

08_recommender

February 28, 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

```
[2]: 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: (14213, 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.

```
[3]: # Retrain W2V on the same corpus as notebook 05
w2v_corpus = list(df_cnhpsx['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 14213 texts with word2vec...

Word2Vec:
100%| |
14213/14213 [00:01<00:00, 7605.24it/s]

[embed_corpus] Done. Shape: (14213, 100)
W2V pool embeddings: (14213, 100)

```
[4]: # 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

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

sbert_model = load_sbert('mpnet')
emb_sbert = embed_corpus(
    texts=list(df_pool['headline'].fillna('')),
    method='sbert',
    model=sbert_model
)
np.save(SBERT_PATH, emb_sbert)
print('SBERT embeddings recomputed and saved:', emb_sbert.shape)
```

```
[SBERT] Loading: all-mpnet-base-v2...
```

Warning: You are sending unauthenticated requests to the HF Hub. Please set a HF_TOKEN to enable higher rate limits and faster downloads.

Loading weights: 0% | 0/199 [00:00<?, ?it/s]

```
MPNetModel LOAD REPORT from: sentence-transformers/all-mpnet-base-v2
Key           | Status   | |
-----+-----+---+
embeddings.position_ids | UNEXPECTED | |
```

Notes:

- UNEXPECTED : can be ignored when loading from different task/architecture; not ok if you expect identical arch.

```
[SBERT] Ready - embedding dim: 768
```

```
[embed_corpus] Embedding 14213 texts with sbert...
```

Batches: 0% | 0/223 [00:00<?, ?it/s]

```
[embed_corpus] Done. Shape: (14213, 768)
```

```
SBERT embeddings recomputed and saved: (14213, 768)
```

```
[6]: # 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

```

[7]: 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

```

[8]: 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

```

```

[9]: 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

```

[10]: 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

```

[11]: 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):
                                         headline
rank
primary_tag similarity final_score
1                  ['K-Electric shows interests in buying stakes in FESCO']
Other      0.973363   0.973363
2                  ['Investors apprised of PSO's profitable growth model']

```

Energy	0.973260	0.973260	
3			['PSO, FFBL earning hopes keep market alive']
Energy	0.960714	0.960714	
4			['PSO's poor assets management causes drop in market share: audit']
Energy	0.960011	0.960011	
5			['TPL Insurance expresses interest in Alico']
Other	0.959858	0.959858	
6			['KSE snaps losing streak as WB announces flood support']
Other	0.959780	0.959780	
7			['KSE seen rising as brokers set to meet President']
Other	0.958585	0.958585	
8			['Interim regulatory changes cause temporary recovery on KSE']
Other	0.958002	0.958002	
9			['KSE rises as investors expect MTS approval without changes']
Other	0.957921	0.957921	
10			['KSE takes a deep plunge as Pak-US row intensifies']
Other	0.957860	0.957860	

User_2 (Construction / infrastructure-heavy):

rank	headline	primary_tag	similarity	final_score	
1	model']	Energy	0.973953	0.973953	['Investors apprised of PSO's profitable growth
2	KSE-100']	Other	0.973545	0.973545	['Expectation of strong earnings lifts up
3	trading']	Other	0.970001	0.970001	['Security concerns keep investors wary,KSE dips amidst dull
4	holidays']	Other	0.968397	0.968397	['KSE-100 off to slow start after Eid
5	FESCO']	Other	0.966831	0.966831	['K-Electric shows interests in buying stakes in
6	audit']	Energy	0.965335	0.965335	['PSO's poor assets management causes drop in market share:
7	Excellence']	Other	0.963590	0.963590	['Silkbank Continues Ascending Towards
8	support']	Other	0.961976	0.961976	['KSE-100 continues to move upward on institutional
9	economy']	Other	0.961570	0.961570	['KSE drops slightly as floods hurt
10	down']	Energy	0.960915	0.960915	['Abandoning of additives in diesel,petrol brings PSO's market share

User_3 (Banking / None):

rank	headline	primary_tag	similarity	final_score
1	['State Bank keeps policy rate unchanged', 'NBP designated branches to facilitate PM's Youth Loan Scheme']	Banking	0.986677	0.986677

