CENTER ZA METODOLOGIJO IN INFORMATIKO
FAKULTETA ZA DRUŽBENE VEDE
UNIVERZA V LJUBLJANI

Identifikacija z rakom povezanih sporočil v spletni zdravstveni skupnosti MedOverNet

Raziskovalno poročilo projekta Priložnosti in nevarnosti spletnih zdravstvenih skupnosti za zdravstvo

Marjan Cugmas, Gregor Petrič

Projekt in raziskavo je financirala Agencija za raziskovalno dejavnost Slovenije (ARRS) (J5-9336)

Projekt in raziskava je potekala v sodelovanju z Onkološkim inštitutom Ljubljana (dr. Rok Petrič)

Ljubljana, september 2020

Kazalo

1 Kazalo tabel ... 3

2 Kazalo slik ... 4

3 Uvod ... 5

3.1 Uporabljene okrajšave ... 6

4 Podatki ... 7

4.1 Pregled števila sporočil po skupinah forumov ... 11

5 Določanje teme raka v sporočilih ... 12

5.1 Korpus ... 12

5.2 Priprava besedila ... 12

5.3 Uporabljene metode nadzorovanega strojnega učenja ... 12

5.3.1 Naivni Bayesov klasifikator (NBC) ... 13

5.3.2 Metoda podpornih vektorjev (SVM) ... 13

5.3.3 Odločitvena drevesa (TREE) ... 13

5.3.4 Slučajni gozdovi (FOREST) ... 14

5.4 Učna množica ... 14

5.5 Evalvacija rezultatov metod za uvrščanje v skupine ... 15

5.6 Ponovno treniranje modela ... 17

5.7 Evalvacija posodobljenih modelov ... 18

6 Opredelitev forumske teme, ki vključuje temo raka ... 22

7 Pojavnost sporočil in forumskih tem s temo raka ... 23

7.1 Pojavnost sporočil in forumskih tem s temo raka po letih ... 32

8 Vzorčenje forumskih tem o raku ... 42

9 Zaključek ... 45

10 Literatura ... 46

1 Kazalo tabel

Tabela 1 Porazdelitev števila sporočil znotraj forumskih tem po forumih 7
Tabela 2 Forumske teme z natanko 1.000 sporočili 10
Tabela 3 Število sporočil po forumskih temah, glede na vrsto forumov 11
Tabela 4 Seznam ključnih besed, uporabljenih za določitev učne množice sporočil 14
Tabela 5 Velikost učne množice sporočil (način vključitve in vsebina) 15
Tabela 6 Vrednosti Pearsonovega koeficienta korelacije med dobljenimi rešitvami (CV-10) 16
Tabela 7 Klasifikacijska tabela za model odločitvenih dreves (CV-10) 16
Tabela 8 Primerjava razvrstitev sporočil (model slučajnih gozdov vs. ročna klasifikacija) 17
Tabela 9 Primerjava rešitev prvotnega in posodobljenega modela slučajnih gozdov 18
Tabela 10 Število sporočil s temo raka po forumih: primerjava prvotnega in posodobljenega modela 19
Tabela 11 Število sporočil s temo raka po forumih 23
Tabela 12 Število sporočil s temo raka, glede na vrsto foruma 26
Tabela 13 Število forumskih tem, glede na vrsto foruma (forumska tema je opredeljena kot forumska tema o raku, če vsaj polovica sporočil vsebuje temo raka) 26
Tabela 14 Število forumskih tem, glede na vrsto foruma (forumska tema je opredeljena kot forumska tema o raku, če vsaj eno sporočilo vsebuje temo raka) 29
Tabela 15 Število koledarskih let, prek katerih se raztezajo forumske teme (ne glede na to, ali vsebujejo temo raka ali ne) 32
Tabela 16 Število forumskih tem glede na leto prvega sporočila, glede na vrsto foruma: upoštevane so forumske teme, ki vsebujejo temo raka (forumska tema vsebuje temo raka, če vsaj eno sporočilo vsebuje temo raka) 32
Tabela 17 Število forumskih tem s temo raka (forumska tema vsebuje temo raka, če vsaj eno sporočilo vsebuje temo raka) po obdobjih 32
Tabela 18 Število forumskih tem s temo raka (forumska tema vsebuje temo raka, če vsaj eno sporočilo vsebuje temo raka) po letih 35
Tabela 19 Število forumskih tem s temo raka po forumih (forumska tema vsebuje temo raka, če vsaj polovica sporočil vsebuje temo raka) (upoštevane so samo forumske teme z več kot enim sporočilom) 38
Tabela 20 Izbor ustreznih forumskih tem za vzorčenje forumskih tem s temo raka 42
Tabela 21 Vzorec izbranih forumov 43

2 Kazalo slik

Slika 1 Število sporočil v času 7
Slika 2 Evalvacija modelov (CV-10) 16
Slika 3 Evalvacija modelov (CV-10) ocenjenih na podlagi dopolnjene učne množice sporočil 18
Slika 4 Skupen delež forumskih tem o raku, glede na različne pragove določitve forumske teme o raku 22
Slika 5 Primer števila forumov in forumskih tem na različnih korakih izbora forumskih tem 43

3 Uvod

Pričujoč dokument je del projekta *Priložnosti in nevarnosti spletnih zdravstvenih skupnosti za zdravstvo*, ki ga financira Agencija za raziskovalno dejavnost Slovenije (ARRS). Gre za del aktivnosti delovnega sklopa DS3 (Merjenje strukturnih značilnosti SZS), katerega namen je identificirati in urediti podatkovno bazo z vsemi sporočil, ki se vsebinsko nanašajo na temo bolezni raka v spletni zdravstveni skupnosti MedOverNet (med.over.net). Baza se v naslednjem koraku uporabi kot izhodišče za določitev sporočil, ki se bodo uporabili v namen merjenja kvalitete z rakom povezanih informacij, ki se objavljajo v spletni zdravstveni skupnosti MedOverNet.

Spletno mesto zdravstvene skupnosti MedOverNet (v času nastanka pričujočega poročila) sodi med najbolj obiskana spletna mesta v Sloveniji (MOSS 2020). Osrednji del spletne zdravstvene skupnosti predstavljajo diskusijski forumi (forumi), znotraj katerih obstaja več forumskih tem, znotraj forumskih tem pa obstaja ena ali več sporočil. Forume na MedOverNet je mogoče razvrstiti v tri skupine (Atanasova 2020):

- **Zdravstvene posvetovalnice** (ZP) so namenjene predvsem dajanju nasvetov, priporočil in informacij uporabnikom s strani moderatorjev, ki so večinoma zdravniki, zdravniki specialisti in drugi strokovnjaki s področja zdravstva in zdravja ali drugih povezanih tem.
- **Forumi spletnih podpornih skupin** (PS) so namenjeni izmenjavi izkušenj, informacij, socialne opore in nasvetov med uporabniki, ki so pogosto pacienti enakih ali podobnih bolezni. Takšni forumi pogosto vključujejo moderatorja, ki usmerja diskusijo in predstavlja pomemben vir informacij za uporabnike, saj so moderatorji pogosto predstavniki društev in združenj pacientov ali pa pacienti s pomembnimi izkušnjami doživljanja bolezni.
- **Družabni forumi** (DF) so namenjeni vsakodnevnim tematikam, kot so prosti čas, potovanja, lepota, politika in druge. Tovrstne forume pogosto moderirajo diskusijski moderatorji, nekateri forumi pa so tudi nemoderirani.

Na MedOverNet so bolezni raka namenjeni številni forumi: nekateri so izključno posvečeni tej temi, nekateri forumi pa poleg drugih tem, povezanih z zdravjem in različnimi področji medicine vključujejo tudi vprašanja, odgovore in diskusije na temo bolezni raka. V večji meri se tema raka pojavlja v forumih zdravstvenih posvetovalnic in forumih podpornih skupin (Atanasova 2020).

Za identificiranje diskusij, ki se nanašajo na temo raka, so bili uporabljeni različni pristopi za analizo besedil. Dobljeni rezultati bodo služili kot izhodišče za kodiranje kakovosti informacij, objavljenih v spletni zdravstveni skupnosti MedOverNet. Natančneje je logika postopka naslednja: 1) Preizkus različnih metod za določitev sporočil, ki se z visoko verjetnostjo nanašajo na temo raka; 2) Opredelitev forumskih tem (ang. *threads*), ki vključujejo vsaj k sporočil na temo raka; 3) Vzorčenje forumskih tem za analizo besedil.

3.1 Uporabljene okrajšave

Okrajšava	*Pomen*	*Okrajšava*	*Pomen*	*Okrajšava*	*Pomen*
N	število enot	Q3	tretji kvartil	NBC	naivni Bayesov klasifikator
M	povprečje	MAX	maksimalna vrednost	SVM	metoda podpornih vektorjev
SD	standardni odklon	ZP	zdravstvene posvetovalnice	TREE	metoda odločitvenih dreves
MIN	minimalna vrednost	PS	podporne skupine	FOREST	metoda slučajnih gozdov
Q1	prvi kvartil	DF	družabni forumi	CV-10	metoda 10-kratnega navzkrižnega preverjanja modela
Q2	drugi kvartil, mediana	TF	frekvence besed	AUC	ploščina pod krivuljo

4 Podatki

Podatki za analizo so bili posredovani v formatu .json s strani MedOverNet. Zajemajo vsa sporočila, ki so bila objavljena v letih vključno med 2015 in vključno 2019. Gre za 2.982.564 sporočil, ki so bila objavljena v 280.357 forumskih temah.

Število sporočil v času rahlo upada (Slika 1). Število sporočil znotraj forumskih tem se pričakovano porazdeljuje zelo asimetrično: okoli 48712 (18 %) forumskih tem vsebuje samo enim sporočilom, 25 % forumskih tem vsebuje 2 sporočila ali manj, 50 % forumskih tem vsebuje 4 sporočila ali manj in 75 % forumskih tem vsebuje 11 sporočil ali manj. Največje število sporočil znotraj ene forumske teme je omejeno na 1.000 sporočil. Forumskih tem s 1.000 sporočili je 15 (

Tabela 2).

Slika 1 Število sporočil v času

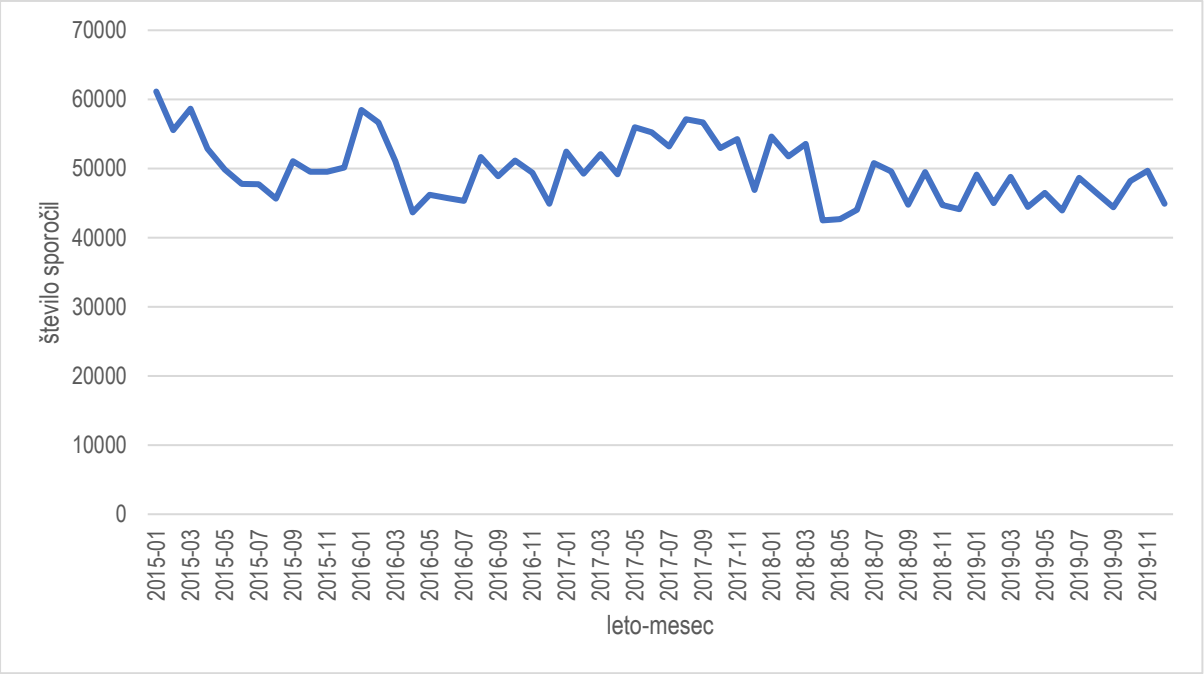


Tabela 1 Porazdelitev števila sporočil znotraj forumskih tem po forumih

Forum	*N*	*M*	*SD*	*MIN*	*Q1*	*Q2*	*Q3*	*MAX*
Zdravljenje s konopljo	116	103,50	68,86	1	32	112	141	198
Pediatrična in fetalna kardiologija	59	75,37	104,6	1	2	7	178	263
Neplodnost	1029	62,60	188,9	1	1	4	16	1000
Hujšanje	785	50,08	182,8	1	2	4	11	1000
Ženski in moški čvek	3336	47,30	113,1	1	3	9	37	1000
Opravljivka	522	42,39	53,44	1	10	26	56	360
Darovanje spolnih celic	74	24,89	33,91	1	2	4	34	97
Politični čvek	4194	23,16	35,92	1	4	10	27	269
Sex v mestu	2851	22,93	33,49	1	4	11	29	316
Finance in kripto valute	829	22,31	52,03	1	1	4	18	372
Estetsko zobozdravstvo	30	19,93	22,46	1	2	10,5	28,25	61
Srčno popuščanje	192	19,66	31,07	1	2	4	17,25	96
Mejna in narcistična osebnostna motnja - za svojce	663	17,53	22,24	1	4	9	23	114

Izgorelost	137	17,10	21,92	1	2	6	24	100
Starševski čvek	213798	15,83	32,06	1	3	7	16	1000
Prehrana in kulinarika	842	15,53	41,16	1	2	5	15	474
Nega, ličenje in style	1268	15,52	29,69	1	2	5	21	457
Ločitev, nezvestoba in ljubezenski nasveti	2661	15,08	20,96	1	3	8	18	201
Čustvena inteligenca, komunikacija in odnosi	1452	14,83	18,35	1	4	9	17	146
Kontracepcija	1055	10,79	27,93	1	2	3	6	140
Kam in kako na počitnice	4584	9,73	17,48	1	1	4	10	226
Kako živeti z rakom?	637	9,54	15,85	1	2	5	9	98
Starševstvo in vzgoja	2682	9,37	30,75	1	1	2	5	422
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	2259	8,00	20,72	1	2	4	6	185
Nosečnost in obporodno obdobje	1844	7,31	37,19	1	1	2	4	373
Nepremičnine	539	7,29	14,14	1	1	3	6	137
Pokojninsko zavarovanje	169	7,22	16,01	1	2	2	4	120
Cepljenja	711	7,17	9,73	1	2	4	8	54
Mladi, ljubezen in spolnost	2921	7,12	15	1	2	3	7	135
Otroško in mladinsko zobozdravstvo	116	7,04	16,84	1	1	2	3,25	71
Limfom in levkemija	253	7,03	12,89	1	2	3	5	52
Proktologija	186	6,78	7,98	1	2	3	7	39
Zasvojenost in pomoč	450	6,61	24,95	1	1,25	2	5,75	265
Gradimo, opremljamo in urejamo dom	1525	6,30	8,82	1	2	3	7	100
Posvojitev otroka in rejništvo	72	6,28	9,85	1	2	3	5	38
ARHIV - Sociala	311	6,24	11,84	1	1	2	5	99
ABC ginekologije in porodništva	10131	6,16	14,4	1	2	3	6	199
Otorinolaringologija	1997	6,04	11,44	1	2	2	5	77
Domače živali in veterina	1398	6,02	10,25	1	2	2	6	96
Avtomobilizem	851	5,88	9,27	1	1	2	6	90
Žalovanje in slovo	93	5,74	9,63	1	1	2	4	51
Glasba, film in fotografija	342	5,57	10,71	1	1	2	5	94
Življenje, ah to življenje	1042	5,47	7,81	1	1	3	6	74
ARHIV - Prosti čas	7	5,29	3,68	1	3	5	7,5	10
Kajenje	85	5,22	8,51	1	2	2	4	35
Pomoč voznikom v prometu	452	5,15	7,46	1	2	3	5	49
Multipla skleroza	330	5,11	10,15	1	1	2	4	64
Neonatologija	411	4,95	4,62	1	2	3	6	27
Revma	138	4,87	9,47	1	1	2	4	40
Slušni aparati in zaščita sluha	67	4,72	5,86	1	2	4	5	35
Kardiologija	1841	4,71	5,29	1	2	3	5	45
Vrnitev k sebi	223	4,71	4,96	1	2	4	5	28
Knjižni molji in pravopis	1013	4,67	7,81	1	1	2	5	90
Depresija, anksioznost in razpoloženjske motnje	552	4,58	3,85	1	2	3	6	23
Motnje hranjenja	128	4,43	5,95	1	2	2	4	48
Delovno pravo in zaposlitev	1602	4,42	11,04	1	1	2	3	91
Ustno zdravje in higiena	118	4,38	4,32	1	2	2	5,75	29
Ortopedija	1381	4,28	4,33	1	2	2	4	34
Nevrokirurgija	416	4,23	4,09	1	2	2	4	23

