```python
texts = []
with open("Reviews.csv", encoding="utf-8") as f:
    for i, line in enumerate(f):
        if i >= 1000:
            break
        if line.strip():
            texts.append(line.strip())

from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(max_features=1000)
tfidf_matrix = vectorizer.fit_transform(texts)
print(tfidf_matrix.toarray())
feature_names = vectorizer.get_feature_names_out()
for idx, word in enumerate(feature_names):
    print(f"{idx}: {word}")
```

```
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
0: 00
1: 10
2: 100
3: 11
4: 12
5: 13
6: 15
7: 16
8: 20
9: 24
10: 25
11: 30
12: 50
13: 99
14: able
15: about
16: absolutely
17: acai
18: acid
19: actual
20: actually
21: add
22: added
23: addicted
24: adding
25: addition
```

```
26: adult
27: advertised
28: after
29: aftertaste
30: again
31: agave
32: ago
33: ahmad
34: all
35: allergic
36: allergies
37: almost
38: along
39: already
40: also
41: alternative
42: although
43: altoids
44: always
45: am
46: amazing
47: amazon
48: amount
49: an
50: and
51: another
52: any
53: anymore
54: anyone
55: anything
56: anyway
57: anywhere
58: apple
59: apples
60: are
61: aren
62: aroma
63: around
64: arrived
65: artificial
66: as
67: ask
68: asparagus
69: at
70: ate
71: available
72: away
73: awesome
74: awful
```

```
75: b000er6yo0
76: b000g6mbx2
77: b000g6ryne
78: b000hdmuq2
79: b000lkzk7c
80: b000s806vm
81: b000vkyktg
82: b0018dqfpc
83: b0019cw0he
84: b001ell6o8
85: b001ell9x6
86: b001eo5qw8
87: b001gvisjm
88: b001ujen6c
89: b0028c44z0
90: b002bcd2og
91: b0035ye9cs
92: b003ao5dlo
93: b003ob0ib8
94: b003se19uk
95: b006f2nyi2
96: babies
97: baby
98: back
99: backyard
100: bad
101: bag
102: bags
103: baked
104: baking
105: balance
106: banana
107: bar
108: bars
109: base
110: based
111: basket
112: bbq
113: be
114: bean
115: beans
116: bears
117: beat
118: because
119: become
120: been
121: beer
122: before
123: being
```

```
124: believe
125: berry
126: best
127: better
128: between
129: big
130: bigger
131: bit
132: bite
133: bitter
134: black
135: bland
136: blend
137: blue
138: body
139: book
140: boost
141: both
142: bottle
143: bottles
144: bottom
145: bought
146: box
147: boxes
148: br
149: brand
150: brands
151: bread
152: breakfast
153: breath
154: brew
155: broken
156: brought
157: brown
158: bulk
159: burn
160: burnt
161: but
162: butter
163: buy
164: buying
165: by
166: caffeine
167: cake
168: called
169: calorie
170: calories
171: came
172: can
```

```
173: candy
174: cane
175: canned
176: cannot
177: caramel
178: carb
179: care
180: carrots
181: carry
182: case
183: cat
184: cats
185: cause
186: certainly
187: change
188: changed
189: cheap
190: cheaper
191: cheddar
192: cheese
193: chemical
194: cherry
195: chicken
196: chip
197: chips
198: chocolate
199: choice
200: christmas
201: cinnamon
202: clean
203: clear
204: close
205: cocker
206: coffee
207: coffees
208: cold
209: color
210: com
211: combination
212: come
213: comes
214: company
215: completely
216: considering
217: consistency
218: contain
219: container
220: contains
221: continue
```

```
222: control
223: convenience
224: convenient
225: cook
226: cooked
227: cookie
228: cookies
229: cooking
230: corn
231: cost
232: could
233: couldn
234: count
235: country
236: couple
237: course
238: cream
239: creamer
240: creamers
241: crisp
242: crispy
243: crunch
244: crunchy
245: cup