2

['India's State Bank Q3 net beats forecast on credit growth'] Banking
0.983775 0.983775
3 ['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 0.981375 0.981375
4 ['State Bank introduces fixed rental rate for Ijara Sukuk', 'UBL's literary excellence awards', 'State Life, Silkbank sign agreement', 'Bank Alfalah Partners with 1Link'] Banking 0.979918
0.979918
5 ['State Bank issues Rs114.23 billion worth currency notes in Ramazan', 'MCB ATMs to remain operational during Eid holidays'] Banking 0.976843 0.976843
6 ['State Bank unveils details of low risk Asaan account'] Banking 0.975748
0.975748
7 ['State Bank revises guidelines for banks' risk management capacity']
Banking 0.974185 0.974185
8 ['State Bank issues criteria to identify fitness of banknotes', 'Habib Bank in talks to buy Barclays Pakistan's business', 'Orix may acquire Standard Chartered Leasing'] Banking 0.973443 0.973443
9 ['State Bank names firms to determine market value of properties'] Banking
0.973317 0.973317
10 ['State Bank to bolster vigilance on foreign banking operations', 'Silkbank's Continues Growth Ascend - 2017 Half Yearly Performance', 'Silkbank concludes Rs2bln TFC issue'] Banking 0.972899
0.972899

User_4 (Energy / None):

rank
headline primary_tag similarity final_score
1 ['Higher food, gas prices to curb US growth', 'Ministry asks FBR to restore PSO accounts', 'PSO announces 30pc interim cash dividend'] Energy
0.984976 0.984976
2 ['PSO buys 520,000 tonnes of fuel oil', 'SSGC issues Rs86,808 bill to residential consumer'] Energy
0.984610 0.984610
3 ['SSGC staff recognised in US'] Other 0.983279 0.983279
4 ['PSO keen to increase revenues from LPG retail, lubricants'] Energy
0.982698 0.982698
5 ['SSGC MD demands revision in UFG benchmarks'] Other 0.981891
0.981891
6 ['Interested OMCs link investment

with TAPI gas field development'', 'PSO holds its 36th AGM']		Energy
0.980758	0.980758	
7		
['SSGC doubles efforts against thieves']	Other	0.980286
8		0.980286
['PSO yet to decide fate of oil import orders']	Energy	0.979367
0.979367		
9	['BoI seeks Malaysian investment for palm oil development', 'SSGC team in BD to learn about reducing gas losses']	Other
0.978592	0.978592	
10		
['PSO puts cap on fuel supply to PIA', 'SSGC recovers Rs194m']	Energy	
0.977133	0.977133	

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

User_1 (Construction / cement-heavy):		
rank		
headline primary_tag similarity final_score		
1		['KSE falls as
investors become cautious on political situation']	Other	0.926293
1.062773		
2	['Cabinet committee approves \$1.06 billion HBL divestment', 'Big appetite for HBL shares likely to augur well for KSE Index']	Banking
1.058090		0.954893
3		
['Mirza becomes PPL acting MD']	Energy	1.048188
4		['KSE to climb further
over likely announcement of good financial results']	Other	0.942829
1.046026		
5		
['Investors apprised of PSO's profitable growth model']	Energy	0.973260
1.039398		
6		
['Hascol EPS at Rs2.85']	Other	1.039284
7		['K-Electric shares success story with
investors', ''KEL's CFL distribution transparent'']	Other	0.932200
1.035397		
8		['Deteriorating
law & order situation breaks bullish spell on KSE']	Other	0.924041
1.031828		
9		
['Hascol reaches settlement with Mena Energy']	Other	0.943910
1.027696		
10		['KSE body formed to resolve issues with
SECP', 'Bearish spell continues, KSE loses 196 points']	Other	0.918127

1.025914

User_2 (Construction / infrastructure-heavy):

rank

headline	primary_tag	similarity	final_score	
1				['KSE falls as
investors become cautious on political situation']			Other 0.933732	
1.070212				
2				
['NBP becomes part of FTSE index series']	Banking	0.914697	1.051176	
3 ['Cabinet committee approves \$1.06 billion HBL divestment', 'Big appetite for HBL shares likely to augur well for KSE Index']	Banking	0.942894		
1.046091				
4				['KSE to climb further
over likely announcement of good financial results']			Other 0.942044	
1.045241				
5				
['Investors apprised of PSO's profitable growth model']	Energy	0.973953		
1.040091				
6				
['Hascol EPS at Rs2.85']	Other	0.953831	1.037617	
7				
['KSE-30 index likely to work from July']	Other	0.929847	1.033044	
8				['Big appetite
for HBL shares likely to augur well for KSE Index']	Banking	0.929200		
1.032398				
9				['LNG
becomes more volatile on heat wave, Trump's trade war']	Other	0.891941		
1.028421				
10				['Local
gross refinery margins likely to fall by 7 percent']	Energy	0.924112		
1.027309				