Starejši so modrejši	276	4,18	7,25	1	1	2	4	79
Zavarovanja	196	4,16	5,54	1	2	2	4	28
Iščete delo, delavca?	2817	4,16	9,99	1	1	1	3	112
Partnerska in družinska posvetovalnica	977	4,08	3,51	1	2	3	5	25
Duševno zdravje	227	4,06	5,78	1	2	2	4	31
Psihiater odgovarja	1565	3,99	4,6	1	2	2	4	37
Podpora pri uporabi forumov OverNet	204	3,96	4,82	1	1	2	5	29
Za ljudi z mejno osebnostno motnjo	21	3,95	2,5	1	2	4	5	9
Računalništvo, telefonija, internet in televizija	892	3,93	6,28	1	1	2	4	72
Težave in bolečine v hrbtenici	546	3,89	3,61	1	2	2	4	32
Mali oglasi	4499	3,85	14,46	1	1	1	2	215
Računovodstvo	476	3,82	5,07	1	2	2	4	48
ARTHRON - Zdravstveno svetovanje v športu	123	3,80	3,44	1	2	2	4	18
Hormonske motnje	1367	3,71	5,07	1	2	2	4	41
Urgenca	3539	3,70	3,41	1	2	2	4	28
Endodontija	91	3,60	3,51	1	1,5	2	5	14
Homeopatija	367	3,58	4,54	1	2	2	4	29
Alergija	496	3,56	6,04	1	1	2	3	33
Oftalmologija	371	3,53	4,11	1	2	2	4	26
Odškodnine in zavarovalnine	224	3,44	4,6	1	1	2	3	29
Dermatologija	2111	3,43	3,3	1	2	2	4	47
Zdravstvo pri nas?	318	3,39	4,22	1	1	2	3	18
Avtomobilska zavarovanja	87	3,37	3,58	1	2	2	3	28
Vrtičkarstvo, zelišča in urejanje okolice	232	3,34	5,15	1	1	2	3	43
Fizioterapija in masaže	485	3,33	2,26	1	2	2	4	18
Za žalujoče po izgubi zaradi samomora	22	3,32	1,46	1	2	4	4	6
Kronična bolečina - fibromialgija	337	3,25	3,24	1	1	2	4	20
Za ženske v stiski	162	3,19	2,74	1	2	2	4	19
ARHIV - Alternativa in zdravo življenje	257	3,19	5,31	1	1	2	3	30
Astma in druge pljučne bolezni	494	3,18	2,38	1	2	2	4	16
Pediatrija	2723	3,18	2,48	1	2	2	4	22
Poroka	160	3,16	3,66	1	1	2	4	28
Laserska korekcija vida	62	3,15	2,05	1	2	2	4	10
Psihoterapija	416	3,12	2,59	1	2	2	4	19
Zobna protetika	206	3,02	1,72	1	2	2	4	8
Rekreacija, oblikovanje telesa in zdrava prehrana	548	3,01	2,59	1	1	2	4	16
Ortodontija - čeljustna in zobna ortopedija	232	2,96	2,17	1	2	2	4	12
Estetska in plastična kirurgija	715	2,96	3,05	1	2	2	3	27
Intimna nega in zdravljenje vaginalnih okužb	716	2,93	2,6	1	2	2	3	23
Manualna medicina - zdravljenje gibalnega sistema	1219	2,92	1,71	1	2	2	3	11
Splošna stomatologija	527	2,89	1,92	1	2	2	4	13
Parodontologija	110	2,80	1,74	1	2	2	3	10
Pravna pomoč bolnikom z rakom	144	2,79	1,46	1	2	2	3	8
ARHIV - Zdravila in zakonodaja	120	2,74	1,63	1	2	2	3	9
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	268	2,72	1,71	1	1	2	4	8
Prehrana in varnost živil	126	2,70	2,68	1	2	2	3	18

ARHIV - Zdravstvo	511	2,65	3,53	1	1	2	3	30
Seksologija - želiva si spolnosti	188	2,62	1,48	1	2	2	3	8
Skrinjica idej!	395	2,60	3,41	1	1	1	3	26
Družinska medicina	3319	2,49	2,45	1	1	2	2	37
Preventivni pregled dojk	148	2,43	1,12	1	2	2	2	6
Farmacevt svetuje	753	2,32	1,36	1	2	2	2	9
Maksilofacialna in oralna kirurgija	247	2,31	2,03	1	1	2	3	14
Genetika v nosečnosti in NIPT	335	2,29	1,45	1	1	2	3	9
Zdravljenje obolenj in deformacij stopala	197	2,25	1,5	1	1	2	2	9
Bolezni prebavil (gastroenterologija)	532	2,24	1,44	1	1	2	3	11
Optika	487	2,22	1,76	1	1	2	2	12
Žilna kirurgija	111	2,20	2,15	1	1	1	2	9
Klimakterij -menopavza	237	2,15	1,44	1	1	2	3	9
Zdravilstvo	264	2,14	1,3	1	1	2	2	6
ARTHRON - fizioterapija	184	2,09	1,16	1	1	2	2	7
Implantologija in zobna protetika	31	2,06	0,81	1	2	2	2	4
Prostovoljna zdravstvena zavarovanja	77	2,05	0,89	1	2	2	2	5
Elektromagnetna terapija	46	1,87	1,11	1	1	2	2	6
Abdominalna kirurgija	211	1,83	1,53	1	1	1	2	7
Radiologija	179	1,53	0,84	1	1	1	2	4
Urologija	168	1,52	3,7	1	1	1	1	35
Pravni nasveti	1811	1,41	1,01	1	1	1	1	9
Varna pot	8	1,12	0,35	1	1	1	1	2

Tabela 2 Forumske teme z natanko 1.000 sporočili

Forum	*Forumska tema*
Ženski in Moški čvek	Zakaj si Nemci želijo toliko beguncev
Starševski čvek	Žlahtna štorija
Starševski čvek	Skrivnost zdravja in lepote
Starševski čvek	štepajte youtube muziko za danes, kaj vam je pri srcu?
Starševski čvek	Kakšno je vaše mnenje o feministkah in o feminizmu na splošno
Starševski čvek	Maja Hausman - pomaga revežem
Neplodnost	IUI/IVF/ICSI SEPTEMBER 2017
Starševski čvek	samo pesmi (zapete) v slovenščini
Starševski čvek	The biggest loser
Hujšanje	Motivacija za vsak dan!
Hujšanje	Ali kdo hujša z luninimi menami?
Neplodnost	ivf 2018
Starševski čvek	Kako ste oblečeni danes?
Neplodnost	IVF/ICSI/IUI 2019
Starševski čvek	Zdravilo za Krisa

4.1 Pregled števila sporočil po skupinah forumov

Forume na MedOverNet je mogoče razvrstiti v tri kategorije: zdravstvene posvetovalnice, podporne skupine ter družabni forumi. Dodali smo tudi skupino »drugo«, ki označuje nekatere arhivirane forumske teme ali teme, ki jih ni mogoče enoznačno razvrstiti v ostale kategorije. Uvrstitev posameznih forumov si je mogoče ogledati v Tabela 11. Glede na vrsto foruma je povprečno število sporočil znotraj forumskih tem najvišje v skupini družabnih forumov, sledijo podporne skupine in zdravstvene posvetovalnice.

Tabela 3 Število sporočil po forumskih temah, glede na vrsto forumov

Vrsta foruma	*N*	*M*	*SD*	*MIN*	*Q1*	*Q2*	*Q3*	*MAX*
Zdravstvena posvetovalnica	52425	5.37	25.88	1	2	2	4	1000
Podporna skupina	15194	13.71	56.16	1	2	3	9	1000
Družabni forum	253455	15.50	33.72	1	3	6	16	1000
Drugo	1206	3.71	7.07	1	1	2	3	99

5 Določanje teme raka v sporočilih

Za določanje prisotnosti teme raka v posameznem sporočilu so bili uporabljeni različni pristopi nadzorovanega strojnega učenja. Ker gre za analizo besedil, je bila opravljena tudi priprava besedil, ki je vključevala lematizacijo, odstranjevanje ključnih besed in druge standardne postopke. Zaradi izjemno velike količine besedila so nekateri koraki analize računsko in časovno zelo zahtevni, zato so nekoliko poenostavljeni. Kljub temu je učinkovitost uporabljenih pristopov ustrezna.

5.1 Korpus

Kot omenjeno v poglavju Podatki, je korpus je sestavljen iz vseh sporočil v spletni zdravstveni skupnosti MedOverNet, ki so bila objavljena v letih med vključno 2015 in vključno 2019. Gre za 2.982.564 sporočil, ki so bila objavljena v 280.357 forumskih temah. Enoto analize v postopku strojnega učenja predstavlja sporočilo.

5.2 Priprava besedila

Priprava besedila je zajemala čiščenje besedila (iz besedila so bile odstranjene hiperpovezave, HTML značke, ločila, števila ter drugi znaki, kot so %, &, € in podobni), lematizacijo (postopek, s katerim določimo osnovno obliko besede), tokenizacijo (razdelitev besedila na posamezne enote) ter odstranitev stop besed (stop besede so besede, ki se v vseh besedilih pojavljajo z isto verjetnostjo in so brez semantične vrednosti).

Za lematizacijo je bil uporabljen označevalnik Obeliks (Grcar, Krek, and Dobrovoljc 2012), katerega natančnost pripisovanja osnovnih oblik besed je 98,55 % brez upoštevanja velikih začetnic. Po opravljenem postopku lematizacije so bila sporočila razdeljena na manjše dele in sicer na besede. Stop besede so bile odstranjene na podlagi seznama stop besed (seznam stop besed je enak kot je bil uporabljen v Erčulj (2020) za napovedovanje izhoda zdravljena neplodnosti na forumu Neplodnost v spletni zdravstveni skupnosti MedOverNet), odstranjene pa so bile tudi besede, ki se v celotnem korpusu besed ne pojavijo več kot petkrat (takih besed je bilo 1.520).

Sporočila so bila predstavljena v obliki matrike dokumentov in značilk (ang. *document term matrix*) (Deerwester et al. 1990), kjer je vsako sporočilo v svoji vrstici, besede pa so predstavljene po stolpcih. Uteži besed (v celicah matrike) so predstavljale frekvence besed v posameznem sporočilu (ang. *Term Frequency*, TF). Obstajajo tudi drugi pristopi za določitev uteži, kot je inverzna frekvenca besed v dokumentih (ang. *inverse document frequency factor*, IDF) (Jones 1972; Robertson 2004) ali verjetnostno utežena količina informacije (ang. *Probability-Weighted Amount of Information*, PWI) (Aizawa 2003), a v pričujoči analizi niso bili uporabljeni, saj so bili rezultati, z upoštevanjem TF zadostni za namene analize.

5.3 Uporabljene metode nadzorovanega strojnega učenja

Za določevanje prisotnosti teme raka so bili uporabljene naslednje metode: naivni Bayesov klasifikator, metoda podpornih vektorjev, odločitvena drevesa in slučajni gozdovi. Vse metode sodijo v sklop metod nadzorovanega strojnega učenja in so v nekaterih drugih študijah že bile uporabljene v kontekstu analize besedil. Sledijo kratki opisi uporabljenih metod.

5.3.1 Naivni Bayesov klasifikator (NBC)

Naivni Bayesov klasifikator (ang. *Naive Bayes classifier*) sodi med enostavnejše pristope za razvrščanje oziroma napovedovanje izidov in je pogosto uporabljen na področjih strojnega učenja. Glavne prednosti klasifikatorja sta relativno kratek izvedbeni čas (Liu et al. 2013) in lahka implementacija (Kononenko and Kukar 2007). Na podlagi Bayesove formule, se za vsako skupino (na primer, »je o raku«, »ni o raku«) določi pogojno verjetnost (glede na dejanske podatke) (Hand et al. 2001)

$$p(c_k|x) \propto p(x|c_k)p(c_k) = p(c_k)\prod_{j=1}^{p} p(x_j|c_k)$$

kjer c_k označuje k-to skupino, x pa je vektor dejanskih podatkov (npr. število posameznih besed) dolžine p. Velja $1 \leq k \leq m$, kjer je m skupno število skupin. Na podlagi dobljenih pogojnih verjetnosti se pripiše pripadnost skupini z največjo verjetnostjo.

Osnovna predpostavka naivnega Bayesovega klasifikatorja je, da so vse spremenljivke pogojno neodvisne, glede na skupino (Hand et al. 2001; Kononenko and Kukar 2007). V primeru klasificiranja sporočil (glede na prisotnost teme raka) to pomeni, da je sopojavljanje različnih besed neodvisno znotraj sporočil, ki vsebujejo temo o raku in ostalimi. Kljub temu, da je predpostavka o pogojni neodvisnosti v praksi pogosto kršena, se naivni Bayesov klasifikator pogosto izkaže kot primerljivo učinkovit z bolj zahtevnimi pristopi (Hand et al. 2001; Ting, Ip, and Tsang 2011).

5.3.2 Metoda podpornih vektorjev (SVM)

Metoda podpornih vektorjev (ang. *Support Vector Machine*) (Basu, Walters, and Shepherd 2003; Wang 2005) je pogosto uporabljena predvsem zaradi relativno nizke računske zahtevnosti. V razumnem času je mogoče analizirati podatke z več kot 300.000 dimenzijami, kar lahko v nekaterih primerih odpravi potrebo po lematizaciji besedila ali odstranitvi stop besed (Leopold and Kindermann 2002).

Metoda deluje tako, da poišče tako hiper ravnino v n-dimenzionalnem prostoru (kjer je n število opazovanih lastnosti skupine, v primeru identifikacije sporočila z vsebino raka so to upoštevane besede), da je razdalja med točkami, ki pripadajo (najbližjim) različnima razredoma (podpornimi vektorji), in hiper ravnino največja. Z uporabo jedrne funkcije je mogoče predpostaviti, ne le linearno razmejitev, pač pa tudi nelinearno razmejitev med skupinama. Nekatere študije kažejo večjo učinkovitost metode podpornih vektorjev, ko je predpostavljena linearna razmejitev, v primerjavi z nelinearno razmejitvijo (Zhang, Yoshida, and Tang 2008), spet drugi poročajo, da ima večji vpliv na kakovost rešitve metode podpornih vektorjev način določitve matrike dokumentov in značilk, kakor vrsta jedrne funkcije (Leopold and Kindermann 2002). V konkretnem primeru je bila uporabljena linearna jedrna funkcija.

