	1.042116		
9					Filmmaker Jami sends legal response to
Hameed Haroon	Other	0.919010	1.039830		
10					Cloud cooking land: Indian housewives become gig
economy chefs	Other	0.895523	1.032003		

User_3 (Banking / None):

rank	headline	primary_tag	similarity	final_score
1	['Mirza becomes PPL acting MD']	Energy	0.951487	1.087967
2	['Extension of PSO chief term becomes mystery']	Energy	0.948382	1.084861
3	['NBP Serving the Nation in Thar Desert']	Banking	0.946235	1.082673
4	['KEL launches net metering service']	Other	0.977341	1.080538
5	['Govt to establish 100 LPG filling stations at PSO outlets', 'Shell suffers loss of Rs223.64m', 'PSO to establish fuel farm at Benazir Bhutto airport']	Energy	0.951362	1.072182
6	['PSO management says selloff unlikely', 'PSO introduces SMS alert service', 'HBL will install over 50 ATMs at Shell fuel stations']	Banking	0.964360	1.067558
7	['KEL receives CSR Award']	Other	0.963330	1.066528
8	['Govt to establish 100 LPG filling stations at PSO outlets', 'PSO to establish fuel farm at Benazir Bhutto airport']	Energy	0.944903	1.065722
9	['Shell Philippines files \$629mln IPO\ may be country's largest in 2016']	Other	0.939614	1.060433
10	['State Bank likely to hike rate as inflation concerns weigh', 'UBL to buy Khushhali Bank shares', 'NBP Commenced Foree Cash & Foree Transfer Home Remittances With Ria Financial Services']	Banking	0.957174	1.060372

User_4 (Energy / None):

rank	headline	primary_tag	similarity	final_score
1	['Govt to establish 100 LPG filling stations at PSO outlets', 'PSO to establish fuel farm at Benazir Bhutto airport']	Energy	0.966061	1.086880
2	['PSO aims to become global energy company', 'PSO assures commitment to development in Balochistan']	Energy	0.943985	1.080465
3	['Govt to establish 100 LPG filling stations at PSO outlets', 'Shell suffers loss of Rs223.64m', 'PSO to establish fuel farm at Benazir Bhutto airport']	Energy	0.957844	1.078663
4	['Oil prices\ rebound', 'Shell Philippines files \$629mln IPO\ may be country's largest in 2016']	Other	0.952685	1.073504
5	['Supplies likely to choke as PIA owes PSO Rs14 bn']	Energy	0.969208	1.072406

6	['PSO management says selloff unlikely', 'PSO introduces SMS alert service', 'HBL will install over 50 ATMs at Shell fuel stations']		
	Banking	0.955165	1.058362
7	['Pakistan to benefit from likely drop in oil prices', 'SSGC to get 100mmcf from Kunnar-Passaki']		
	Other	0.949039	1.052236
8	['PSO management says selloff unlikely', 'PSO introduces SMS alert service']		
	Energy	0.948594	1.051791
9	['Vitol raises stake in Hascol to 25pc']		
	Other	0.964535	1.048321
10	['Hascol Petroleum acquires 11.4pc shares of PRL']		
	Other	0.964441	1.048227

CONFIG: SBERT - no boost

User_1 (Construction / cement-heavy):				
rank	headline	primary_tag	similarity	final_score
1	Pakistan	Other	0.788898	0.788898
2	Pakistan	Other	0.788898	0.788898
3	Pakistan	Construction	0.788898	0.788898
4	Pakistan	Other	0.788898	0.788898
5	Pakistan	Other	0.788898	0.788898
6	Pakistan	Other	0.788898	0.788898
7	Pakistan	Other	0.788898	0.788898
8	Pakistan	Other	0.788898	0.788898
9	Pakistan	Other	0.788898	0.788898
10	Pakistan	Other	0.788898	0.788898

