

Keliamo turinio strategijos analizė socialinio tinklo „Facebook“ puslapiuose

Dominykas Eirošius

Vilniaus universitetas, Matematikos ir informatikos fakultetas,
Didlaukio g. 47, LT-08303 Vilnius
dominykas.eirosius@mif.stud.vu.lt

Santrauka. Straipsnyje yra pristatoma socialinio tinklo „Facebook“ puslapiuose keliamo turinio strategijos analizė. Turinio strategija apibūdinama kaip turinio publikavimo planavimas, kūrimas pasirinktomis tematikomis bei kitoks turinio valdymas. Ši analizė buvo atlikta analizuojant tris „Facebook“ puslapius ir naudojantis latentiniu Dirichlė paskirstymo algoritmu, skirtu temų modeliavimui. Temų modeliavimas padeda analizuoti ir geriau suprasti nestruktūrizuotų tekstų rinkinius, kurie savyje slepia daugybę naudingos informacijos. Tyrimo metu, iš beveik 9 tūkst. vaizdo įrašų pavadinimų buvo sudarytos 10 temų, kurios leido detaliau pažvelgti į keliamą turinį. Išanalizavus duomenis buvo atskleistos keliamo turinio strategijos bei jų pasikeitimai nagrinėjamu laikotarpiu.

Raktiniai žodžiai: socialiniai tinklai; „Facebook“; latentinis Dirichlė paskirstymas; temų modeliavimas.

1 Įvadas

Temų modeliavimas yra vienas iš galingiausių teksto apdorojimo metodų, skirtų latentiniams duomenims aptikti bei tekstinių dokumentų ir kitų duomenų ryšiams rasti. Mokslininkai yra paskelbę daugybę įvairių temų modeliavimo straipsnių įvairiausiose srityse [1], [2], [3], [4], [5]. Šiam uždaviniui atlikti yra įvairių metodų, tačiau latentinis Dirichlė paskirstymas (angl. Latent Dirichlet allocation (LDA)) yra vienas populiariausių naudojamų algoritmų, leidžiančių išskirti iš žodžių rinkinių pagrindines temas be papildomos žmogaus intervencijos.

LDA galime įvardinti kaip metodą, kuris automatiškai suformuoja temas, kurioms priklauso tam tikri žodžiai. Kitaip tariant, tai yra temų, kurios sudarytos iš žodžių su tam tikromis tikimybėmis, rinkinys. Taigi, įvardinto metodo tikslas – iš žodžių ir dokumentų sukurti tiesiogiai nestebimą (latentinę) temą. Iš struktūrizuoto tekstų rinkinio (angl. corpus) LDA pirmiausia atkuria esančius dokumentus. To veikimas yra paremtas iš žodžių ir dokumentų

sudarytų temų svarba [6]. Prieš nurodant LDA specifikacijas, apibrėžkime sąvokas:

- Pagrindinis diskrečių duomenų vienetas yra žodis $\{1, \dots, V\}$, kuris įvardinamas kaip svarbiausias reikšminis kalbos vienetas. Žodžiai yra vaizduojami kaip vienetiniai vektoriai, o v -asis žodis yra vaizduojamas vektoriumi w , kur $w^v = 1$ ir $w^u = 0$, kai $v \neq u$.
- Dokumentas apibrėžiamas kaip N žodžių seka $\mathbf{w} = (w_1, w_2, \dots, w_N)$, kur w_n yra n -tasis sekos žodis.
- Struktūrizuotas teksto rinkinys (angl. corpus) yra M dokumentų kolekcija, žymima $D = \{w_1, w_2, \dots, w_M\}$.

LDA generavimo proceso specifikacijas kiekvienam dokumentui \mathbf{w} , esančiame struktūrizuotame teksto rinkinyje D galime užrašyti taip:

1. $N \sim \text{Poisson}(\xi)$
2. $\theta \sim \text{Dir}(\alpha)$ – dokumento temos skirstinys.
3. Kiekvienam žodžiui w_n iš N :
 - 3.1. Pasirenkama tema $z_n \sim \text{Multinomial}(\theta)$.
 - 3.2. Pasirenkamas žodis w_n iš $p(w_n | z_n, \beta)$ – tikimybės, sąlygotos temos z_n .

Taip pat modelyje yra daromos ir kelios prielaidos. Dirichlė skirstinio dimensija k ir temos kintamojo dimensija z yra fiksuoti ir žinomi. Žodžių tikimybės apibrėžiamos kaip $k \times V$ matrica β , kur $\beta_{ij} = p(w^j = 1 | z^i = 1)$ yra laikomu fiksuotu kiekiu. Tuo tarpu N nepriklauso nuo visų duomenis generuojančių kintamųjų, tokių kaip θ ir z .