5.3.3 Odločitvena drevesa (TREE)

Osnovni princip metode odločitvenih dreves (ang. *Decision Trees*) je razvrščanje enot v skupine na način, da so si enote znotraj istih skupin čim bolj podobne, glede na vrednost izbrane spremenljivke. Enote se razvrščajo v skupine glede na vrednosti nekih dodatnih spremenljivk v več korakih. Na vsakem koraku se upošteva spremenljivka, ki najbolj ločuje med skupinami izbrane spremenljivke. V odločitvenem drevesu je lahko toliko ravni, kolikor je enot z različnimi vrednostmi pri dodatnih

spremenljivkah (Hand et al. 2001). Model odločitvenih dreves dobro deluje tudi v primeru podatkov s kompleksno strukturo (ko so prisotne interakcije ter nelinearne povezave) (Williams 2011). Obstaja več algoritmov za gradnjo odločitvenega drevesa. V našem primeru je bil uporabljan algoritem CART (Breiman et al. 1984; Breiman and Shang 1996). Kot kriterij za izbiro spremenljivk na vsakem koraku gradnje drevesa je bil uporabljen Ginijev koeficient. Obrezovanje[1] (Patil, Wadhai, and Gokhale 2010) drevesa ni bilo opravljeno.

5.3.4 Slučajni gozdovi (FOREST)

Slučajni gozdovi (ang. *Random Forests*) (Breiman 2001) temeljijo na množici odločitvenih dreves. Dobljeni so s postopkom ponovnega vzorčenja: iz osnovnega vzorca ponovno vzorčimo (s ponavljanjem) mnogo vzorcev in za vsak vzorec določimo odločitveno drevo. Dano enoto, na podlagi dobljenih odločitvenih dreves, razvrstimo v tisto skupino, v katero bi bila uvrščena na podlagi največ odločitvenih dreves (po načelu večinskega glasu, ang. *Majority Vote approach*). Namen gradnje odločitvenih dreves je izboljšati stabilnost in natančnost odločitvenih dreves ter zmanjšati problem prekomernega prileganja učni množici podatkov (Tibshirani, Walther, and Hastie 2001). Nekateri avtorji poročajo o visoki stopnji natančnosti odločitvenih dreves na primerih analize besedil (Chen and Ho 1999; Kılınç et al. 2017; Xue and Li 2015). Razvite so tudi posebne različice slučajnih gozdov, ki z uporabo metode podpornih vektorjev rešujejo problem neuravnoteženih vzorcev (Wu et al. 2014) (v tej analizi niso bile uporabljene).

5.4 Učna množica

Osnovni pogoj za uporabo metod nadzorovanega strojnega učenja je obstoj tako imenovane učne množice, ki služi učenju algoritmov strojnega učenja. Učna množica vsebuje izbrana sporočila ter podatek o kategoriji sporočila (torej o tem, ali izbrano sporočilo vsebuje temo raka ali ne).

Tabela 4 Seznam ključnih besed, uporabljenih za določitev učne množice sporočil

benigen	radiolog	biološko zdravilo	prostata/prostate	levkemija
biopsija	marker	konopljina smola	zatrdlina	limfom
mozeg	maligen	rak grla	samopregledovanje	metastaza
rak	radioterapija	HPV	mamografija	onkolog
karcinom	transplantacija	rak danke	punkcija	obsevanje
kemoterapija	mozeg	rak mod	stadij	sarkoma
hormonska terapija	dojka	konzilij	progestini	tumor
imunoterapija	kemoprevencija	blastom	BRCA	rekonstrukcija dojk
rak na prsih	celice	genska terapija	hospic	mamogram
osteosarkom	rak dojk	rak pljuč	rak maternične vratu	rak debelega črevesa
PSA	rak kože	kožni rak	hodgkinov	

Učno množico predstavlja 368 slučajno izbranih sporočil ter 298 ročno izbranih sporočil, ki so poiskana z uporabo ključnih besed, za katere smo domnevali, da se pogosteje pojavljajo v sporočilih s temo o raku. Oba pristopa sta že bila uporabljena v sorodnih raziskavah (Cousineau, Rancourt, and Green

[1] Dobljeno odločitveno drevo z upoštevanjem vseh enot in vseh spremenljivk je lahko izjemno kompleksno v primeru velikega števila spremenljivk, prav tako pa so nižji deli drevesa pogosto nezanesljivi. Problem se običajno rešuje s tako imenovanim obrezovanjem, pri katerem določimo ravni, ki jih bomo upoštevali. S tem omejimo velikost drevesa, odpravimo posledice prekomernega prileganja ter dosežemo bolj enostavno rešitev.

2006; Hoffman-Goetz, Donelle, and Thomson 2009; Hwang et al. 2010; Macias, Lewis, and Smith 2005; Mo and Coulson 2008).

Za iskanje sporočil z uporabo ključnih besed (Tabela 4) je bil uporabljen spletni iskalnik na spletnem mestu MedOverNet. Upoštevana so bila tako sporočila z vsebino o raku, kakor tudi sporočila, ki ne vsebujejo temo raka. Določevanje učne množice sporočil sta izvedla dva ocenjevalca. Vsako sporočilo je določil en ocenjevalec, preverjanje zanesljivosti zaradi enostavnosti problema ni bilo potrebno.

Ročno je bilo izbranih 298 sporočil, od tega jih je 199 (67 %) vsebovalo temo raka. Slučajno je bilo izbranih dodatnih 368 sporočil, od katerih jih večina (367 sporočil) ni vsebovala teme raka (Tabela 5).

Tabela 5 Velikost učne množice sporočil (način vključitve in vsebina)

	Število sporočil o raku	*Število sporočil, ki niso o raku*	*Skupaj*
Slučajno izbrana množica sporočil	1	367	368
Množica sporočil, izbrana na podlagi ključnih besed o raku	199	99	298
Skupaj	200	466	666

5.5 Evalvacija rezultatov metod za uvrščanje v skupine

Rezultati napovedi oziroma razvrščanja sporočil med sporočila, ki vsebujejo temo raka ter ostale, so evalvirani po metodi navzkrižnega preverjanja. Pri navzkrižnem preverjanju (CV-10) podatke slučajno razdelimo na 10 približno enakih delov. Nato ocenimo model z upoštevanjem vseh, razen enega dela. Na podlagi tako ocenjenih modelov napovemo prisotnost teme raka za sporočila, ki pripadajo delu, ki ni bil uporabljen za oceno modelov. Enak postopek ponovimo na vseh desetih delih podatkov.

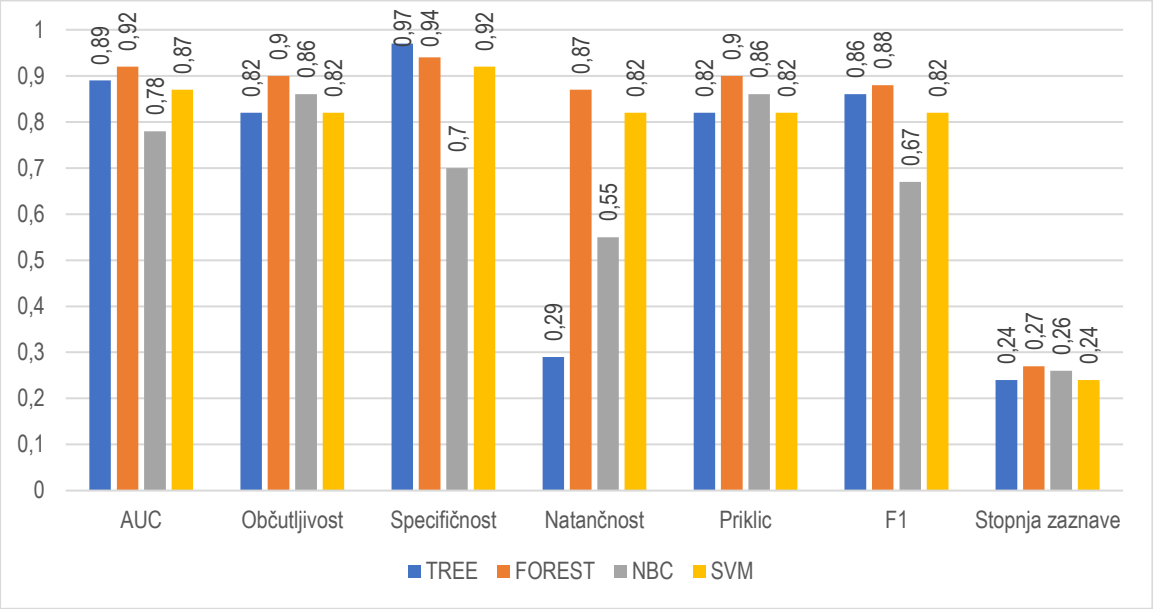
Prednost CV-10 je v tem, da so vrednosti ocenjenih parametrov kakovosti modelov nepristranske, saj modeli niso evalvirani na isti množici podatkov, kot je bila uporabljena za oceno modelov.

Za evalvacijo modelov so uporabljene naslednje mere:

- AUC: ploščina pod krivuljo;
- Občutljivost: delež pravilno napovedanih sporočil z rakom izmed vseh sporočil z rakom;
- Specifičnost: delež pravilno napovedanih sporočil brez raka izmed vseh sporočil brez raka;
- Natančnost: delež pravilno napovedanih sporočil z rakom izmed vseh napovedanih sporočil z rakom;
- Priklic: zaznan delež sporočil z rakom (izmed vseh sporočil z rakom);
- F1: harmonična sredina priklica in natančnosti;
- Stopnja zaznave: delež pravilno zaznanih sporočil z rakom izmed vseh sporočil.

Kot najboljši se izkaže model slučajnih gozdov, sicer z zanemarljivo nižjo specifičnostjo (v primerjavi z modelom odločitvenih dreves). Model odločitvenih dreves daje v splošnem dobre rezultate, a ima izjemno nizko stopnjo natančnosti, model podpornih vektorjev prav tako daje relativno dobre rezultate, medtem ko je naivni Bayesov klasifikator najslabši, glede na vse opazovane mere.

Slika 2 Evalvacija modelov (CV-10)



* prevalenca števila sporočil s temo raka je 30 %

Različni modeli dajejo v splošnem zelo podobne rezultate (z izjemo naivnega Bayesovega klasifikatorja), še posebej modeli, ki temeljijo na odločitvenih drevesih (torej model odločitvenih dreves in model slučajnih gozdov) (Tabela 6).

Tabela 6 Vrednosti Pearsonovega koeficienta korelacije med dobljenimi rešitvami (CV-10)

	TREE	*FOREST*	*SVM*	*NB*
TREE	1.0	0.8	0.7	0.5
FOREST	0.8	1.0	0.8	0.6
SVM	0.7	0.8	1.0	0.5
NB	0.5	0.6	0.5	1.0

Ko model slučajnih gozdov evalviramo s pristopom CV-10, za 21 sporočil napačno napovemo, da ne vsebujejo teme raka in za 27 sporočil napačno napovemo, da vsebujejo temo raka. Glede na vrednosti povprečnega znižanja natančnosti (ang. *Mean Decreased Accuracy*) in povprečnega znižanja vrednosti Gini indeksa (ang. *Mean Decrease Gini Index*), med najpomembnejše besede, ki ločujejo med sporočili o raku in ostalimi, sodijo rak, tumor, kemoterapija, zdravljenje, bezgavka, bolezen, maligen, dojka, zdravnik, operacija, preiskava, limfom, pregled, diagnoza, možnost, karcinom, bolnik, pomeniti, obsevanje, levkemija, celica, metastaza in druge.

Tabela 7 Klasifikacijska tabela za model odločitvenih dreves (CV-10)

		Dejanske vrednosti	
		Ne vsebuje teme raka	Vsebuje temo raka
Napoved	Ne vsebuje teme raka	439	21
	Vsebuje temo raka	27	179

5.6 Ponovno treniranje modela

Kljub relativno dobrim rezultatom odločitvenih dreves smo povečali učno množico sporočil za treniranje algoritmov strojnega učenja. Namen povečanja učne množice je bil zmanjšati število napačno določenih sporočil, kot sporočil, ki vsebujejo temo raka.

V nadaljnjih analizah nismo upoštevali hrvaških forumov (Mršavljenje, Virtualni laboratorij kemije, Kutak za biologiju, Ultrazvuk u ginekologiji, Matične stanice, Neplodnost, Implantologija i zubna protetika Nekirurška estetika lica, Djeca s posebnim potrebama, Obiteljska medicina) in testnih forumov (TEST - Testni forum 5, TEST - Minister za zdravje obvešča TEST - Testni Forum 2, TEST - Testni forum 3).

Za ponovno treniranje modela smo izbrali sporočila po naslednjih kriterijih:

- vsa sporočila, ki vsebujejo temo raka in se nahajajo v forumskih temah, ki imajo 25 ali manj sporočil, ki se nanašajo na temo raka (464 sporočil),
- vsa sporočila, ki niso označena kot sporočila s temo o raku, a se nahajajo v forumskih temah, kjer je več kot 75 % sporočil s temo raka (68 sporočil),
- 480 slučajno izbranih sporočil s temo raka (480 sporočil),
- 25 slučajno izbranih sporočil s temo raka iz vsakega izmed naslednjih forumov: Kako živeti z rakom, Preventivni pregled dojk, Limfom in levkemija, Pravna pomoč bolnikom z rakom (100 sporočil).

Sporočila s temo raka so bila določena na podlagi modela slučajnih gozdov. Skupaj je bilo izbranih 1.100 sporočil (vsota zgoraj navedenega števila izbranih spročil je 1.112, a so bila nekatera sporočila izbrana večkrat), a je bilo zaradi omejenih resursov kodiranih zgolj 621 slučajno izbranih sporočil.

Tabela 8 Primerjava razvrstitev sporočil (model slučajnih gozdov vs. ročna klasifikacija)

		Ročna klasifikacija		
		Vsebuje temo raka	Ne vsebuje teme raka	Skupaj
Klasifikacija glede na model slučajnih gozdov	Vsebuje temo raka	250	337	587
	Ne vsebuje teme raka	18	16	34
	Skupaj	268	353	621

Ker so bila pravila za izbor sporočil določena na način, da so bile izbrana sporočila ki imajo višjo verjetnost, da so bile na podlagi modela slučajnih gozdov razvrščene napačno na podlagi zgornje tabele (Tabela 8) ni mogoče ovrednotiti kakovosti modela slučajnih gozdov. Pri ročnem klasificiranju sporočil se je sicer pokazalo, da se je veliko sporočil nanašalo na zobozdravstvene težave. Za tovrstna sporočila je značilno, da vsebujejo besede, kot je računalniška tomografija in druge diagnostične pristope, ki se uporabljajo tudi pri odkrivanju bolezni raka. Pogost izraz v tovrstnih sporočilih je tudi cista.

Ročno klasificirana sporočila smo dodali k prvotni učni množici sporočil za treniranje modelov. Posodobljeni modeli so tako ocenjeni na skupno 1.287 sporočil (468 sporočil, ki vsebujejo temo raka in 819 sporočil, ki ne vsebujejo teme raka).