User_2 (Construction / infrastructure-heavy):				
rank	headline	primary_tag	similarity	final_score
1	Pakistan	Other	0.795416	0.795416
2	Pakistan	Other	0.795416	0.795416
3	Pakistan	Other	0.795416	0.795416
4	Pakistan	Other	0.795416	0.795416
5	Pakistan	Other	0.795416	0.795416
6	Pakistan	Other	0.795416	0.795416
7	Pakistan	Other	0.795416	0.795416
8	Pakistan	Other	0.795416	0.795416
9	Pakistan	Construction	0.795416	0.795416
10	Pakistan	Other	0.795416	0.795416

```
User_3 (Banking / None):
rank                                     headline
primary tag  similarity  final_score
```

1			['NBP earns Rs18.2bn profit']
Banking	0.829100	0.829100	
2			['NBP earns Rs4.2bn net profit']
Banking	0.814201	0.814201	
3			['NBP posts profit of Rs5.5bn']
Banking	0.812567	0.812567	
4			['NBP disburses Rs7.1 billion']
Banking	0.801543	0.801543	
5			['NBP gives Rs7.5 dividend, profit reaches Rs17.7bn']
Banking	0.788068	0.788068	
6			['NBP earns Rs8.20bn profits', 'NBP's entire branch network goes online']
Banking	0.787973	0.787973	
7			['NBP's profit doubles to Rs6.06bln in April-June']
Banking	0.784394	0.784394	
8			['NBP to start due diligence of Burj Bank']
Banking	0.781016	0.781016	
9			['MCB Bank posts Q1 profit of Rs7.9bln', 'Engro Foods, NBP sign MoU']
Banking	0.776461	0.776461	
10			['NBP share sale next fiscal']
Banking	0.776022	0.776022	

User_4 (Energy / None):

rank	headline	primary_tag	similarity	final_score
1				['PSO posts Rs5bn profit in first half', 'PSO buys up to 780,000T fuel oil']
		Energy	0.830479	0.830479
2				['PSO buys 520,000 tonnes of fuel oil', 'SSGC issues Rs86,808 bill to residential consumer']
		Energy	0.798750	0.798750
3				['POL imports rise due to energy crisis', 'PSO earns profits of Rs6.3 billion']
		Energy	0.794886	0.794886
4				['Oil prices steady after nearing \$124', 'PSO bankrolls NICH']
		Energy	0.790129	0.790129
5				['PSO buys 520,000 tonnes of fuel oil']
		Energy	0.789124	0.789124
6				['PSO in process of buying 30pc shares of Pakistan Refinery']
		Energy	0.783322	0.783322
7				['Oil prices down', 'PSO to buy 63 million right shares offered by PRL']
		Energy	0.781043	0.781043
8				['PPL discovers hydrocarbon in Sanghar district', 'Shareholders laud PSO's performance']
		Energy	0.771086	0.771086
9				['PSO profit up 304pc', 'No payout for POL shareholders']
		Energy	0.770224	0.770224
10				['Oil prices slump below \$55', 'PSO ready for privatisation: PM']
		Energy	0.768864	0.768864

=====
CONFIG: SBERT - recency boost
=====

User_1 (Construction / cement-heavy):

rank	headline	primary_tag	similarity	final_score
1	Pakistan	Other	0.788898	0.788898
2	Pakistan	Other	0.788898	0.788898
3	Pakistan	Other	0.788898	0.788898
4	Pakistan	Other	0.788898	0.788898
5	Pakistan	Other	0.788898	0.788898
6	Pakistan	Other	0.788898	0.788898
7	Pakistan	Other	0.788898	0.788898
8	Pakistan	Other	0.788898	0.788898
9	Pakistan	Other	0.788898	0.788898
10	Pakistan	Other	0.788898	0.788898

User_2 (Construction / infrastructure-heavy):

rank	headline	primary_tag	similarity	final_score
1	Pakistan	Other	0.795416	0.795416
2	Pakistan	Other	0.795416	0.795416
3	Pakistan	Other	0.795416	0.795416
4	Pakistan	Other	0.795416	0.795416
5	Pakistan	Other	0.795416	0.795416
6	Pakistan	Other	0.795416	0.795416
7	Pakistan	Other	0.795416	0.795416
8	Pakistan	Other	0.795416	0.795416
9	Pakistan	Other	0.795416	0.795416
10	Pakistan	Other	0.795416	0.795416