User_3 (Banking / None):

rank

headline	primary_tag	similarity	final_score	
1				
['Silk Bank allots right shares to CEO to fulfill capital requirements']				
Other 0.965662		1.086481		
2 ['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	0.981375	1.084573	
3				
['State Bank proposes task force to promote return filing']	Banking			
0.952819		1.073639		
4				
['State Bank likely to maintain status quo on policy rates']	Banking			
0.956239		1.059436		

5
 ['State Bank likely to reduce discount rate: analysts'] Banking 0.950172
 1.053370

6
 ['State Bank likely to keep policy rate unchanged'] Banking 0.948102
 1.051300

7
 ['Yaseen likely to be new State Bank governor'] Banking 0.943132
 1.046329

8
 ['State Bank revises guidelines for banks' risk management capacity']
 Banking 0.974185 1.040323

9
 ['NBP Serving the Nation in Thar Desert'] Banking 0.903365 1.039803
 10

10
 ['State Bank issues guidelines for agri water management financing', 'Standard Chartered to screen'] Banking 0.967331 1.033469

User_4 (Energy / None):
 rank
 headline primary_tag similarity final_score

1 ['PSO aims to become global energy company', 'PSO assures commitment to development in Balochistan']
 Energy 0.961126 1.097605

2 [''Govt to establish 100 LPG filling stations at PSO outlets'', 'PSO to establish fuel farm at Benazir Bhutto airport'] Energy 0.952543 1.073363

3
 ['Supplies likely to choke as PIA owes PSO Rs14 bn'] Energy 0.966123
 1.069320

4
 ['NBP staff seek demands fulfillment'] Banking 0.941243 1.062062

5
 ['SSGC to quadruple capital expenditure in FY 2007-08'] Other 0.959027
 1.061060

6 [''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 0.937608 1.058428

7
 ['PSO management says sell-off unlikely', 'PSO introduces SMS alert service']
 Energy 0.951517 1.054714

8 ['PSO management says sell-off unlikely', 'PSO introduces SMS alert service', 'HBL will install over 50 ATMs at Shell fuel stations'] Banking 0.948218 1.051416

9
 ['Vitol raises stake in Hascol to 25pc'] Other 0.967331 1.051117

10
 ['Nepra approves 12.2 paisas/kWh reduction for KEL consumers'] Other

0.946964 1.050162

=====
CONFIG: SBERT - no boost
=====

User_1 (Construction / cement-heavy):
rank
headline primary_tag similarity final_score
1 ['KSE
further up on buying in cement stocks'] Construction 0.782182 0.782182
2 ['KSE
ends up cement, power stocks in focus'] Construction 0.767803 0.767803
3 ['KSE committees not yet formed', 'Share
prices surge on strong valuations at KSE'] Other 0.754280 0.754280
4 ['KSE extraordinary general meeting adjourned till 22nd', 'KSE up 84
points on buying in energy sector stocks'] Other 0.752579 0.752579
5 ['PPL, govt sign concession agreements', 'KSE index
rises 304 points on better PSO payout'] Energy 0.751302 0.751302
6 ['KSE above 14,000
points on buying in cement, E&P stocks'] Construction 0.751109 0.751109
7 ['KSE reaches 7-week high on positive sentiments',
'Refineries under-perform KSE in 2011'] Other 0.747929 0.747929
8 ['46-floor KSE Tower project approved', 'Profit-selling
pushes KSE into consolidation phase'] Other 0.747094 0.747094
9 ['KSE up 26
points on interest in cement stocks'] Construction 0.742544 0.742544
10 ['Banking, energy scrips fuel KSE rise'] Other 0.741890 0.741890

User_2 (Construction / infrastructure-heavy):
rank
headline primary_tag similarity final_score
1 ['PPL, govt sign concession
agreements', 'KSE index rises 304 points on better PSO payout'] Energy
0.764775 0.764775
2 ['Cement, banking, oil scrips push KSE up by
147 points', 'Range bound activity witnessed at KSE'] Construction 0.763431
0.763431
3 ['Foreign buying of \$13m pushes up
KSE to 19-month high', 'KSE recomposes 100-share index'] Other
0.763021 0.763021
4 ['KSE committees not yet
formed', 'Share prices surge on strong valuations at KSE'] Other
0.760865 0.760865
5 ['Foreigners hold 30pc of \$11bn free float KSE market', 'KSE index crosses
17,600 points on foreign interest in banking shares'] Other 0.759074
0.759074