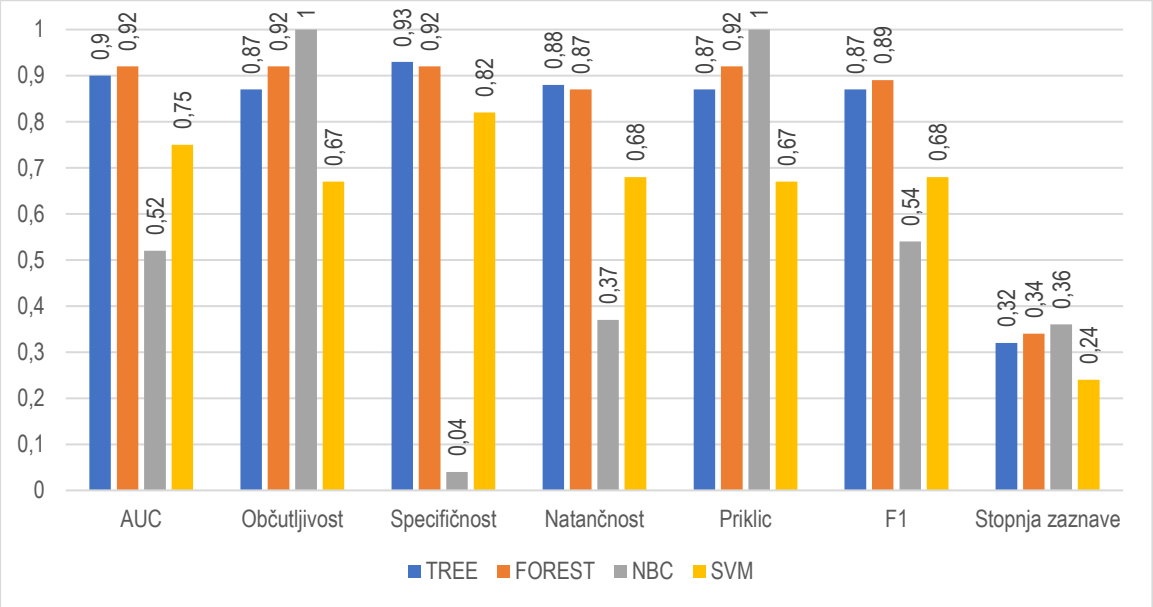
Tako pri evalvaciji prvotnega, kakor tudi posodobljenih modelov so upoštevana vsa sporočila (razen sporočila v forumih, ki niso upoštevana pri evalvaciji posodobljenega modela slučajnih gozdov), tudi

zelo kratka sporočila, kar lahko zniža (podceni) vrednosti mer za evalvacijo modelov strojnega učenja, če bi naknadno kot veljavna sporočila za analizo upoštevali zgolj sporočila z več znaki ali forumske teme z več sporočili.

5.7 Evalvacija posodobljenih modelov

Na podlagi izbranih mer za evalvacijo modelov lahko zaključimo, da se je z upoštevanjem dodatnih sporočil kakovost modela slučajnih gozdov zanemarljivo izboljšala. Zelo se je sicer poslabšal rezultat naivnega Bayesovega klasifikatorja, ki sodeč po merah občutljivosti in specifičnosti pretirano teži k uvrščanju sporočil med sporočila s temo raka.

Slika 3 Evalvacija modelov (CV-10) ocenjenih na podlagi dopolnjene učne množice sporočil



* prevalenca števila sporočil s temo raka je 36 %

Z uporabo posodobljenega modela slučajnih gozdov smo na novo prepoznali 6.183 sporočil, kot sporočil s temo raka, za 12.202 sporočil pa smo, na podlagi posodobljenega modela določili, da ne vsebujejo teme raka, čeprav so bile prej določene kot sporočila s temo raka. 99 % sporočil smo klasificirali enako na podlagi obeh modelov (Tabela 9).

Tabela 9 Primerjava rešitev prvotnega in posodobljenega modela slučajnih gozdov

		Posodobljen model		
		Ne vsebuje teme raka	Vsebuje temo raka	Skupaj
Prvotni model	Ne vsebuje teme raka	2950002	6183	2956185
	Vsebuje temo raka	12202	11518	23720
	Skupaj	2962204	17701	2979905

Tabela 10 Število sporočil s temo raka po forumih: primerjava prvotnega in posodobljenega modela

Forum	*Prvotni model*	*Posodobljen model*	*Sprememba*
ABC ginekologije in porodništva	2150	1030	-1120
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	849	45	-804
Manualna medicina - zdravljenje gibalnega sistema	689	24	-665
Kardiologija	511	27	-484
Družinska medicina	626	294	-332
Pediatrija	411	102	-309
Urgenca	607	366	-241
Ortopedija	270	41	-229
Otorinolaringologija	497	304	-193
Hormonske motnje	301	109	-192
Psihiater odgovarja	190	40	-150
Bolezni prebavil (gastroenterologinja)	230	81	-149
Dermatologija	297	166	-131
Multipla skleroza	130	38	-92
Težave in bolečine v hrbtenici	98	11	-87
Srčno popuščanje	88	4	-84
Proktologija	115	34	-81
Neonatologija	93	14	-79
Kronična bolečina - fibromialgija	96	21	-75
Nevrokirurgija	96	28	-68
Neplodnost	91	27	-64
Mejna in narcistična osebnostna motnja - za svojce	75	13	-62
Partnerska in družinska posvetovalnica	70	14	-56
Astma in druge pljučne bolezni	109	53	-56
Depresija, anksioznost in razpoloženjske motnje	75	21	-54
Homeopatija	62	9	-53
ARTHRON - Zdravstveno svetovanje v športu	54	2	-52
ARHIV - Zdravstvo	89	39	-50
Genetika v nosečnosti in NIPT	56	11	-45
Farmacevt svetuje	57	14	-43
Cepljenja	78	35	-43
Endodontija	46	3	-43
Maksilofacialna in oralna kirurgija	63	21	-42
Intimna nega in zdravljenje vaginalnih okužb	73	32	-41
Fizioterapija in masaže	44	6	-38
Abdominalna kirurgija	52	15	-37
Alergija	45	9	-36
Ortodontija - čeljustna in zobna ortopedija	37	1	-36
Splošna stomatologija	44	8	-36
Oftalmologija	39	5	-34
Starševstvo in vzgoja	54	21	-33
Politični čvek	96	65	-31
ARHIV - Zdravila in zakonodaja	35	5	-30
Čustvena inteligenca, komunikacija in odnosi	55	26	-29

Zdravljenje obolenj in deformacij stopala	29	0	-29
Parodontologija	30	2	-28
Izgorelost	34	6	-28
Revma	28	1	-27
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	25	0	-25
ARTHRON - fizioterapija	29	5	-24
Zdravilstvo	30	8	-22
Domače živali in veterina	136	114	-22
Zasvojenost in pomoč	26	5	-21
Odškodnine in zavarovalnine	20	0	-20
Klimakterij -menopavza	39	19	-20
Otroško in mladinsko zobozdravstvo	21	2	-19
Psihoterapija	24	6	-18
Radiologija	42	25	-17
Darovanje spolnih celic	17	1	-16
Ustno zdravje in higiena	18	3	-15
Optika	17	2	-15
Pediatrična in fetalna kardiologija	14	1	-13
Žilna kirurgija	17	4	-13
ARHIV - Sociala	16	4	-12
Urologija	45	33	-12
Limfom in levkemija	191	180	-11
Zobna protetika	13	2	-11
Pravni nasveti	16	5	-11
Mali oglasi	18	8	-10
Rekreacija, oblikovanje telesa in zdrava prehrana	9	0	-9
Motnje hranjenja	10	1	-9
Starejši so modrejši	9	0	-9
Vrnitev k sebi	9	1	-8
Zavarovanja	8	2	-6
Zdravstvo pri nas?	14	8	-6
Seksologija - želiva si spolnosti	8	2	-6
Hujšanje	32	26	-6
Ločitev, nezvestoba in ljubezenski nasveti	28	22	-6
Prostovoljna zdravstvena zavarovanja	7	2	-5
Pomoč voznikom v prometu	5	0	-5
Za ženske v stiski	5	0	-5
ARHIV - Alternativa in zdravo življenje	14	9	-5
Za ljudi z mejno osebnostno motnjo	4	0	-4
Delovno pravo in zaposlitev	8	5	-3
Knjižni molji in pravopis	10	7	-3
Računalništvo, telefonija, internet in televizija	3	0	-3
Duševno zdravje	4	1	-3
Slušni aparati in zaščita sluha	3	1	-2
Iščete delo, delavca?	2	0	-2
Gradimo, opremljamo in urejamo dom	2	0	-2

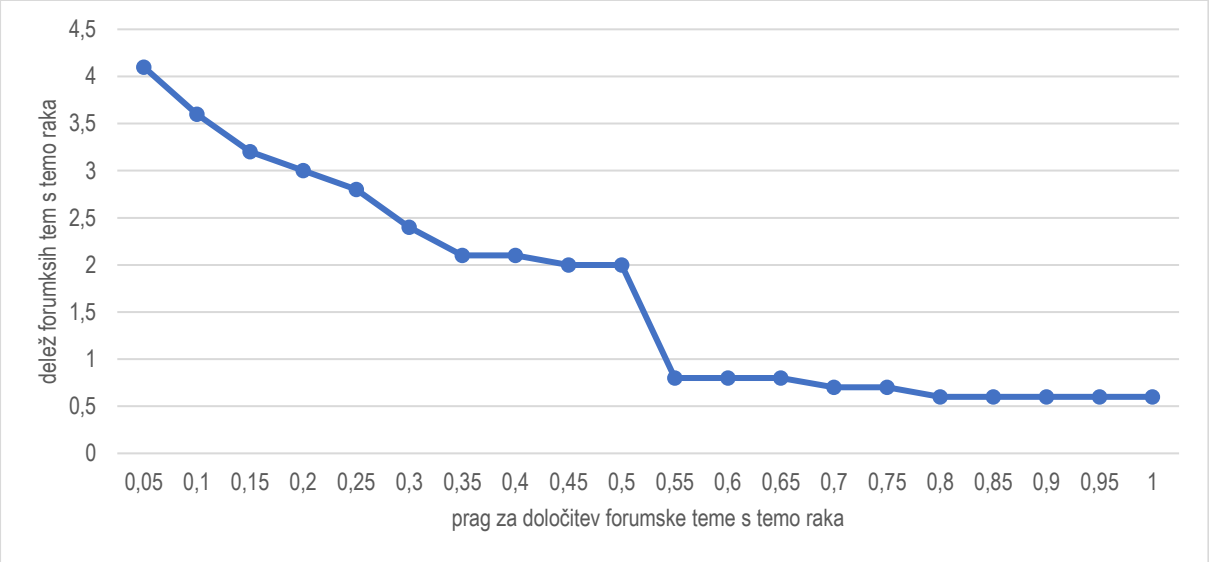
Avtomobilizem	2	0	-2
Žalovanje in slovo	7	6	-1
Finance in kripto valute	1	0	-1
Kontracepcija	31	30	-1
Elektromagnetna terapija	2	1	-1
Za žalujoče po izgubi zaradi samomora	1	0	-1
Implantologija in zobna protetika	1	0	-1
Estetsko zobozdravstvo	1	0	-1
Skrinjica idej!	0	0	0
Avtomobilska zavarovanja	0	0	0
Posvojitev otroka in rejništvo	0	0	0
Računovodstvo	0	0	0
Kajenje	3	3	0
Varna pot	0	0	0
Pokojninsko zavarovanje	1	1	0
Podpora pri uporabi forumov Over.Net	1	1	0
ARHIV - Prosti čas	0	0	0
Laserska korekcija vida	0	0	0
Poroka	0	0	0
Glasba, film in fotografija	0	0	0
Nepremičnine	0	1	1
Prehrana in varnost živil	3	4	1
Pravna pomoč bolnikom z rakom	62	63	1
Vrtičkarstvo, zelišča in urejanje okolice	3	4	1
Mladi, ljubezen in spolnost	12	15	3
Kam in kako na počitnice	15	18	3
Življenje, ah to življenje	22	28	6
Prehrana in kulinarika	14	21	7
Sex v mestu	50	63	13
Nega, ličenje in style	5	19	14
Zdravljenje s konopljo	166	181	15
Estetska in plastična kirurgija	53	68	15
Opravljivka	43	60	17
Ženski in moški čvek	86	106	20
Nosečnost in obporodno obdobje	103	138	35
Preventivni pregled dojk	90	142	52
Kako živeti z rakom?	730	797	67
Starševski čvek	11280	12174	894

6 Opredelitev forumske teme, ki vključuje temo raka

Znotraj posameznih forumskih tem se lahko pojavijo sporočila, ki vsebujejo temo raka in sporočila, ki ne vsebujejo teme raka. Eden izmed načinov za določanje prisotnosti teme raka na ravni forumske teme je, da določimo delež sporočil s temo raka znotraj forumske teme, na podlagi katerega določimo ali se posamezna forumska tema nanaša na temo raka ali ne.

Na sliki spodaj je delež forumskih tem o raku, določen glede na različne pragove za določitev forumske teme kot teme o raku. Če bi kot prag za določitev forumske teme kot teme o raku uporabili vrednost 5 % (to je, vsaj 5 % sporočil znotraj dane forumske teme mora vsebovati temo raka, da lahko zaključimo, da forumska tema vsebuje temo raka), potem bi nekaj več kot 4 % forumskih tem vsebovalo temo raka. Zanimiv je preskok v deležu forumskih tem s temo raka, ko je prag določen pri 50 %, kar je verjetno posledica dejstva, da veliko forumskih tem vsebuje dve sporočili. V takem primeru je lahko eno sporočilo s temo raka, druga pa ne.

Slika 4 Skupen delež forumskih tem o raku, glede na različne pragove določitve forumske teme o raku



7 Pojavnost sporočil in forumskih tem s temo raka

Rezultati analiz, ki sledijo, so omejeni na forumske teme z med (vključno) 2 in (vključno) 50 sporočili. Na tak način je upoštevanih 68 % vseh sporočil oziroma 79 % vseh forumskih tem. Z modelom slučajnih gozdov smo zaznali 17616 (0,87 %) sporočil o raku in 2018336 (99,13 %) sporočil, ki niso o raku.

Tabela 11 Število sporočil s temo raka po forumih

Forum	*# ni rak*	*# je rak*	*N*	*% ni rak*	*% je rak*	*Vrsta foruma*
Preventivni pregled dojk	59	142	201	29,35	70,65	ZP
Kako živeti z rakom?	831	797	1628	51,04	48,96	PS
Limfom in levkemija	261	180	441	59,18	40,82	PS
Pravna pomoč bolnikom z rakom	124	63	187	66,31	33,69	ZP
Zdravljenje s konopljo	717	181	898	79,84	20,16	ZP
Urologija	173	33	206	83,98	16,02	ZP
Radiologija	165	25	190	86,84	13,16	ZP
Bolezni prebavil (gastroenterologija)	680	81	761	89,36	10,64	ZP
Otorinolaringologija	3205	304	3509	91,34	8,66	ZP
Proktologija	459	34	493	93,1	6,9	ZP
Abdominalna kirurgija	223	15	238	93,7	6,3	ZP
Klimakterij -menopavza	288	19	307	93,81	6,19	ZP
Estetska in plastična kirurgija	1034	68	1102	93,83	6,17	ZP
Multipla skleroza	641	38	679	94,4	5,6	PS
Maksilofacialna in oralna kirurgija	372	21	393	94,66	5,34	ZP
Hormonske motnje	2034	109	2143	94,91	5,09	ZP
Družinska medicina	5604	294	5898	95,02	4,98	ZP
ARHIV – Zdravstvo	753	39	792	95,08	4,92	Drugo
Astma in druge pljučne bolezni	1038	53	1091	95,14	4,86	ZP
Urgenca	8933	366	9299	96,06	3,94	ZP
Kronična bolečina - fibromialgija	575	21	596	96,48	3,52	ZP
ABC ginekologije in porodništva	29158	1030	30188	96,59	3,41	ZP
Nosečnost in obporodno obdobje	3993	138	4131	96,66	3,34	PS
Dermatologija	5087	166	5253	96,84	3,16	ZP
Nevrokirurgija	897	28	925	96,97	3,03	ZP
Žilna kirurgija	132	4	136	97,06	2,94	ZP
Intimna nega in zdravljenje vaginalnih okužb	1071	32	1103	97,1	2,9	ZP
Zdravilstvo	283	8	291	97,25	2,75	ZP
Žalovanje in slovo	217	6	223	97,31	2,69	PS
Domače živali in veterina	4305	114	4419	97,42	2,58	DF
Genetika v nosečnosti in NIPT	456	11	467	97,64	2,36	ZP
Prehrana in varnost živil	167	4	171	97,66	2,34	DF
ARHIV - Alternativa in zdravo življenje	399	9	408	97,79	2,21	Drugo
ARHIV - Zdravila in zakonodaja	253	5	258	98,06	1,94	Drugo
ARTHRON - fizioterapija	270	5	275	98,18	1,82	ZP
Endodontija	169	3	172	98,26	1,74	ZP
Pediatrija	6030	102	6132	98,34	1,66	ZP
Zdravstvo pri nas?	497	8	505	98,42	1,58	DF