User_3 (Banking / None):

rank			headline
primary_tag	similarity	final_score	
1			['NBP earns Rs18.2bn profit']
Banking	0.829100	0.858209	
2			['NBP earns Rs4.2bn net profit']
Banking	0.814201	0.843309	
3			['NBP posts profit of Rs5.5bn']
Banking	0.812567	0.841676	
4			['NBP disburses Rs7.1 billion']
Banking	0.801543	0.830651	
5			['NBP aims to be top bank of region']
Banking	0.768110	0.823865	
6			['NBP gives Rs7.5 dividend, profit reaches Rs17.7bn']
Banking	0.788068	0.817177	
7	['NBP earns Rs8.20bn profits', 'NBP's entire branch network goes online']		
Banking	0.787973	0.817082	
8	['UBL announces 10pc cash dividend', 'NBP aims to be top bank of region']		
Banking	0.758388	0.814143	
9			['NBP's profit doubles to Rs6.06bln in April-June']
Banking	0.784394	0.813503	

```

10                                     ['NBP staff seek demands fulfillment']
Banking      0.690484      0.811303

User_4 (Energy / None):
rank
headline primary_tag similarity final_score
1
['PSO vying to become single largest shareholder in Pakistan Refinery Ltd']
Energy      0.751878      0.888357
2                                     ['PSO aims to become
global energy company', 'PSO assures commitment to development in Balochistan']
Energy      0.709375      0.845855
3
['PSO posts Rs5bn profit in first half', 'PSO buys up to 780,000T fuel oil']
Energy      0.830479      0.835477
4 ['Govt to establish 100 LPG filling stations at PSO outlets', 'Shell
suffers loss of Rs223.64m', 'PSO to establish fuel farm at Benazir Bhutto
airport']      Energy      0.713429      0.834249
5                                     ['PPL
discovers hydrocarbon in Sanghar district', 'Shareholders laud PSO's
performance']      Energy      0.771086      0.824257
6                                     ['PSO buys
520,000 tonnes of fuel oil', 'SSGC issues Rs86,808 bill to residential
consumer']      Energy      0.798750      0.809287
7                                     ['Oil
prices\rebound', 'Shell Philippines files $629mln IPO\may be country's
largest in 2016']      Other      0.684695      0.805514
8
['POL imports rise due to energy crisis', 'PSO earns profits of Rs6.3 billion']
Energy      0.794886      0.803739
9
['PSO seeks 143,000 cubic meters of LNG for July delivery']      Energy
0.733168      0.799306
10                                     ['Govt to establish 100 LPG filling
stations at PSO outlets', 'PSO to establish fuel farm at Benazir Bhutto
airport']      Energy      0.676886      0.797705

```

1.6 Step 6 — Evaluate and compare

```

[30]: eval_rows = []

for config_name, user_recs in all_recs.items():
    users = users_w2v if 'W2V' in config_name else users_sbert
    for user in users:
        metrics = evaluate_recommendations(user_recs[user.user_id], user)
        metrics['config'] = config_name
        eval_rows.append(metrics)

```

```
df_eval = pd.DataFrame(eval_rows)
print(df_eval[['config', 'user_id', 'sector', 'sector_precision', 'diversity', 'mean_similarity']]).
↳to_string(index=False))
```