246: cups
247: customer
248: cut
249: cute
250: daily
251: dark
252: date
253: daughter
254: day
255: days
256: deal
257: decent
258: decided
259: definitely
260: delicious
261: delivered
262: delivery
263: did
264: didn
265: diet
266: difference
267: different
268: difficult
269: dijon
270: dinner
```

```
271: dip
272: disappointed
273: disappointing
274: dish
275: do
276: does
277: doesn
278: dog
279: dogs
280: don
281: done
282: down
283: dried
284: drink
285: drinking
286: drinks
287: dry
288: due
289: during
290: each
291: earl
292: earth
293: easier
294: easily
295: easy
296: eat
297: eaten
298: eating
299: eats
300: edible
301: eggs
302: either
303: else
304: end
305: energy
306: enjoy
307: enjoyed
308: enough
309: entire
310: especially
311: etc
312: even
313: ever
314: every
315: everyday
316: everyone
317: everything
318: exactly
319: excellent
```

```
320: excited
321: expect
322: expected
323: expensive
324: experience
325: expiration
326: extra
327: extract
328: extremely
329: fact
330: fair
331: fall
332: family
333: fan
334: fantastic
335: far
336: fast
337: fat
338: fats
339: favorite
340: favorites
341: feeding
342: feel
343: felidae
344: few
345: figured
346: filled
347: finally
348: find
349: finding
350: fine
351: finish
352: first
353: fish
354: five
355: flavor
356: flavored
357: flavorful
358: flavoring
359: flavors
360: flavour
361: flour
362: flower
363: food
364: foods
365: for
366: form
367: formula
368: found
```

```
369: free
370: fresh
371: fried
372: friend
373: friends
374: from
375: frosting
376: fruit
377: full
378: fun
379: garlic
380: gas
381: gave
382: get
383: gets
384: getting
385: gift
386: ginger
387: gingerbread
388: give
389: given
390: gives
391: giving
392: glad
393: glass
394: gluten
395: go
396: goes
397: going
398: gold
399: gone
400: good
401: got
402: gourmet
403: gp
404: grain
405: grape
406: gravy
407: greasy
408: great
409: green
410: grey
411: grocery
412: guess
413: gum
414: had
415: half
416: hand
417: happy
```

```
418: hard
419: has
420: hate
421: have
422: haven
423: having
424: he
425: health
426: healthier
427: healthy
428: heard
429: heart
430: heat
431: heavy
432: help
433: helps
434: her
435: here
436: high
437: higher
438: highly
439: him
440: hint
441: his
442: hit
443: holistic
444: home
445: honey
446: hooked
447: hope
448: horrible
449: hot
450: hour
451: house
452: how
453: however
454: href
455: http
456: huge
457: husband
458: ice
459: icing
460: idea
461: if
462: immediately
463: in
464: ingredient
465: ingredients
466: inside
```

```
467: instant
468: instead
469: into
470: introduced
471: is
472: isn
473: issue
474: issues
475: it
476: item
477: items
478: its
479: itself
480: jalapeno
481: japanese
482: jar
483: jars
484: john
485: juice
486: just
487: keep
488: ketchup
489: kettle
490: kick
491: kids
492: kind
493: kitchen
494: knew
495: know
496: kona
497: label
498: large
499: larger
500: last
501: lays
502: learned
503: least
504: left
505: lemon
506: less
507: let
508: licorice
509: life
510: light
511: like
512: liked
513: likes
514: line
515: liquid
516: list
```

```
517: little
518: live
519: ll
520: loaded
521: local
522: long
523: longer
524: look
525: looked
526: looking
527: looks
528: loose
529: lose
530: lot
531: lots
532: love
533: loved
534: lover
535: loves
536: low
537: lower
538: lunch
539: made
540: make
541: makes
542: making