Kajenje	188	3	191	98,43	1,57	ZP
Prostovoljna zdravstvena zavarovanja	126	2	128	98,44	1,56	DF
Depresija, anksioznost in razpoloženjske motnje	1340	21	1361	98,46	1,54	ZP
Neonatologija	963	14	977	98,57	1,43	ZP
Cepljenja	2464	35	2499	98,6	1,4	ZP
Elektromagnetna terapija	73	1	74	98,65	1,35	ZP
Homeopatija	700	9	709	98,73	1,27	ZP
Psihiater odgovarja	3143	40	3183	98,74	1,26	ZP
Ortopedija	3294	41	3335	98,77	1,23	ZP
Farmacevt svetuje	1163	14	1177	98,81	1,19	ZP
Alergija	786	9	795	98,87	1,13	ZP
Življenje, ah to življenje	2597	28	2625	98,93	1,07	DF
Manualna medicina - zdravljenje gibalnega sistema	2222	24	2246	98,93	1,07	ZP
Kontracepcija	3027	30	3057	99,02	0,98	PS
Ustno zdravje in higiena	304	3	307	99,02	0,98	ZP
Parodontologija	216	2	218	99,08	0,92	ZP
Težave in bolečine v hrbtenici	1216	11	1227	99,1	0,9	ZP
Seksologija - želiva si spolnosti	247	2	249	99,2	0,8	ZP
Splošna stomatologija	1013	8	1021	99,22	0,78	ZP
ARTHRON - Zdravstveno svetovanje v športu	266	2	268	99,25	0,75	ZP
Vrtičkarstvo, zelišča in urejanje okolice	539	4	543	99,26	0,74	DF
Psihoterapija	801	6	807	99,26	0,74	ZP
Oftalmologija	705	5	710	99,3	0,7	ZP
Partnerska in družinska posvetovalnica	2111	14	2125	99,34	0,66	ZP
Srčno popuščanje	609	4	613	99,35	0,65	ZP
Slušni aparati in zaščita sluha	152	1	153	99,35	0,65	ZP
Otroško in mladinsko zobozdravstvo	312	2	314	99,36	0,64	ZP
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	7207	45	7252	99,38	0,62	ZP
Fizioterapija in masaže	984	6	990	99,39	0,61	ZP
Kardiologija	5073	27	5100	99,47	0,53	ZP
Zobna protetika	386	2	388	99,48	0,52	ZP
Hujšanje	5054	26	5080	99,49	0,51	ZP
Opravljivka	12169	60	12229	99,51	0,49	DF
Starševski čvek	2508663	12174	2520837	99,52	0,48	DF
Zasvojenost in pomoč	1046	5	1051	99,52	0,48	ZP
Zavarovanja	449	2	451	99,56	0,44	DF
Revma	248	1	249	99,6	0,4	PS
Darovanje spolnih celic	248	1	249	99,6	0,4	ZP
Motnje hranjenja	280	1	281	99,64	0,36	ZP
Izgorelost	1651	6	1657	99,64	0,36	ZP
ARHIV – Sociala	1167	4	1171	99,66	0,34	Drugo
Mejna in narcistična osebnostna motnja - za svojce	3777	13	3790	99,66	0,34	PS
Optika	611	2	613	99,67	0,33	ZP
Prehrana in kulinarika	7285	21	7306	99,71	0,29	DF
Ortodontija - čeljustna in zobna ortopedija	362	1	363	99,72	0,28	ZP
Sex v mestu	23005	63	23068	99,73	0,27	DF

Nega, ličenje in style	6922	19	6941	99,73	0,27	DF
Neplodnost	9993	27	10020	99,73	0,27	PS
Knjižni molji in pravopis	2672	7	2679	99,74	0,26	DF
Čustvena inteligenca, komunikacija in odnosi	10032	26	10058	99,74	0,26	PS
Duševno zdravje	403	1	404	99,75	0,25	ZP
Pravni nasveti	2110	5	2115	99,76	0,24	ZP
Mladi, ljubezen in spolnost	7335	15	7350	99,8	0,2	DF
Vrnitev k sebi	502	1	503	99,8	0,2	PS
Podpora pri uporabi forumov Over,Net	590	1	591	99,83	0,17	DF
Ženski in moški čvek	68093	106	68199	99,84	0,16	DF
Starševstvo in vzgoja	13048	21	13069	99,84	0,16	PS
Pediatrična in fetalna kardiologija	617	1	618	99,84	0,16	ZP
Ločitev, nezvestoba in ljubezenski nasveti	15132	22	15154	99,85	0,15	PS
Delovno pravo in zaposlitev	3554	5	3559	99,86	0,14	PS
Pokojninsko zavarovanje	751	1	752	99,87	0,13	DF
Politični čvek	63273	65	63338	99,9	0,1	DF
Mali oglasi	7719	8	7727	99,9	0,1	DF
Kam in kako na počitnice	28743	18	28761	99,94	0,06	DF
Nepremičnine	1914	1	1915	99,95	0,05	DF
ARHIV - Prosti čas	17	0	17	100	0	Drugo
Skrinjica idej!	869	0	869	100	0	DF
Avtomobilska zavarovanja	189	0	189	100	0	DF
Računovodstvo	1231	0	1231	100	0	DF
Finance in kripto valute	4374	0	4374	100	0	DF
Odškodnine in zavarovalnine	424	0	424	100	0	DF
Pomoč voznikom v prometu	1187	0	1187	100	0	DF
Računalništvo, telefonija, internet in televizija	2291	0	2291	100	0	DF
Iščete delo, delavca?	4055	0	4055	100	0	DF
Poroka	282	0	282	100	0	DF
Glasba, film in fotografija	1453	0	1453	100	0	DF
Starejši so modrejši	634	0	634	100	0	DF
Gradimo, opremljamo in urejamo dom	5200	0	5200	100	0	DF
Avtomobilizem	2295	0	2295	100	0	DF
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	408	0	408	100	0	PS
Posvojitev otroka in rejništvo	138	0	138	100	0	PS
Varna pot	9	0	9	100	0	PS
Za ženske v stiski	323	0	323	100	0	PS
Za žalujoče po izgubi zaradi samomora	42	0	42	100	0	PS
Zdravljenje obolenj in deformacij stopala	275	0	275	100	0	ZP
Rekreacija, oblikovanje telesa in zdrava prehrana	907	0	907	100	0	ZP
Laserska korekcija vida	133	0	133	100	0	ZP
Implantologija in zobna protetika	37	0	37	100	0	ZP
Estetsko zobozdravstvo	178	0	178	100	0	ZP
Za ljudi z mejno osebnostno motnjo	55	0	55	100	0	ZP

Tabela 12 Število sporočil s temo raka, glede na vrsto foruma

Vrsta foruma	*# ni o raku*	*# je o raku*	*N*	*% ni o raku*	*% je o raku*
Zdravstvena posvetovalnica	121141	3618	124759	97,1	2,9
Podporna skupina	66176	1305	67481	98,07	1,93
Družabni forum	2772298	12721	2785019	99,54	0,46
Drugo	2589	57	2646	97,85	2,15

Tabela 13 Število forumskih tem, glede na vrsto foruma (forumska tema je opredeljena kot forumska tema o raku, če vsaj polovica sporočil vsebuje temo raka)

Forum	*% forumskih tem o raku*	*# forumskih tem o raku*	*Skupno število forumskih tem*	*Vrsta foruma*
Preventivni pregled dojk	82,42	75	91	ZP
Kako živeti z rakom?	55,66	177	318	PS
Pravna pomoč bolnikom z rakom	47,44	37	78	ZP
Limfom in levkemija	44,97	67	149	PS
Urologija	14,37	23	160	ZP
Radiologija	14,19	21	148	ZP
Zdravljenje s konopljo	13,04	3	23	ZP
Bolezni prebavil (gastroenterologija)	11,84	47	397	ZP
Proktologija	10,48	13	124	ZP
Otorinolaringologija	9,55	107	1120	ZP
Klimakterij -menopavza	8,52	15	176	ZP
Estetska in plastična kirurgija	8,25	41	497	ZP
Hormonske motnje	8,17	68	832	ZP
Abdominalna kirurgija	7,95	14	176	ZP
Astma in druge pljučne bolezni	6,85	27	394	ZP
Žalovanje in slovo	6,45	4	62	PS
Družinska medicina	6,14	171	2785	ZP
Maksilofacialna in oralna kirurgija	5,83	12	206	ZP
Kronična bolečina - fibromialgija	5,35	13	243	ZP
Multipla skleroza	4,91	11	224	PS
Urgenca	4,71	139	2951	ZP
Žilna kirurgija	4,55	4	88	ZP
Nosečnost in obporodno obdobje	4,46	63	1413	PS
Dermatologija	4,46	79	1771	ZP
ABC ginekologije in porodništva	4,37	348	7969	ZP
ARHIV - Zdravstvo	4,21	17	404	Drugo
Nevrokirurgija	4,07	12	295	ZP
ARHIV - Zdravila in zakonodaja	3,88	4	103	Drugo
Zdravilstvo	3,49	6	172	ZP
Kajenje	3,33	2	60	ZP
Neonatologija	3,33	9	270	ZP
Genetika v nosečnosti in NIPT	3,29	8	243	ZP
ARHIV - Alternativa in zdravo življenje	3,17	6	189	Drugo
Pediatrična in fetalna kardiologija	2,94	1	34	ZP

Intimna nega in zdravljenje vaginalnih okužb	2,88	14	486	ZP
ARTHRON - fizioterapija	2,72	4	147	ZP
Prehrana in varnost živil	2,63	2	76	DF
Zdravstvo pri nas?	2,5	6	240	DF
Ustno zdravje in higiena	2,27	2	88	ZP
Pediatrija	2,21	49	2217	ZP
Psihiater odgovarja	2,11	23	1090	ZP
Domače živali in veterina	1,97	21	1067	DF
Seksologija - želiva si spolnosti	1,8	2	111	ZP
Cepljenja	1,8	10	555	ZP
Farmacevt svetuje	1,72	10	580	ZP
Depresija, anksioznost in razpoloženjske motnje	1,57	6	383	ZP
Splošna stomatologija	1,5	6	401	ZP
Prostovoljna zdravstvena zavarovanja	1,49	1	67	DF
Endodontija	1,45	1	69	ZP
Homeopatija	1,45	4	276	ZP
Ortopedija	1,38	14	1013	ZP
Psihoterapija	1,29	4	310	ZP
ARHIV - Sociala	1,26	3	238	Drugo
Kontracepcija	1,23	10	815	PS
Parodontologija	1,19	1	84	ZP
ARTHRON - Zdravstveno svetovanje v športu	1,16	1	86	ZP
Oftalmologija	1,15	3	260	ZP
Motnje hranjenja	1,1	1	91	ZP
Vrtičkarstvo, zelišča in urejanje okolice	1,04	2	192	DF
Alergija	1,04	4	386	ZP
Manualna medicina - zdravljenje gibalnega sistema	1,02	9	880	ZP
Revma	0,98	1	102	PS
Kardiologija	0,98	14	1422	ZP
Izgorelost	0,98	1	102	ZP
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	0,97	17	1756	ZP
Srčno popuščanje	0,93	1	107	ZP
Hujšanje	0,82	4	490	ZP
Pokojninsko zavarovanje	0,78	1	128	DF
Partnerska in družinska posvetovalnica	0,78	5	638	ZP
Težave in bolečine v hrbtenici	0,77	3	391	ZP
Čustvena inteligenca, komunikacija in odnosi	0,73	7	956	PS
Ortodontija - čeljustna in zobna ortopedija	0,68	1	146	ZP
Zasvojenost in pomoč	0,67	2	297	ZP
Zavarovanja	0,62	1	161	DF
Fizioterapija in masaže	0,58	2	347	ZP
Podpora pri uporabi forumov Over,Net	0,57	1	176	DF
Optika	0,57	2	350	ZP
Knjižni molji in pravopis	0,53	4	758	DF

Življenje, ah to življenje	0,42	3	715	DF
Pravni nasveti	0,3	5	1677	ZP
Opravljivka	0,27	1	375	DF
Mladi, ljubezen in spolnost	0,25	5	1976	DF
Delovno pravo in zaposlitev	0,23	3	1308	PS
Starševski čvek	0,22	424	193990	DF
Neplodnost	0,16	1	633	PS
Nega, ličenje in style	0,12	1	819	DF
Mali oglasi	0,1	4	3811	DF
Ženski in moški čvek	0,08	2	2383	DF
Sex v mestu	0,06	1	1764	DF
Ločitev, nezvestoba in ljubezenski nasveti	0,06	1	1586	PS
Starševstvo in vzgoja	0,05	1	2078	PS
Politični čvek	0,03	1	3534	DF
ARHIV - Prosti čas	0	0	4	Drugo
Skrinjica idej!	0	0	356	DF
Avtomobilska zavarovanja	0	0	67	DF
Računovodstvo	0	0	369	DF
Finance in kripto valute	0	0	508	DF
Nepremičnine	0	0	393	DF
Prehrana in kulinarika	0	0	687	DF
Kam in kako na počitnice	0	0	3695	DF
Odškodnine in zavarovalnine	0	0	172	DF
Pomoč voznikom v prometu	0	0	331	DF
Računalništvo, telefonija, internet in televizija	0	0	726	DF
Iščete delo, delavca?	0	0	2222	DF
Poroka	0	0	118	DF
Glasba, film in fotografija	0	0	278	DF
Starejši so modrejši	0	0	204	DF
Gradimo, opremljamo in urejamo dom	0	0	1141	DF
Avtomobilizem	0	0	618	DF
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	0	0	185	PS
Posvojitev otroka in rejništvo	0	0	38	PS
Varna pot	0	0	8	PS
Za ženske v stiski	0	0	120	PS
Mejna in narcistična osebnostna motnja - za svojce	0	0	365	PS
Vrnitev k sebi	0	0	145	PS
Za žalujoče po izgubi zaradi samomora	0	0	15	PS
Darovanje spolnih celic	0	0	38	ZP
Zdravljenje obolenj in deformacij stopala	0	0	141	ZP
Zobna protetika	0	0	145	ZP
Otroško in mladinsko zobozdravstvo	0	0	99	ZP
Rekreacija, oblikovanje telesa in zdrava prehrana	0	0	392	ZP
Elektromagnetna terapija	0	0	43	ZP

Duševno zdravje	0	0	158	ZP
Laserska korekcija vida	0	0	49	ZP
Implantologija in zobna protetika	0	0	20	ZP
Slušni aparati in zaščita sluha	0	0	38	ZP
Estetsko zobozdravstvo	0	0	15	ZP
Za ljudi z mejno osebnostno motnjo	0	0	16	ZP

Tabela 14 Število forumskih tem, glede na vrsto foruma (forumska tema je opredeljena kot forumska tema o raku, če vsaj eno sporočilo vsebuje temo raka)