6						
KSE', 'KSE Index rises as fresh investment pours in']					Other	0.758151
0.758151						
7						
by year-end', 'KSE 100-share index gains 133 points']					Other	0.755779
0.755779						
8						
'KSE up 157 points as institutions,retailers enter oversold market', 'European firm to market KSE indices abroad']					Other	0.754246
0.754246						
9						
'Europe, ME companies eye stake in KSE', 'Stock market stages recovery, KSE-100 up 109 points']					Other	0.752164
0.752164						
10						
'KSE index gains 324 points on hike in discount rate']					Other	0.752026
0.752026						

User_3 (Banking / None):

rank						
headline primary_tag similarity final_score						
1						
lowering discount rate']	Banking	0.821997	0.821997			['State Bank hints at
2 ['State Bank amends rules to combat money laundering', 'SBP approves NIB						
Bank's merger into MCB']	Banking	0.809805	0.809805			
3						
rate by 50bps to 13.5pc']	Banking	0.799379	0.799379			['State Bank raises discount
4						
Bank's security tightened']	Banking	0.775912	0.775912			
5 ['Senate committee opposes discretionary powers for State Bank', 'MCB Bank						
posts Rs4.1bn profit']	Banking	0.775625	0.775625			
6						
['State Bank announces clean note policy', 'NIB						
selling NAFA stake to NBP']	Banking	0.775069	0.775069			
7						
['UBL announces 10pc cash dividend', 'NBP aims to						
be top bank of region']	Banking	0.771155	0.771155			
8						
['State Bank injects						
Rs267bn into banking system']	Banking	0.760011	0.760011			
9						
['State Bank injects						
Rs603bn into banking system']	Banking	0.758191	0.758191			
10						
gives warning signals']	Banking	0.757321	0.757321			['State Bank

User_4 (Energy / None):

rank						
headline primary_tag similarity final_score						
1 ['US production is putting brakes on oil prices: IEA', 'Mari Petroleum,						
KUFPEC sign MoU', 'PPL donates Rs9.02mln to LDH']					Energy	0.778932
0.778932						
2						
['PSO yet to decide fate of oil import orders']					Energy	0.764404

0.764404
 3
 ['PSO buys 520,000 tonnes of fuel oil'] Energy 0.757625 0.757625
 4
 ['PSO not to buy Pakistan Refinery'] Energy 0.750168 0.750168
 5 ['PSO buys 520,000 tonnes of fuel oil', 'SSGC issues Rs86,808 bill to residential consumer'] Energy 0.749107
 0.749107
 6
 ['PSO seeks offers to import fuel oil'] Energy 0.742193 0.742193
 7 ['POL imports rise due to energy crisis', 'PSO earns profits of Rs6.3 billion'] Energy 0.741916
 0.741916
 8 ['PSO posts Rs5bn profit in first half', 'PSO buys up to 780,000T fuel oil'] Energy 0.741904
 0.741904
 9 ['PSO in process of buying 30pc shares of Pakistan Refinery'] Energy 0.741214
 0.741214
 10 ['Oil prices steady after nearing \$124', 'PSO bankrolls NICH'] Energy 0.739284
 0.739284

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

User_1 (Construction / cement-heavy):
 rank
 headline primary_tag similarity final_score
 1 ['Late buying in energy, telecom scrips minimises losses on KSE', 'KSE MD elected SAFE vice-chairman'] Other 0.739673
 0.807564
 2 ['KSE committees not yet formed', 'Share prices surge on strong valuations at KSE'] Other 0.754280 0.807451
 0.754280
 3 ['KSE reaches 7-week high on positive sentiments', 'Refineries under-perform KSE in 2011'] Other 0.747929
 0.801100
 4 ['Hubco to invest Rs1bn in coal-fired power plant', 'KEL profits up 81pc'] Other 0.799986 0.696789
 0.799986
 5 ['KSE index down 150 points on uncertainty over budget', 'KSE approves formal listing of Hascol'] Other 0.786932 0.703146
 0.786932
 6 ['Earnings, political scene likely to set KSE direction'] Other 0.679839 0.783036