543: many
544: maple
545: marinade
546: market
547: may
548: maybe
549: mccann
550: me
551: meal
552: mean
553: meat
554: melt
555: melted
556: mess
557: middle
558: might
559: mild
560: milk
561: mind
562: mine
563: mint
564: mints
565: minutes
```

```
566: miss
567: missing
568: mix
569: mixed
570: mom
571: money
572: month
573: months
574: more
575: morning
576: most
577: mostly
578: mother
579: mountain
580: mouth
581: msg
582: much
583: must
584: my
585: myself
586: name
587: natural
588: naturally
589: nectar
590: need
591: needed
592: never
593: new
594: next
595: nice
596: nicely
597: no
598: non
599: none
600: normal
601: not
602: note
603: nothing
604: now
605: oatmeal
606: oats
607: of
608: off
609: often
610: oh
611: oil
612: oily
613: ok
614: old
```

```
615: on
616: once
617: one
618: ones
619: onion
620: online
621: only
622: open
623: opened
624: opinion
625: option
626: or
627: order
628: ordered
629: ordering
630: organic
631: original
632: other
633: others
634: otherwise
635: ounce
636: ounces
637: our
638: out
639: outside
640: over
641: overly
642: own
643: oz
644: pack
645: package
646: packaged
647: packaging
648: packed
649: packs
650: paid
651: pancakes
652: part
653: party
654: past
655: pay
656: peanut
657: peanuts
658: people
659: pepper
660: per
661: perfect
662: perfectly
663: perhaps
```

```
664: person
665: picky
666: pieces
667: pineapple
668: place
669: plain
670: plan
671: plastic
672: please
673: pleased
674: plocky
675: plus
676: pocky
677: pop
678: potato
679: potatoes
680: pounds
681: powder
682: prefer
683: prepare
684: pretty
685: price
686: prices
687: prime
688: probably
689: problem
690: problems
691: product
692: products
693: protein
694: purchase
695: purchased
696: purchasing
697: pure
698: put
699: quality
700: quick
701: quickly
702: quite
703: rancid
704: rather
705: raw
706: re
707: read
708: real
709: really
710: reason
711: received
712: recently
```

```
713: recipe
714: recommend
715: recommended
716: red
717: regular
718: rest
719: results
720: review
721: reviewer
722: reviews
723: rice
724: rich
725: right
726: room
727: run
728: runny
729: said
730: sale
731: salsa
732: salt
733: salted
734: salty
735: same
736: sandwich
737: sassafras
738: satisfying
739: sauce
740: save
741: saw
742: say
743: says
744: school
745: scissors
746: sea
747: seasoning
748: second
749: see
750: seem
751: seems
752: seen
753: select
754: sell
755: seller
756: selling
757: sensitive
758: service
759: serving
760: set
761: several
```

```
762: share
763: she
764: ship
765: shipment
766: shipped
767: shipping
768: shop
769: short
770: shot
771: shots
772: should
773: side
774: similar
775: simple
776: simply
777: since
778: single
779: six
780: size
781: sized
782: slightly
783: small
784: smaller
785: smell
786: smooth
787: snack
788: snacking
789: snacks
790: so
791: soda
792: sodium
793: soft
794: solomon
795: some
796: someone
797: something
798: sometimes
799: somewhat
800: son
801: soon
802: sort
803: soup
804: sour
805: soy
806: special
807: spice
808: spices
809: spicy
810: stale
```

```
811: star
812: stars
813: start
814: started
815: state
816: stay
817: steaz
818: stick
819: sticks
820: sticky
821: still
822: stock
823: stop
824: stopped
825: store
826: stores
827: strawberry
828: strong
829: stuck
830: stuff
831: style
832: subscribe
833: substitute
834: subtle
835: such