Forum	*# forumskih z rakom*	*% forumskih tem z rakom*	*Skupno # tem*	*Vrsta foruma*
Preventivni pregled dojk	77	85	91	ZP
Kako živeti z rakom?	238	75	318	PS
Pravna pomoč bolnikom z rakom	53	68	78	ZP
Limfom in levkemija	94	63	149	PS
Zdravljenje s konopljo	11	48	23	ZP
Radiologija	23	16	148	ZP
Otorinolaringologija	163	15	1120	ZP
Bolezni prebavil (gastroenterologija)	58	15	397	ZP
Proktologija	19	15	124	ZP
Urologija	24	15	160	ZP
Hormonske motnje	90	11	832	ZP
Klimakterij -menopavza	18	10	176	ZP
Estetska in plastična kirurgija	50	10	497	ZP
Urgenca	254	9	2951	ZP
Abdominalna kirurgija	15	9	176	ZP
Astma in druge pljučne bolezni	36	9	394	ZP
Kronična bolečina - fibromialgija	19	8	243	ZP
Žalovanje in slovo	5	8	62	PS
Multipla skleroza	19	8	224	PS
ABC ginekologije in porodništva	604	8	7969	ZP
Nevrokirurgija	23	8	295	ZP
Družinska medicina	224	8	2785	ZP
Nosečnost in obporodno obdobje	105	7	1413	PS
Dermatologija	120	7	1771	ZP
Maksilofacialna in oralna kirurgija	15	7	206	ZP
ARHIV - Zdravstvo	28	7	404	Drugo
Kajenje	3	5	60	ZP
Prehrana in varnost živil	4	5	76	DF
Neonatologija	13	5	270	ZP
Žilna kirurgija	4	5	88	ZP
Intimna nega in zdravljenje vaginalnih okužb	23	5	486	ZP
Cepljenja	26	5	555	ZP
ARHIV - Zdravila in zakonodaja	4	4	103	Drugo
Zdravilstvo	7	4	172	ZP
Domače živali in veterina	47	4	1067	DF

Pediatrija	85	4	2217	ZP
Depresija, anksioznost in razpoloženjske motnje	16	4	383	ZP
Genetika v nosečnosti in NIPT	10	4	243	ZP
ARHIV - Alternativa in zdravo življenje	7	4	189	Drugo
Izgorelost	4	4	102	ZP
Zdravstvo pri nas?	7	3	240	DF
Darovanje spolnih celic	1	3	38	ZP
Pediatrična in fetalna kardiologija	1	3	34	ZP
Kontracepcija	25	3	815	PS
Endodontija	2	3	69	ZP
Homeopatija	8	3	276	ZP
Mejna in narcistična osebnostna motnja - za svojce	10	3	365	PS
Srčno popuščanje	3	3	107	ZP
ARTHRON - fizioterapija	4	3	147	ZP
Slušni aparati in zaščita sluha	1	3	38	ZP
Opravljivka	11	3	375	DF
Psihiater odgovarja	35	3	1090	ZP
Neplodnost	21	3	633	PS
Hujšanje	14	3	490	ZP
Kardiologija	27	2	1422	ZP
Manualna medicina - zdravljenje gibalnega sistema	22	2	880	ZP
Ženski in moški čvek	50	2	2383	DF
Alergija	7	2	386	ZP
Partnerska in družinska posvetovalnica	14	2	638	ZP
Sex v mestu	33	2	1764	DF
Starševski čvek	4110	2	193990	DF
Čustvena inteligenca, komunikacija in odnosi	16	2	956	PS
Ortopedija	23	2	1013	ZP
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	36	2	1756	ZP
Seksologija - želiva si spolnosti	2	2	111	ZP
Farmacevt svetuje	13	2	580	ZP
Parodontologija	2	2	84	ZP
Elektromagnetna terapija	1	2	43	ZP
Težave in bolečine v hrbtenici	8	2	391	ZP
Psihoterapija	6	2	310	ZP
Ustno zdravje in higiena	2	2	88	ZP
ARHIV - Sociala	4	2	238	Drugo
Oftalmologija	5	2	260	ZP
Vrtičkarstvo, zelišča in urejanje okolice	3	2	192	DF
Splošna stomatologija	8	2	401	ZP
Zavarovanja	2	1	161	DF
Starševstvo in vzgoja	15	1	2078	PS
Zasvojenost in pomoč	4	1	297	ZP
Mladi, ljubezen in spolnost	13	1	1976	DF
Politični čvek	38	1	3534	DF
Revma	1	1	102	PS

Prehrana in kulinarika	7	1	687	DF
Prostovoljna zdravstvena zavarovanja	1	1	67	DF
Ortodontija - čeljustna in zobna ortopedija	1	1	146	ZP
Zobna protetika	2	1	145	ZP
Življenje, ah to življenje	5	1	715	DF
Otroško in mladinsko zobozdravstvo	1	1	99	ZP
Knjižni molji in pravopis	6	1	758	DF
Pokojninsko zavarovanje	1	1	128	DF
Fizioterapija in masaže	5	1	347	ZP
Motnje hranjenja	1	1	91	ZP
Duševno zdravje	1	1	158	ZP
Vrnitev k sebi	1	1	145	PS
ARTHRON - Zdravstveno svetovanje v športu	1	1	86	ZP
Podpora pri uporabi forumov Over.Net	1	1	176	DF
Nega, ličenje in style	11	1	819	DF
Optika	2	1	350	ZP
Ločitev, nezvestoba in ljubezenski nasveti	16	1	1586	PS
Skrinjica idej!	0	0	356	DF
Avtomobilska zavarovanja	0	0	67	DF
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	0	0	185	PS
Posvojitev otroka in rejništvo	0	0	38	PS
Delovno pravo in zaposlitev	5	0	1308	PS
Računovodstvo	0	0	369	DF
Finance in kripto valute	0	0	508	DF
Nepremičnine	1	0	393	DF
Kam in kako na počitnice	14	0	3695	DF
Varna pot	0	0	8	PS
Zdravljenje obolenj in deformacij stopala	0	0	141	ZP
Odškodnine in zavarovalnine	0	0	172	DF
Pravni nasveti	5	0	1677	ZP
Rekreacija, oblikovanje telesa in zdrava prehrana	0	0	392	ZP
Pomoč voznikom v prometu	0	0	331	DF
Računalništvo, telefonija, internet in televizija	0	0	726	DF
Za ženske v stiski	0	0	120	PS
ARHIV - Prosti čas	0	0	4	Drugo
Laserska korekcija vida	0	0	49	ZP
Za žalujoče po izgubi zaradi samomora	0	0	15	PS
Implantologija in zobna protetika	0	0	20	ZP
Estetsko zobozdravstvo	0	0	15	ZP
Iščete delo, delavca?	0	0	2222	DF
Za ljudi z mejno osebnostno motnjo	0	0	16	ZP
Mali oglasi	6	0	3811	DF
Poroka	0	0	118	DF
Glasba, film in fotografija	0	0	278	DF
Starejši so modrejši	0	0	204	DF

Gradimo, opremljamo in urejamo dom	0	0	1141	DF
Avtomobilizem	0	0	618	DF

7.1 Pojavnost sporočil in forumskih tem s temo raka po letih

Večina (96,5 %) forumskih tem se začne in zaključi znotraj enega koledarskega leta. 2 % forumskih tem se razteza med dve koledarski leti. Upoštevati je potrebno tudi dejstvo, da če se je forumska tema začela leta 2019, se ni mogla nadaljevati v leto 2020, ker ga še ni bilo. Prav tako niso bila upoštevana sporočila iz let pred 2015.

Tabela 15 Število koledarskih let, prek katerih se raztezajo forumske teme (ne glede na to, ali vsebujejo temo raka ali ne)

Število koledarskih let	*Frekvenca*	*Kumulativna frekvenca*	*Delež*	*Kumulativni delež*
1	265643	265643	96,5	96,5
2	5696	271339	2,07	98,57
3	2277	273616	0,83	99,4
4	1121	274737	0,41	99,81
4	531	275268	0,19	100

Tabela 16 Število forumskih tem glede na leto prvega sporočila, glede na vrsto foruma: upoštevane so forumske teme, ki vsebujejo temo raka (forumska tema vsebuje temo raka, če vsaj eno sporočilo vsebuje temo raka)

Vrsta foruma	*2015*	*2016*	*2017*	*2018*	*2019*
Drugo	20	12	9	2	0
Družabni forum	919	961	868	776	847
Podporna skupina	165	122	106	80	98
Zdravstvena posvetovalnica	518	469	542	436	390
Skupaj	*1622*	*1564*	*1525*	*1294*	*1335*

Tabela 17 Število forumskih tem s temo raka (forumska tema vsebuje temo raka, če vsaj eno sporočilo vsebuje temo raka) po obdobjih

Forum	*# raka (2015-2017)*	*% raka (2015-2017)*	*# raka (2018-2019)*	*% raka (2018-2019)*
Zdravljenje s konopljo	7	88	9	43
Preventivni pregled dojk	43	84	35	85
Kako živeti z rakom?	165	80	84	68
Pravna pomoč bolnikom z rakom	33	77	22	59
Limfom in levkemija	56	60	40	69
Radiologija	18	19	6	11
Proktologija	11	15	10	17
Urologija	14	14	10	16
Otorinolaringologija	93	13	79	17
Hormonske motnje	64	11	28	11
Bolezni prebavil (gastroenterologija)	17	11	43	17

Klimakterij -menopavza	15	11	3	7
Estetska in plastična kirurgija	37	10	14	10
Žalovanje in slovo	5	10	0	0
Abdominalna kirurgija	11	9	4	7
Kajenje	3	9	0	0
ABC ginekologije in porodništva	404	8	229	8
Urgenca	189	8	74	12
Družinska medicina	167	8	62	8
Nosečnost in obporodno obdobje	79	8	29	7
Kronična bolečina - fibromialgija	13	8	7	8
Dermatologija	83	7	40	6
ARHIV - Zdravstvo	26	7	3	6
Nevrokirurgija	10	7	14	9
Multipla skleroza	8	7	13	12
Astma in druge pljučne bolezni	17	6	19	14
Opravljivka	1	6	11	3
Maksilofacialna in oralna kirurgija	7	5	8	11
Prehrana in varnost živil	3	5	1	3
Srčno popuščanje	2	5	2	3
Darovanje spolnih celic	1	5	0	0
Pediatrija	63	4	25	4
Domače živali in veterina	35	4	14	5
Cepljenja	14	4	15	6
Intimna nega in zdravljenje vaginalnih okužb	12	4	11	6
Depresija, anksioznost in razpoloženjske motnje	9	4	8	6
ARHIV - Alternativa in zdravo življenje	7	4	0	0
Genetika v nosečnosti in NIPT	6	4	4	5
ARHIV - Zdravila in zakonodaja	4	4	0	0
ARTHRON - fizioterapija	4	4	0	0
Žilna kirurgija	2	4	3	7
Parodontologija	2	4	0	0
Endodontija	2	4	0	0
Ženski in moški čvek	44	3	10	1
Psihiater odgovarja	22	3	14	4
Manualna medicina - zdravljenje gibalnega sistema	15	3	8	2
Kontracepcija	15	3	12	4
Ortopedija	14	3	12	2
Neplodnost	14	3	11	6
Zdravstvo pri nas?	6	3	1	2
Neonatologija	6	3	7	7
Alergija	4	3	4	2
Elektromagnetna terapija	1	3	0	0
Starševski čvek	2582	2	1729	2
Sex v mestu	21	2	23	2
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	18	2	22	3
Čustvena inteligenca, komunikacija in odnosi	16	2	3	1

Hujšanje	9	2	7	6
Farmacevt svetuje	7	2	6	3
Homeopatija	5	2	3	7
Mejna in narcistična osebnostna motnja - za svojce	5	2	8	5
Psihoterapija	4	2	2	2
ARHIV - Sociala	4	2	0	0
Zdravilstvo	3	2	4	7
Zavarovanja	2	2	1	2
Zobna protetika	2	2	0	0
Motnje hranjenja	1	2	0	0
ARTHRON - Zdravstveno svetovanje v športu	1	2	0	0
Ustno zdravje in higiena	1	2	1	3
Politični čvek	18	1	20	1
Starševstvo in vzgoja	13	1	2	0
Ločitev, nezvestoba in ljubezenski nasveti	13	1	5	1
Mladi, ljubezen in spolnost	11	1	2	0
Nega, ličenje in style	4	1	9	2
Zasvojenost in pomoč	3	1	1	1
Prehrana in kulinarika	3	1	4	1
Knjižni molji in pravopis	3	1	4	2
Težave in bolečine v hrbtenici	3	1	5	3
Partnerska in družinska posvetovalnica	2	1	12	4
Fizioterapija in masaže	2	1	3	2
Vrtičkarstvo, zelišča in urejanje okolice	2	1	1	3
Optika	2	1	0	0
Splošna stomatologija	2	1	6	5
Ortodontija - čeljustna in zobna ortopedija	1	1	0	0
Seksologija - želiva si spolnosti	1	1	1	3
Otroško in mladinsko zobozdravstvo	1	1	0	0
Pokojninsko zavarovanje	1	1	0	0
Vrnitev k sebi	1	1	0	0
Kam in kako na počitnice	8	0	6	0
Pravni nasveti	5	0	0	0
Delovno pravo in zaposlitev	3	0	2	1
Mali oglasi	3	0	3	0
Nepremičnine	1	0	0	0
Skrinjica idej!	0	0	0	0
Avtomobilska zavarovanja	0	0	0	0
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	0	0	0	0
Posvojitev otroka in rejništvo	0	0	0	0
Računovodstvo	0	0	0	0
Finance in kripto valute	0	0	0	0
Revma	0	0	1	4
Prostovoljna zdravstvena zavarovanja	0	0	1	7
Varna pot	0	0		
Zdravljenje obolenj in deformacij stopala	0	0	0	0

Pediatrična in fetalna kardiologija	0	0	1	5
Življenje, ah to življenje	0	0	5	2
Odškodnine in zavarovalnine	0	0	0	0
Rekreacija, oblikovanje telesa in zdrava prehrana	0	0	0	0
Pomoč voznikom v prometu	0	0	0	0
Računalništvo, telefonija, internet in televizija	0	0	0	0
Za ženske v stiski	0	0	0	0
Duševno zdravje	0	0	1	1
Podpora pri uporabi forumov Over.Net	0	0	1	1
ARHIV - Prosti čas	0	0		
Oftalmologija	0	0	5	3
Laserska korekcija vida	0	0	0	0
Za žalujoče po izgubi zaradi samomora	0	0	0	0
Implantologija in zobna protetika	0	0	0	0
Izgorelost	0	0	4	5
Slušni aparati in zaščita sluha	0	0	1	6
Estetsko zobozdravstvo	0	0	0	0
Iščete delo, delavca?	0	0	0	0
Poroka	0	0	0	0
Glasba, film in fotografija	0	0	0	0
Starejši so modrejši	0	0	0	0
Gradimo, opremljamo in urejamo dom	0	0	0	0
Avtomobilizem	0	0	0	0
Za ljudi z mejno osebnostno motnjo	NA	NA	0	0

Tabela 18 Število forumskih tem s temo raka (forumska tema vsebuje temo raka, če vsaj eno sporočilo vsebuje temo raka) po letih

Forum	*2015 #*	*2015 %*	*2016 #*	*2016 %*	*2017 #*	*2017 %*	*2018 #*	*2018 %*	*2019 #*	*2019 %*
Zdravljenje s konopljo	1	100	4	80	6	86	6	60	7	47
Preventivni pregled dojk	14	88	16	80	15	88	23	96	16	76
Kako živeti z rakom?	63	73	77	87	52	84	39	91	51	59
Pravna pomoč bolnikom z rakom	13	72	12	80	14	88	11	69	13	57
Limfom in levkemija	19	59	25	54	15	79	19	66	23	74
Proktologija	5	33	5	13	4	16	6	19	4	13
Elektromagnetna terapija	1	25	0	0	0	0	0	0	0	0
Abdominalna kirurgija	4	14	6	11	3	6	4	11	0	0
Otorinolaringologija	26	12	32	13	52	17	47	19	40	16
Estetska in plastična kirurgija	15	12	7	4	18	23	4	7	11	12
Nosečnost in obporodno obdobje	43	11	24	7	19	7	16	6	16	8
Bolezni prebavil (gastroenterologija)	13	11	7	17	2	50	13	17	32	18
Klimakterij -menopavza	4	11	6	12	7	13	2	6	1	8
Kronična bolečina - fibromialgija	6	10	1	3	8	10	4	7	4	9
Žalovanje in slovo	2	10	2	9	2	17	0	0	0	0
Hormonske motnje	18	9	23	11	24	13	14	10	14	10
Prehrana in varnost živil	2	9	2	8	1	4	0	0	1	6