	config	user_id	sector	sector_precision	diversity	mean_similarity
0.9521	W2V - no boost	User_1	Construction	0.0	2	
0.9647	W2V - no boost	User_2	Construction	0.0	1	
0.9883	W2V - no boost	User_3	Banking	1.0	1	
0.9816	W2V - no boost	User_4	Energy	0.8	2	
0.9128	W2V - recency boost	User_1	Construction	0.0	2	
0.9246	W2V - recency boost	User_2	Construction	0.0	1	
0.9544	W2V - recency boost	User_3	Banking	0.3	3	
0.9572	W2V - recency boost	User_4	Energy	0.5	3	
0.7889	SBERT - no boost	User_1	Construction	0.1	2	
0.7954	SBERT - no boost	User_2	Construction	0.1	2	
0.7951	SBERT - no boost	User_3	Banking	1.0	1	
0.7878	SBERT - no boost	User_4	Energy	1.0	1	
0.7889	SBERT - recency boost	User_1	Construction	0.0	1	
0.7954	SBERT - recency boost	User_2	Construction	0.0	1	
0.7835	SBERT - recency boost	User_3	Banking	1.0	1	
0.7465	SBERT - recency boost	User_4	Energy	0.9	2	

```
[18]: # Summary table - mean per config
summary = df_eval.
↳groupby('config')[['sector_precision', 'diversity', 'mean_similarity', 'mean_final_score']].
↳mean().round(4)
summary = summary.reindex(['W2V - no boost', 'W2V - recency boost', 'SBERT - no_
↳boost', 'SBERT - recency boost'])
```

```

print('=== Mean Metrics per Configuration ===')
print(summary.to_string())
print('\n→ Best sector precision:', summary['sector_precision'].idxmax())
print('→ Best diversity:          ', summary['diversity'].idxmax())

```

=== Mean Metrics per Configuration ===

	sector_precision	diversity	mean_similarity
mean_final_score			
config			
W2V - no boost	0.450	1.50	0.9702
0.9702			
W2V - recency boost	0.175	2.25	0.9375
1.0530			
SBERT - no boost	0.550	1.50	0.7918
0.7918			
SBERT - recency boost	0.475	1.25	0.7786
0.8090			

→ Best sector precision: SBERT - no boost

→ Best diversity: W2V - recency boost

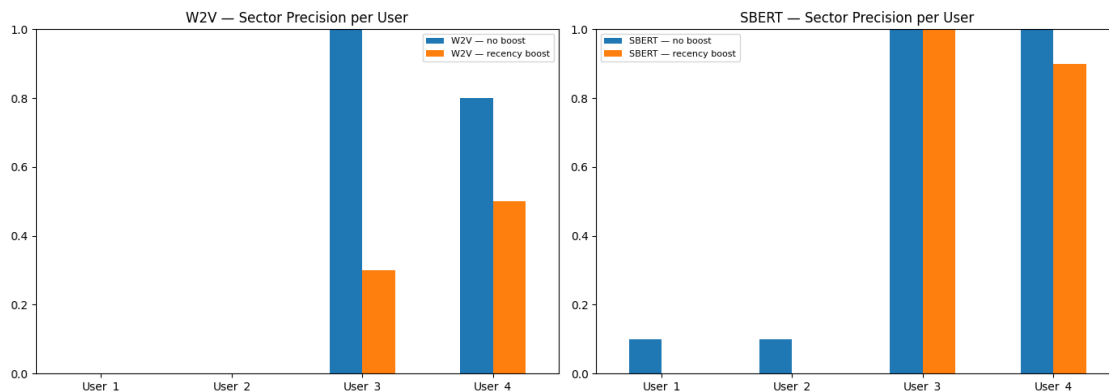
```

[31]: # Sector precision per user - W2V vs SBERT
fig, axes = plt.subplots(1, 2, figsize=(14, 5))

for ax, model in [(axes[0], 'W2V'), (axes[1], 'SBERT')]:
    subset = df_eval[df_eval['config'].str.startswith(model)]
    pivot = subset.pivot(index='user_id', columns='config',
        values='sector_precision')
    pivot.plot(kind='bar', ax=ax, title=f'{model} - Sector Precision per User',
        ylim=(0, 1))
    ax.set_xlabel('')
    ax.tick_params(axis='x', rotation=0)
    ax.legend(fontsize=8)

plt.tight_layout()
plt.show()

```



1.7 Conclusions

- ☐ W2V no boost — mean sector precision: ...
- ☐ W2V recency boost — mean sector precision: ...
- ☐ SBERT no boost — mean sector precision: ...
- ☐ SBERT recency boost — mean sector precision: ...
- ☐ Best overall config: ...
- ☐ Does recency boost improve results: yes / no / marginally
- ☐ Ready for evaluation notebook (09): yes/no