7

['KSE further up on buying in cement stocks']	Construction	0.782182
0.782182		
8 ['Cabinet committee approves \$1.06 billion HBL divestment', 'Big appetite for HBL shares likely to augur well for KSE Index']	Banking	0.676491
0.779689		
9		['KSE falls as investors become cautious on political situation']
0.778863	Other	0.642383
10		['KSE delists 7 companies', 'Turnover at KSE improved by 25 million shares']
0.712385	Other	0.778523

User_2 (Construction / infrastructure-heavy):

rank		
headline primary_tag similarity final_score		
1		['KSE committees not yet formed', 'Share prices surge on strong valuations at KSE']
	Other	0.760865 0.814036
2		['KSE index down 150 points on uncertainty over budget', 'KSE approves formal listing of Hascol']
	Other	0.724205 0.807991
3		['KSE index likely to surpass 15,500 points on foreign buying']
	Other	0.702639 0.805836
4		['KSE index up 117 points on budget concerns', 'Trading in Hascol shares begins at KSE']
	Other	0.714683 0.798469
5		['KSE approves formal listing of Al-Shaheer Corporation', 'KSE Index fall, dragged by massive foreign selling']
	Other	0.743869 0.797040
6		['KSE continued stellar performance in 2014', 'KSE 100 rises by 47 points on buying in second-tier stocks']
	Other	0.739429 0.792600
7		['KSE index up by 84 points on buying in telecom, cement stocks']
	Construction	0.724435 0.792327
8		['Late buying in energy, telecom scrips minimises losses on KSE', 'KSE MD elected SAFE vice-chairman']
	Other	0.723596 0.791487
9		['KSE falls as investors become cautious on political situation']
	Other	0.649296 0.785775
10		['KSE outperforms regional, world markets']
	Other	0.732394 0.785565

User_3 (Banking / None):

rank		
headline primary_tag similarity final_score		
1		['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	0.735380 0.838578
2		['State Bank likely to reduce discount rate: analysts']
0.830933	Banking	0.727735
3		['UBL announces 10pc cash dividend', 'NBP aims to be top bank of region']

Banking 0.771155 0.826909
 4
 ['State Bank hints at lowering discount rate'] Banking 0.821997
 0.821997
 5
 ['State Bank amends rules to combat money laundering', 'SBP approves NIB Bank's merger into MCB'] Banking 0.809805 0.818249
 6
 ['State Bank delays monetary data', 'JCR-VIS reaffirms NBP AAA/A-1+ rating', 'Askari Bank suffers loss'] Banking 0.743345 0.809484
 7
 ['State Bank likely to maintain status quo on policy rates'] Banking
 0.701640 0.804837
 8
 ['State Bank announces clean note policy', 'NIB selling NAFA stake to NBP']
 Banking 0.775069 0.804178
 9
 ['State Bank raises discount rate by 50bps to 13.5pc'] Banking 0.799379
 0.799379
 10
 ['Senate committee opposes discretionary powers for State Bank', 'MCB Bank posts Rs4.1bn profit'] Banking 0.775625 0.784069

User_4 (Energy / None):
 rank
 headline primary_tag similarity final_score
 1
 ['PSO vying to become single largest shareholder in Pakistan Refinery Ltd']
 Energy 0.706650 0.843130
 2 ['PSO aims to become global energy company', 'PSO assures commitment to development in Balochistan']
 Energy 0.697662 0.834142
 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 0.677670 0.798489
 4
 ['PSO seeks 143,000 cubic meters of LNG for July delivery'] Energy
 0.720584 0.786722
 5 ['US production is putting brakes on oil prices: IEA', 'Mari Petroleum, KUFPEC sign MoU', 'PPL donates Rs9.02mln to LDH']
 Energy 0.778932 0.783281
 6
 ['PTCL hikes UAN tariff', 'PTCL donates Rs10m for BISP'] Other 0.727038
 0.778673
 7
 ['MD PSO discusses prospects with dealers in Quetta visit'] Energy
 0.691051 0.774840
 8

```

['Supplies likely to choke as PIA owes PSO Rs14 bn']      Energy    0.670828
0.774026
9
[''Govt to establish 100 LPG filling
stations at PSO outlets'', 'PSO to establish fuel farm at Benazir Bhutto
airport']      Energy    0.650659    0.771478
10
['PSO yet to decide fate of oil import orders']      Energy    0.764404
0.769402