Astma in druge pljučne bolezni	7	9	5	4	5	6	13	16	7	12
Parodontologija	2	9	0	0	0	0	0	0	0	0
ARHIV - Zdravstvo	14	9	9	7	7	8	3	6	0	0
ABC ginekologije in porodništva	153	8	138	7	149	8	149	9	103	8
Kajenje	1	8	2	20	0	0	0	0	0	0
Maksilofacialna in oralna kirurgija	5	8	2	4	2	6	4	10	4	13
Družinska medicina	49	8	48	6	75	10	27	7	37	9
Radiologija	2	8	6	17	12	24	2	7	5	16
Dermatologija	32	7	16	5	42	10	18	6	24	7
Urgenca	55	7	67	8	78	11	44	13	39	12
Genetika v nosečnosti in NIPT	3	7	2	4	1	1	1	2	3	10
Multipla skleroza	3	6	1	2	5	14	6	9	7	14
Neonatologija	5	6	2	3	1	2	5	9	2	4
Cepljenja	4	6	6	5	6	4	8	6	10	9
Kontracepcija	9	5	7	3	4	2	7	4	6	5
Intimna nega in zdravljenje vaginalnih okužb	5	5	6	5	3	3	6	7	8	7
Manualna medicina - zdravljenje gibalnega sistema	5	4	6	3	7	4	6	3	3	1
Ženski in moški čvek	32	4	14	2	6	2	7	2	6	2
Žilna kirurgija	1	4	0	0	1	6	3	11	0	0
Endodontija	1	4	1	7	0	0	0	0	0	0
Domače živali in veterina	14	4	12	4	16	7	9	5	7	6
Depresija, anksioznost in razpoloženjske motnje	4	4	4	4	2	3	6	8	4	5
Psihiater odgovarja	7	4	8	2	9	3	8	5	7	3
Neplodnost	7	4	5	2	4	4	8	7	6	7
Zdravstvo pri nas?	3	3	3	4	0	0	1	3	0	0
Zdravilstvo	2	3	0	0	1	3	3	9	2	9
Zobna protetika	1	3	1	3	0	0	0	0	0	0
Otroško in mladinsko zobozdravstvo	1	3	0	0	0	0	0	0	0	0
Homeopatija	3	3	1	1	1	2	2	7	1	7
Motnje hranjenja	1	3	0	0	0	0	0	0	0	0
Pediatrija	23	3	24	5	24	5	16	4	12	4
Kardiologija	7	2	9	3	9	3	6	2	2	1
Sex v mestu	4	2	12	4	17	4	20	3	14	3
Starševski čvek	840	2	982	2	971	2	936	3	949	3
Čustvena inteligenca, komunikacija in odnosi	5	2	3	2	10	3	3	1	0	0
Ortopedija	1	2	6	4	8	3	9	3	4	1
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	6	2	13	3	5	1	14	4	11	3
Ortodontija - čeljustna in zobna ortopedija	1	2	0	0	0	0	0	0	0	0
Mejna in narcistična osebnostna motnja - za svojce	2	2	2	2	3	4	7	8	4	6
Starševstvo in vzgoja	9	1	1	0	3	1	2	1	0	0
Zasvojenost in pomoč	1	1	1	1	1	1	1	2	0	0
Mladi, ljubezen in spolnost	5	1	3	0	5	1	1	0	1	0
Nepremičnine	1	1	0	0	0	0	0	0	0	0

Politični čvek	10	1	3	1	6	1	10	1	11	2
Prehrana in kulinarika	3	1	0	0	0	0	3	1	2	2
Nega, ličenje in style	2	1	2	1	3	2	2	2	9	3
Optika	1	1	1	1	0	0	0	0	0	0
Splošna stomatologija	1	1	1	1	0	0	2	4	5	8
Hujšanje	1	1	2	1	6	5	6	7	1	2
Zavarovanja	0	0	2	4	1	4	1	4	0	0
Skrinjica idej!	0	0	0	0	0	0	0	0	0	0
Avtomobilska zavarovanja	0	0	0	0	0	0	0	0	0	0
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	0	0	0	0	0	0	0	0	0	0
Darovanje spolnih celic	0	0	1	9	0	0	0	0	0	0
Alergija	0	0	1	3	3	4	3	3	1	1
Partnerska in družinska posvetovalnica	0	0	0	0	2	1	3	2	9	6
Posvojitev otroka in rejništvo	0	0	0	0	0	0	0	0	0	0
Delovno pravo in zaposlitev	1	0	0	0	2	1	1	0	1	1
Računovodstvo	0	0	0	0	0	0	0	0	0	0
Finance in kripto valute	0	0	0	0	0	0	0	0	0	0
Revma	0	0	0	0	0	0	1	5	0	0
Kam in kako na počitnice	1	0	4	1	4	0	3	0	4	1
Prostovoljna zdravstvena zavarovanja	0	0	0	0	0	0	1	10	0	0
Zdravljenje obolenj in deformacij stopala	0	0	0	0	0	0	0	0	0	0
Nevrokirurgija	0	0	3	4	8	10	8	11	7	8
Pediatrična in fetalna kardiologija	0	0	0	0	0	0	0	0	1	5
Življenje, ah to življenje	0	0	0	0	0	0	0	0	5	4
Seksologija - želiva si spolnosti	0	0	0	0	1	3	1	3	0	0
Farmacevt svetuje	0	0	5	4	2	2	4	3	2	2
Odškodnine in zavarovalnine	0	0	0	0	0	0	0	0	0	0
Pravni nasveti	1	0	4	1	0	0	0	0	0	0
Knjižni molji in pravopis	1	0	1	1	2	2	3	3	2	2
Rekreacija, oblikovanje telesa in zdrava prehrana	0	0	0	0	0	0	0	0	0	0
Pomoč voznikom v prometu	0	0	0	0	0	0	0	0	0	0
Računalništvo, telefonija, internet in televizija	0	0	0	0	0	0	0	0	0	0
Za ženske v stiski	0	0	0	0	0	0	0	0	0	0
Pokojninsko zavarovanje	0	0	1	2	0	0	0	0	0	0
Fizioterapija in masaže	0	0	1	1	1	1	1	1	2	4
Težave in bolečine v hrbtenici	0	0	2	2	1	1	2	2	3	3
Duševno zdravje	0	0	0	0	0	0	0	0	1	2
ARTHRON - fizioterapija	0	0	2	5	3	9	0	0	0	0
Vrnitev k sebi	0	0	0	0	1	2	0	0	0	0
Psihoterapija	0	0	2	3	2	2	0	0	2	4
ARTHRON - Zdravstveno svetovanje v športu	0	0	0	0	1	7	0	0	0	0
Ustno zdravje in higiena	0	0	1	6	0	0	0	0	1	5
Oftalmologija	0	0	0	0	0	0	2	2	3	3
Izgorelost	0	0	0	0	0	0	2	3	2	9

Slušni aparati in zaščita sluha	0	0	0	0	0	0	1	7	0	0
Iščete delo, delavca?	0	0	0	0	0	0	0	0	0	0
Opravljivka	0	0	0	0	1	6	1	3	11	3
Vrtičkarstvo, zelišča in urejanje okolice	0	0	1	2	1	4	0	0	1	7
Mali oglasi	1	0	0	0	2	0	1	0	2	1
Poroka	0	0	0	0	0	0	0	0	0	0
Glasba, film in fotografija	0	0	0	0	0	0	0	0	0	0
Starejši so modrejši	0	0	0	0	0	0	0	0	0	0
Ločitev, nezvestoba in ljubezenski nasveti	2	0	3	1	8	2	3	1	2	1
Gradimo, opremljamo in urejamo dom	0	0	0	0	0	0	0	0	0	0
Avtomobilizem	0	0	0	0	0	0	0	0	0	0

Tabela 19 Število forumskih tem s temo raka po forumih (forumska tema vsebuje temo raka, če vsaj polovica sporočil vsebuje temo raka) (upoštevane so samo forumske teme z več kot enim sporočilom)

Forum	*% forumskih tem o raku*	*# forumskih tem o raku*	*Skupno število forumskih tem*
Preventivni pregled dojk	83,54	66	79
Kako živeti z rakom?	55,65	133	239
Pravna pomoč bolnikom z rakom	46,88	30	64
Limfom in levkemija	43,01	40	93
Urologija	27,27	3	11
Zdravljenje s konopljo	14,29	3	21
Radiologija	13,79	4	29
Bolezni prebavil (gastroenterologija)	11,35	26	229
Estetska in plastična kirurgija	10,7	35	327
Otorinolaringologija	9,86	103	1045
Žalovanje in slovo	9,68	3	31
Proktologija	9,4	11	117
Hormonske motnje	8,22	60	730
Klimakterij -menopavza	8,11	6	74
Maksilofacialna in oralna kirurgija	6,9	6	87
Astma in druge pljučne bolezni	6,01	19	316
Družinska medicina	5,73	100	1746
Multipla skleroza	5,5	6	109
Kronična bolečina - fibromialgija	5,08	6	118
ARHIV - Zdravstvo	4,85	8	165
Dermatologija	4,64	78	1681
Nosečnost in obporodno obdobje	4,58	34	742
Urgenca	4,51	117	2594
ABC ginekologije in porodništva	4,48	331	7382
Žilna kirurgija	4,35	1	23
Nevrokirurgija	4,24	12	283
ARHIV - Zdravila in zakonodaja	4,17	4	96
Pediatrična in fetalna kardiologija	3,23	1	31
Neonatologija	3,14	8	255
ARHIV - Alternativa in zdravo življenje	2,67	2	75

Intimna nega in zdravljenje vaginalnih okužb	2,52	8	317
Domače živali in veterina	2,47	18	730
Seksologija - želiva si spolnosti	2,44	2	82
Zdravilstvo	2,41	2	83
Ustno zdravje in higiena	2,41	2	83
Zdravstvo pri nas?	2,25	2	89
Alergija	2,23	4	179
ARTHRON - fizioterapija	2,22	2	90
Pediatrija	2,21	47	2123
Psihiater odgovarja	2,16	20	927
Revma	2,08	1	48
Prostovoljna zdravstvena zavarovanja	2	1	50
Prehrana in varnost živil	2	1	50
Farmacevt svetuje	1,94	8	412
Depresija, anksioznost in razpoloženjske motnje	1,71	6	351
Splošna stomatologija	1,63	6	367
Motnje hranjenja	1,49	1	67
Ortopedija	1,35	13	964
Psihoterapija	1,35	3	223
Cepljenja	1,33	7	525
Parodontologija	1,33	1	75
Kontracepcija	1,29	10	775
Homeopatija	1,29	3	232
ARTHRON - Zdravstveno svetovanje v športu	1,27	1	79
Oftalmologija	1,23	3	244
Srčno popuščanje	1,16	1	86
Manualna medicina - zdravljenje gibalnega sistema	1,09	9	829
Zasvojenost in pomoč	1,09	2	184
Pokojninsko zavarovanje	1,09	1	92
Izgorelost	1,06	1	94
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	0,98	17	1733
Podpora pri uporabi forumov Over,Net	0,91	1	110
Hujšanje	0,85	3	353
Težave in bolečine v hrbtenici	0,83	3	360
Ortodontija - čeljustna in zobna ortopedija	0,77	1	130
Genetika v nosečnosti in NIPT	0,75	1	134
Kardiologija	0,73	8	1096
Zavarovanja	0,73	1	137
Fizioterapija in masaže	0,63	2	319
Knjižni molji in pravopis	0,6	3	499
Življenje, ah to življenje	0,46	2	435
Mladi, ljubezen in spolnost	0,32	4	1254
Opravljivka	0,29	1	350
Neplodnost	0,27	1	372
Čustvena inteligenca, komunikacija in odnosi	0,25	2	805
Partnerska in družinska posvetovalnica	0,2	1	507

Starševski čvek	0,2	332	168750
Nega, ličenje in style	0,18	1	543
Starševstvo in vzgoja	0,08	1	1196
Ločitev, nezvestoba in ljubezenski nasveti	0,08	1	1265
Sex v mestu	0,07	1	1437
Ženski in moški čvek	0,05	1	1823
Politični čvek	0,03	1	3132
Skrinjica idej!	0	0	126
Avtomobilska zavarovanja	0	0	58
Otroci s posebnimi potrebami in dolgotrajno bolni otroci	0	0	109
Darovanje spolnih celic	0	0	30
Posvojitev otroka in rejništvo	0	0	26
Delovno pravo in zaposlitev	0	0	723
Računovodstvo	0	0	266
Finance in kripto valute	0	0	283
Nepremičnine	0	0	239
Kajenje	0	0	45
Prehrana in kulinarika	0	0	532
Kam in kako na počitnice	0	0	2543
Varna pot	0	0	1
Zdravljenje obolenj in deformacij stopala	0	0	82
Abdominalna kirurgija	0	0	26
Zobna protetika	0	0	139
Odškodnine in zavarovalnine	0	0	107
Pravni nasveti	0	0	287
Otroško in mladinsko zobozdravstvo	0	0	65
Rekreacija, oblikovanje telesa in zdrava prehrana	0	0	199
Pomoč voznikom v prometu	0	0	304
Računalništvo, telefonija, internet in televizija	0	0	438
Za ženske v stiski	0	0	85
Elektromagnetna terapija	0	0	24
Endodontija	0	0	46
Mejna in narcistična osebnostna motnja - za svojce	0	0	317
Duševno zdravje	0	0	102
Vrnitev k sebi	0	0	113
ARHIV - Sociala	0	0	93
ARHIV - Prosti čas	0	0	2
Laserska korekcija vida	0	0	47
Za žalujoče po izgubi zaradi samomora	0	0	13
Implantologija in zobna protetika	0	0	13
Slušni aparati in zaščita sluha	0	0	31
Estetsko zobozdravstvo	0	0	13
Iščete delo, delavca?	0	0	557
Za ljudi z mejno osebnostno motnjo	0	0	12
Vrtičkarstvo, zelišča in urejanje okolice	0	0	83

Mali oglasi	0	0	953
Optika	0	0	161
Poroka	0	0	50
Glasba, film in fotografija	0	0	156
Starejši so modrejši	0	0	81
Gradimo, opremljamo in urejamo dom	0	0	830
Avtomobilizem	0	0	357

8 Vzorčenje forumskih tem o raku

Glede na zgornje ugotovitve in strukturo objavljanja sporočil v spletni zdravstveni skupnosti MedOverNet je bila sprejeta odločitev, da se za preučevanje kvalitete sporoči o raku v spletnih zdravstvenih sporočil vzorči forumske teme. Odločitev izhaja iz predvsem iz dveh ugotovitev:

- Razprave v spletnih zdravstvenih skupnostih pogosto potekajo na nestrukturiran način, kjer eno sporočilo lahko predstavlja vprašanje, odgovor, spodbudo za nadaljnjo razpravo, delno pojasnitev v odnosu do drugega sporočila ipd. Ker se vzpostavljajo zelo različne relacije med sporočili, je za analizo vsebine ključna tudi vsebina predhodnih in sledečih sporočil nekemu sporočilu, ki je bil identificiran kot sporočilo, ki se navezuje na temo raka. Zato je pomembno, da se za analizo vsebine upošteva celotno forumsko temo in znotraj ročno kodira relevantna sporočila.
- Relativno nizek odstotek zaznanih sporočil s temo raka v forumskih temah, ki se eksplicitno navezujejo na temo raka, sugerira, da uporabljeni pristopi sporočil s temo raka ne zaznajo dovolj učinkovito.