836: sugar
837: summer
838: super
839: supermarket
840: sure
841: surprise
842: surprised
843: sweet
844: sweetener
845: sweetness
846: syrup
847: taffy
848: take
849: takes
850: taking
851: tangy
852: tart
853: tarts
854: tassimo
855: taste
856: tasted
857: tastes
858: tasting
859: tasty
```

```
860: tea
861: teas
862: teeth
863: tell
864: terrible
865: texture
866: thai
867: than
868: thank
869: thanks
870: that
871: the
872: their
873: them
874: themselves
875: then
876: there
877: these
878: they
879: thick
880: thicker
881: thin
882: thing
883: things
884: think
885: this
886: those
887: though
888: thought
889: three
890: through
891: throw
892: time
893: times
894: tin
895: tiny
896: tired
897: to
898: together
899: told
900: too
901: took
902: top
903: tortilla
904: totally
905: touch
906: trans
907: treat
908: treats
```

```
909: tried
910: true
911: try
912: trying
913: tuna
914: turkey
915: twizzlers
916: two
917: type
918: under
919: unfortunately
920: unique
921: unless
922: unsalted
923: until
924: up
925: us
926: use
927: used
928: using
929: usually
930: value
931: vanilla
932: varieties
933: variety
934: ve
935: vegetable
936: veggies
937: vernor
938: version
939: very
940: vet
941: vinegar
942: vitamin
943: want
944: wanted
945: wants
946: was
947: wasn
948: waste
949: water
950: way
951: we
952: weather
953: website
954: week
955: weeks
956: weight
957: well
```

```
958: went
959: were
960: what
961: whatever
962: wheat
963: when
964: where
965: which
966: while
967: white
968: who
969: whole
970: why
971: will
972: wine
973: wish
974: with
975: within
976: without
977: won
978: wonderful
979: work
980: worked
981: works
982: world
983: worse
984: worst
985: worth
986: would
987: wouldn
988: wow
989: wrong
990: www
991: year
992: years
993: yes
994: yet
995: you
996: your
997: yourself
998: yum
999: yummy
```

```
from gensim.models import Word2Vec
from nltk.tokenize import word_tokenize
import nltk

nltk.download('punkt')
sentences = [doc.lower().split() for doc in texts]
model = Word2Vec(sentences, vector_size=100, window=5, min_count=1)  #
```

```
min_count=1 żeby działało nawet na małym zbiorze
print("Podobieństwo 'food' i 'good':", model.wv.similarity('food',
'good'))  # wartości od -1 do 1
print("Najbardziej podobne do 'food':")
print(model.wv.most_similar('food'))
```

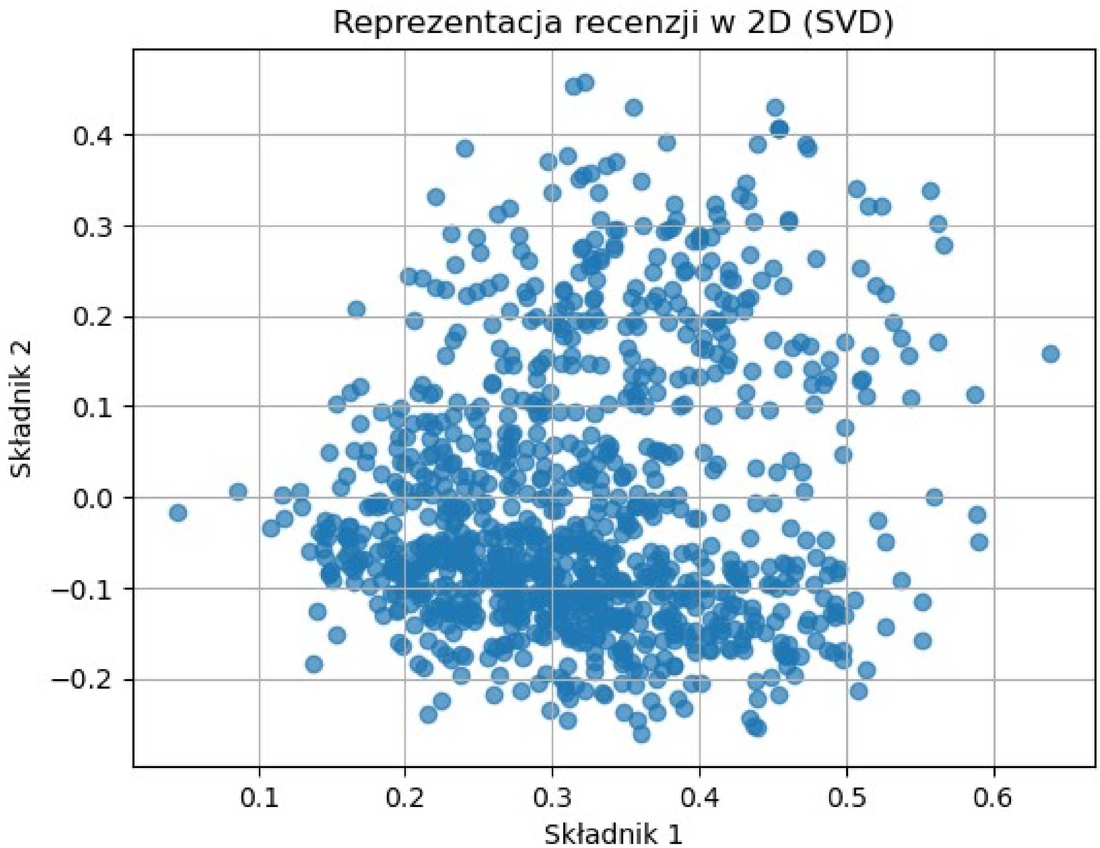
```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\48664\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

```
Podobieństwo 'food' i 'good': 0.99985147
Najbardziej podobne do 'food':
[('just', 0.999893307685852), ('in', 0.9998893737792969), ('and',
0.9998839497566223), ('also', 0.9998823404312134), ('from',
0.9998819231987), ('our', 0.9998814463615417), ('for',
0.9998805522918701), ('was', 0.9998801350593567), ('or',
0.999879777431488), ('but', 0.9998793601989746)]
```

```
from sklearn.decomposition import TruncatedSVD
import matplotlib.pyplot as plt

# Redukcja wymiarowości do 2D
svd = TruncatedSVD(n_components=2)
X_2d = svd.fit_transform(tfidf_matrix)

# Rysowanie wykresu
plt.scatter(X_2d[:, 0], X_2d[:, 1], alpha=0.7)
plt.title("Reprezentacja recenzji w 2D (SVD)")
plt.xlabel("Składnik 1")
plt.ylabel("Składnik 2")
plt.grid(True)
plt.show()
```


Reprezentacja recenzji w 2D (SVD)
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
Składnik 2
0.1
0.2
0.3
0.4
0.5
0.6
Składnik 1