```

1.6 Step 6 — Evaluate and compare

```
[12]: 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					
0.9619	W2V - no boost	User_1	Construction	0.0	2
0.9666	W2V - no boost	User_2	Construction	0.0	2
0.9778	W2V - no boost	User_3	Banking	1.0	1
0.9814	W2V - no boost	User_4	Energy	0.6	2
0.9383	W2V - recency boost	User_1	Construction	0.0	3
0.9336	W2V - recency boost	User_2	Construction	0.0	3
0.9542	W2V - recency boost	User_3	Banking	0.9	2
0.9532	W2V - recency boost	User_4	Energy	0.5	3
0.7539	SBERT - no boost	User_1	Construction	0.4	3
0.7584	SBERT - no boost	User_2	Construction	0.1	3
0.7804	SBERT - no boost	User_3	Banking	1.0	1

SBERT - no boost	User_4	Energy	1.0	1
0.7507				
SBERT - recency boost	User_1	Construction	0.1	3
0.7135				
SBERT - recency boost	User_2	Construction	0.1	2
0.7215				
SBERT - recency boost	User_3	Banking	1.0	1
0.7661				
SBERT - recency boost	User_4	Energy	0.9	2
0.7085				

```
[46]: # 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('→ 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	0.425	1.50	0.9711
0.9711			
W2V - recency boost	0.200	2.25	0.9380
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

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

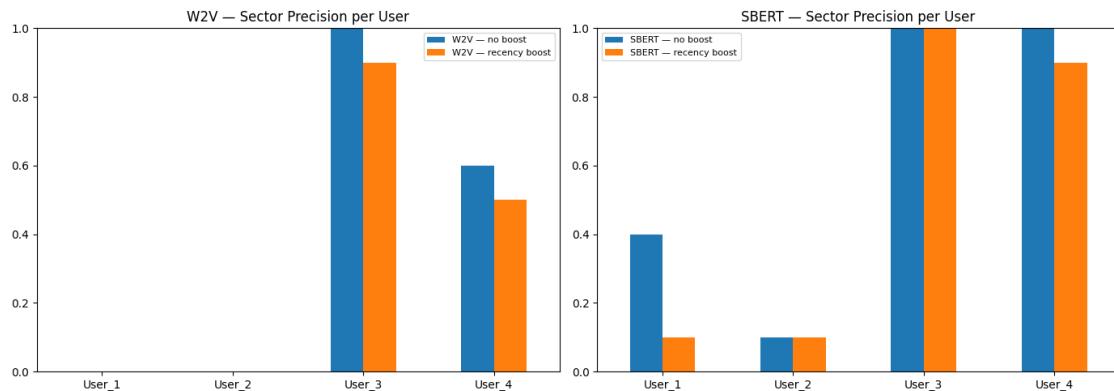
for ax, model in [(axes[0], 'W2V'), (axes[1], 'SBERT')]:
    subset = df_eval[df_eval['config'].str.startswith(model)]
    pivot = subset.pivot(index='user_id', columns='config',u
    ↪values='sector_precision')
    pivot.plot(kind='bar', ax=ax, title=f'{model} - Sector Precision per User',u
    ↪ylim=(0, 1))
```

```

    ax.set_xlabel('')
    ax.tick_params(axis='x', rotation=0)
    ax.legend(fontsize=8)

plt.tight_layout()
plt.show()

```



1.7 Step 7 — Save

```

[14]: CONFIG_FILERAMES = {
    'W2V - no boost': 'recs_w2v_no_boost.csv',
    'W2V - recency boost': 'recs_w2v_boost.csv',
    'SBERT - no boost': 'recs_sbert_no_boost.csv',
    'SBERT - recency boost': 'recs_sbert_boost.csv',
}

for config_name, user_recs in all_recs.items():
    frames = []
    for user_id, recs in user_recs.items():
        recs = recs.copy()
        recs.insert(0, 'user_id', user_id)
        frames.append(recs)
    combined = pd.concat(frames, ignore_index=True)
    path = f'../data/processed/recommendations/{CONFIG_FILERAMES[config_name]}'
    combined.to_csv(path, index=False)
    print(f'Saved: {path} ({len(combined)} rows)')

print('\nAll recommendation CSVs saved.')

```

```

Saved: ../data/processed/recommendations/recs_w2v_no_boost.csv (40 rows)
Saved: ../data/processed/recommendations/recs_w2v_boost.csv (40 rows)
Saved: ../data/processed/recommendations/recs_sbert_no_boost.csv (40 rows)
Saved: ../data/processed/recommendations/recs_sbert_boost.csv (40 rows)

```

All recommendation CSVs saved.

1.8 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