Predlog izbora ustreznih forumskih tem za vzorčenje je opisan v tabeli spodaj ter v naslednjem odstavku.

Iz baze odstranimo neustrezne (na primer Hrvaške forumske teme). Nato odstranimo vse forumske teme z enim sporočilom. Nato odstranimo vse forumske teme, ki imajo več kot 101 sporočil. Na koncu odstranimo vse forumske teme, ki so v forumih, ki imajo manj kot 10 forumskih tem (z drugimi besedami: obdržimo samo forume z več kot 9 sporočili).

Tabela 20 Izbor ustreznih forumskih tem za vzorčenje forumskih tem s temo raka

Raven vzorčenja	*Ustreznost*	*Število enot*
Izbor forumov	upoštevamo tiste forume, ki imajo vsaj 10 forumskih tem z 2 sporočiloma ter največ 100 sporočili in vsaj 1 sporočilom o raku	vsi ustrezni forumi
Izbor forumskih tem po forumih	upoštevamo tiste forumske teme, ki imajo vaj 2 sporočila in največ 100 sporočil ter vsaj 1 sporočilo o raku	5 forumskih tem (ostale ustrezne forumske teme bodo za nadomestno vzorčenje)

Za ilustracijo smo izvedli vzorčenje forumskih tem po zgoraj opisanem postopku. Največji upad števila forumskih tem je bil zabeležen po odstranitvi vseh forumskih tem, ki nimajo sporočila o raku ter po odstranitvi forumskih tem z manj kot 10 sporočili (Slika 5). Izbor vsebuje 46 forumov (v vsakem slučajno izbranih 5 forumskih tem), 230 forumskih tem ter 2481 sporočil (Tabela 21).

Slika 5 Primer števila forumov in forumskih tem na različnih korakih izbora forumskih tem

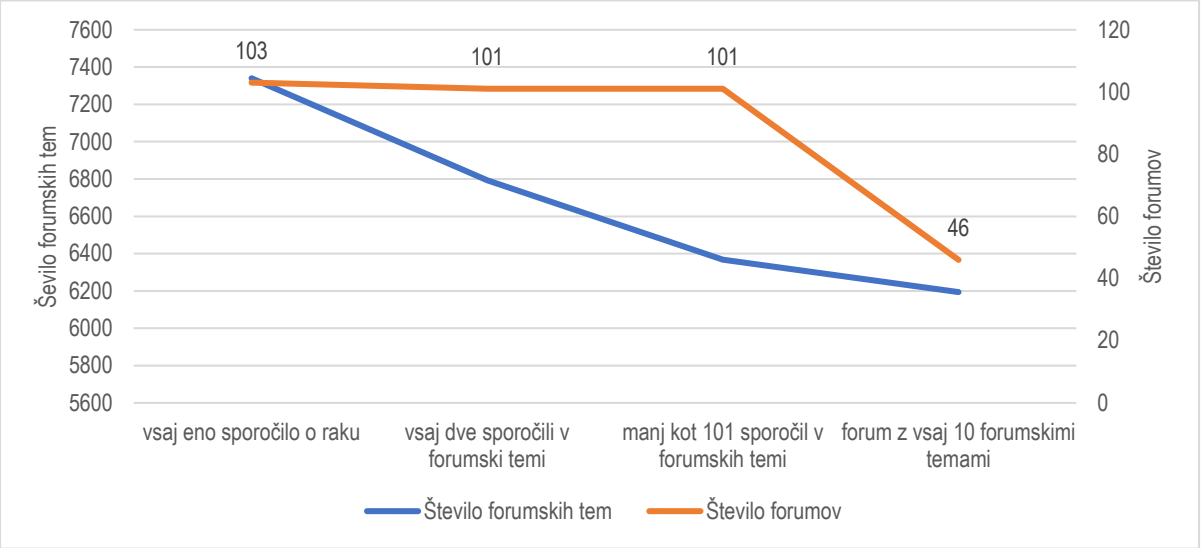


Tabela 21 Vzorec izbranih forumov

Forum	*Vrsta foruma*	*Število tem*	*Število sporočil*	*Število sporočil o raku*
ARHIV – Zdravstvo	Drugo	14	267	18
Nega, ličenje in style	DF	194	1411	753
Politični čvek	DF	83	308	100
Domače živali in veterina	DF	82	292	101
Opravljivka	DF	37	106	60
Sex v mestu	DF	32	120	37
Mladi, ljubezen in spolnost	DF	23	78	28
Kam in kako na počitnice	DF	12	95	14
Starčevski čvek	DF	11	220	12
Ženski in moški čvek	DF	11	163	21
Ločitev, nezvestoba in ljubezenski nasveti	PS	119	483	165
Čustvena inteligenca, komunikacija in odnosi	PS	75	384	107
Nosečnost in obporodno obdobje	PS	46	129	56
Kontracepcija	PS	36	1137	63
Neplodnost	PS	25	109	30
Kako živeti z rakom?	PS	22	110	40
Limfom in levkemija	PS	14	453	18
Multipla skleroza	PS	12	43	13
Starševstvo in vzgoja	PS	11	398	19
Farmacevt svetuje	ZP	3651	99146	10152
Kronična bolečina - fibromialgija	ZP	586	3419	1010
Manualna medicina - zdravljenje gibalnega sistema	ZP	232	1161	344
Hormonske motnje	ZP	159	842	300
Partnerska in družinska posvetovalnica	ZP	153	593	223
Cepljenja	ZP	68	166	133
Intimna nega in zdravljenje vaginalnih okužbe	ZP	67	316	153
Pediatrija	ZP	44	452	111
Urgenca	ZP	44	141	62
Kardiologija	ZP	35	230	43

Estetska in plastična kirurgija	ZP	29	938	53
Neonatologija	ZP	28	109	45
Depresija, anksioznost in razpoloženjske motnje	ZP	26	804	55
Otorinolaringologija	ZP	23	251	32
Bolezni prebavil (gastroenterologija)	ZP	22	81	24
ABC ginekologije in porodništva	ZP	21	107	21
Proktologija	ZP	19	113	30
Preventivni pregled dojk	ZP	17	111	32
Ortopedija	ZP	17	79	26
Pravna pomoč bolnikom z rakom	ZP	16	85	21
Dermatologija	ZP	14	212	33
ARTHRON Klinika za sklepe, hrbtenico in poškodbe	ZP	12	45	14
Astma in druge pljučne bolezni	ZP	11	302	14
Nevrokirurgija	ZP	11	29	12
Psihiater odgovarja	ZP	10	478	58
Hujšanje	ZP	10	215	18
Družinska medicina	ZP	10	59	10

9 Zaključek

Pričujoče raziskovalno poročilo je del delovnega sklopa DS3 projekta *Priložnosti in nevarnosti spletnih zdravstvenih skupnosti za zdravstvo*, ki jo financira Agencija za raziskovalno dejavnost Slovenije (ARRS). Namen tega dela raziskave je bil identificirati in urediti podatkovno bazo z vsemi diskusijami, ki se vsebinsko nanašajo na temo bolezni raka v spletni zdravstveni skupnosti MedOverNet, ki v času nastanka poročila sodi med najpogosteje obiskana spletna mesta v Sloveniji.

Analiziranih je bilo 2.982.564 sporočil, ki so bile objavljene v različnih forumih spletne zdravstvene skupnosti v letih med vključno 2015 in vključno 2019. Za ugotavljanje prisotnosti teme raka v posameznih sporočilih so bili uporabljeni različni pristopi nadzorovanega strojnega učenja, ki so v preteklosti že bili uporabljeni na področju analize besedil: naivni Bayesov klasifikator, metoda podpornih vektorjev, metoda odločitvenih dreves in metoda slučajnih gozdov. Učna množica podatkov (sporočil) je bila določena ročno (sporočila z rakom so bila ročno izbrana s pomočjo ključnih besed, za katere smo domnevali, da se pogosto pojavljajo v sporočilih, ki vsebujejo temo raka), pred uporabo pristopov strojnega učenja pa je bilo besedilo ustrezno urejeno (besedilo je bilo prečiščeno html znakov in ostalih znakov, opravljena je bila lematizacija, odstranjene so bile stop besede). Ocenjujemo, da bi bilo uspešnost uporabljenih metod še mogoče izboljšati z bolj optimalno izbiro parametrov metod ter z bolj optimalno določitvijo matrike dokumentov in značilk, a so bili rezultati bolj parsimonega pristopa zadovoljivi, zato (ob upoštevanju dejstva, da gre za zelo velik obseg besedila) nadaljnjih prilagoditev nismo izvedli.

Kot najbolj učinkovit se je izkazal model slučajnih gozdov z 92 % občutljivostjo in specifičnostjo ($F1 = 89$ %). Izmed 2.979.905 sporočil (število je nižje od izhodiščnega zaradi izločitve nekaterih forumov) jih je 17701 (0,6 %) vsebovalo temo raka. Največji delež sporočil s temo raka je pričakovano v skupini forumov »zdravstvena posvetovalnica« (2,9 %), najmanjši delež pa v skupni družabnih forumov (0,5 %). V skupini forumov podpornih skupin je delež sporočil, ki vsebujejo temo raka 1,93 %. Glede na posamezne forume je največji delež sporočil s temo raka v forumih Preventivni pregled dojk (70,6 %), Kako živeti z rakom? (49,0 %), Limfom in levkemija (40,8 %), Prva pomoč bolnikom z rakom (33,7 %), Zdravljenje s konopljo (20,2 %), Urologija (16,0 %), Radiologija (13,2 %) in Bolezni prebavil (gastoenterologija) (10,6 %).

Sklep za odvisni naslednji delovni sklop (DS4 – Analiza vsebine za merjenje kvalitete spletnih z zdravjem povezanih informacij) je, da je potrebno vzorčiti forumske teme in znotraj njih ročno identificirati sporočila (izmed vseh sporočil v forumski temi), ki se nanašajo na bolezen raka.

10 Literatura

Aizawa, Akiko. 2003. “An Information-Theoretic Perspective of Tf–Idf Measures.” *Information Processing & Management* 39(1):45–65.

Atanasova, Sara. 2020. *Kvalitativna Raziskava Projekta Priložnosti in Nevarnosti Spletnih Zdravstvenih Skupnosti Za Zdravstvo : Raziskovalno Poročilo*. Ljubljana: Fakulteta za družbene vede, Center za metodologijo in informatiko.

Basu, A., C. Walters, and M. Shepherd. 2003. “Support Vector Machines for Text Categorization.” P. 7 pp.- in *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*.

Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45(1):5–32.

Breiman, Leo, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. 1984. *Classification and Regression Trees*. CRC press.

Breiman, Leo, and Nong Shang. 1996. “Born Again Trees.” *University of California, Berkeley, Berkeley, CA, Technical Report* 1:2.

Chen, Hao, and Tin Kam Ho. 1999. “Evaluation of Decision Forests on Text Categorization.” Pp. 191–99 in *Document Recognition and Retrieval VII*. Vol. 3967. International Society for Optics and Photonics.

Cousineau, Tara M., Diana Rancourt, and Traci Craig Green. 2006. “Web Chatter before and after the Women’s Health Initiative Results: A Content Analysis of on-Line Menopause Message Boards.” *Journal of Health Communication* 11(2):133–147.

Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. “Indexing by Latent Semantic Analysis.” *Journal of the American Society for Information Science* 41(6):391–407.

Erčulj, Vanja. 2020. “Analiza Diskusij Spletnih Podpornih Skupin z Metodami Strojnega Učenja Za Namen Pridobivanja Informacij o Psiholoških Vidikih Zdravljenja.” University of Ljubljana, Ljubljana.

Grcar, Miha, Simon Krek, and Kaja Dobrovoljc. 2012. “Obeliks: Statisticni Oblikoskladenjski Oznacevalnik in Lematizator Za Slovenski Jezik.” in *Zbornik Osme konference Jezikovne tehnologije, Ljubljana, Slovenia*.

Hand, David J., Professor in the Department of Statistics David J. Hand, Heikki Mannila, and Padhraic Smyth. 2001. *Principles of Data Mining*. MIT Press.

Hoffman-Goetz, Laurie, Lorie Donelle, and Maria D. Thomson. 2009. “Clinical Guidelines about Diabetes and the Accuracy of Peer Information in an Unmoderated Online Health Forum for Retired Persons.” *Informatics for Health and Social Care* 34(2):91–99.

Hwang, Kevin O., Allison J. Ottenbacher, Angela P. Green, M. Roseann Cannon-Diehl, Oneka Richardson, Elmer V. Bernstam, and Eric J. Thomas. 2010. “Social Support in an Internet Weight Loss Community.” *International Journal of Medical Informatics* 79(1):5–13.

Jones, Karen Sparck. 1972. “A Statistical Interpretation of Term Specificity and Its Application in Retrieval.” *Journal of Documentation*.

Kılınç, Deniz, Akın Özçift, Fatma Bozyigit, Pelin Yıldırım, Fatih Yücalar, and Emin Borandag. 2017. “TTC-3600: A New Benchmark Dataset for Turkish Text Categorization.” *Journal of Information Science* 43(2):174–85.

Kononenko, Igor, and Matjaž Kukar. 2007. “Chapter 3 - Machine Learning Basics.” Pp. 59–105 in *Machine Learning and Data Mining*, edited by I. Kononenko and M. Kukar. Woodhead Publishing.

Leopold, Edda, and Jörg Kindermann. 2002. “Text Categorization with Support Vector Machines. How to Represent Texts in Input Space?” *Machine Learning* 46(1):423–44.

Liu, Bingwei, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. 2013. “Scalable Sentiment Classification for Big Data Analysis Using Naïve Bayes Classifier.” Pp. 99–104 in *2013 IEEE International Conference on Big Data*.

Macias, Wendy, Liza Stavchansky Lewis, and Tenikka L. Smith. 2005. “Health-Related Message Boards/Chat Rooms on the Web: Discussion Content and Implications for Pharmaceutical Sponsorships.” *Journal of Health Communication* 10(3):209–223.

Mo, Phoenix KH, and Neil S. Coulson. 2008. “Exploring the Communication of Social Support within Virtual Communities: A Content Analysis of Messages Posted to an Online HIV/AIDS Support Group.” *Cyberpsychology & Behavior* 11(3):371–374.

MOSS. 2020. “Med.Over.Net.” *MOSS, Merjenje Obiskanosti Spletnih Strani*. Retrieved September 16, 2020 (https://www.moss-soz.si/spletna-stran/med-over-net/).

Patil, Dipti D., V. M. Wadhai, and J. A. Gokhale. 2010. “Evaluation of Decision Tree Pruning Algorithms for Complexity and Classification Accuracy.” *International Journal of Computer Applications* 11(2):23–30.

Robertson, Stephen. 2004. “Understanding Inverse Document Frequency: On Theoretical Arguments for IDF.” *Journal of Documentation* 60(5):503–20.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. “Estimating the Number of Clusters in a Data Set via the Gap Statistic.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2):411–423.

Ting, S. L., W. H. Ip, and Albert H. C. Tsang. 2011. “Is Naïve Bayes a Good Classifier for Document Classification?” *International Journal of Software Engineering and Its Applications* 5(3):37–46.

Wang, Lipo. 2005. *Support Vector Machines: Theory and Applications*. Springer Science & Business Media.

Williams, Graham. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. New York: Springer-Verlag.

Wu, Qingyao, Yunming Ye, Haijun Zhang, Michael K. Ng, and Shen-Shyang Ho. 2014. “ForesTexter: An Efficient Random Forest Algorithm for Imbalanced Text Categorization.” *Knowledge-Based Systems* 67:105–16.

Xue, Dashen, and Fengxin Li. 2015. “Research of Text Categorization Model Based on Random Forests.” Pp. 173–76 in *2015 IEEE International Conference on Computational Intelligence Communication Technology*.

Zhang, Wen, Taketoshi Yoshida, and Xijin Tang. 2008. “Text Classification Based on Multi-Word with Support Vector Machine.” *Knowledge-Based Systems* 21(8):879–86.