θ yra k -matis Dirichlė atsitiktinis kintamasis, kuris įgyja reikšmes $(k - 1)$ -simplekse (k - vektorius θ įgyja reikšmes $(k - 1)$ -simplekse, jei $\theta_i \geq 0$, $\sum_{i=1}^k \theta_i = 1$, ir turi tikimybių tankį, lygų:

$$p(\theta | \alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \dots \theta_k^{\alpha_k - 1},$$

kur parametras α yra k -vektorius, kurio komponentai $\alpha_i > 0$, o $\Gamma(x)$ yra gama funkcija. Turėdami α ir β parametrus, bendras θ , temų rinkinio \mathbf{z} ir \mathbf{w} paskirstymas išreiškiamas:

$$p(\theta, \mathbf{z}, \mathbf{w} | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta),$$

kur $p(z_n | \theta)$ yra θ_i unikaliam i , kad $z_n^i = 1$. Integruodami θ ir sumuodami z , gauname ribinį dokumento skirstinį:

$$p(\mathbf{w}|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta.$$

Galiausiai, paėmę atskirų dokumentų ribinių tikimybių sandaugą, gauname struktūrizuoto teksto rinkinio tikymbę:

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{d=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d.$$

Šio tyrimo tikslas – pasinaudojant latentiniu Dirichlė paskirstymo algoritmu ištirti trijų, socialinio tinklo „Facebook“, puslapių keliamo turinio strategijas. „Facebook“ puslapis yra apibūdinamas kaip profilis, kurį gali sukurti įmonės, organizacijos ar individualūs asmenys. Šiuose puslapiuose gali būti skelbiama įvairi informacija, kuri yra prieinama ir matoma visiems socialinės platformos vartotojams [7]. Socialinių tinklų kontekste, turinys apima ne tik puslapyje rašomus tekstus, bet ir naudojamas nuotraukas bei vaizdo įrašus. Todėl turinio strategija apibūdinama kaip turinio publikavimo planavimas, kūrimas pasirinktomis tematikomis bei kitoks turinio valdymas [8]. Tyrimas buvo atliekamas trimis etapais: pirmiausia duomenys buvo paruošiami analizei: išvalytos trūkstamos reikšmės bei sukurti papildomi išvestiniai kintamieji. Po to atliekamas vaizdo įrašų temų modeliavimas. Šiame etape sutvarkomos tekstinio kintamojo reikmės, atlikta tokenizacija ir žodžių redukavimo procesas iki žodžių kamienų. Sekančiu žingsniu buvo pritaikytas latentinio Dirichlė paskirstymo algoritmas ir susisteminamos keliamo turinio tematikos. Galiausiai, buvo atliekama „Facebook“ keliamo turinio skirtingais pjūviais statistinė analizė bei pateikiamos gautos išvados.

2 Duomenų aprašas

Tyrimo metu buvo naudojami duomenys, ištraukti iš „Tubular“ platformos [9]. Ši platforma yra skirta vaizdo įrašų parametrų sekimui ir jų analizėms. Naudojama duomenų imtis yra subalansuota, kurią sudaro 14 kintamųjų: 4 kokybiniai bei 10 kiekybiniai, ir 8792 skirtingų vaizdo įrašų iš 3 skirtingų „Facebook“ puslapių. Kintamieji apibūdina vaizdo įrašų pagrindinę informaciją: universalų identifikacijos numerį, puslapio ir vaizdo įrašo pavadinimą, paskelbimo laiką, įrašo trukmę, peržiūrų skaičių bei auditorijos skirtingų interakcijų kiekį.

Pradėjus analizuoti duomenis buvo pastebėtas papildomų išvestinių kintamųjų poreikis, be kurių būtų sudėtinga atlikti analizę. Vienas iš trūkstamųjų kintamųjų - bendras vaizdo įrašo surinktų reakcijų skaičius. 2016 m. „Facebook“ įvykdė pakeitimus savo platformoje ir išleido naują ypatybę „Reakcijas“ [10]. Ši funkcija yra esamos funkcijos paspausti mėgti ant vaizdo įrašo papildymas: vartotojai dabar gali pasirinkti vieną iš penkių papildomų animuotų jaustukų, kuriais gali išreikšti savo požiūrį į pamatytą vaizdo įrašą. Taigi, dabar egzistuoja šešios tokios reakcijos: „Patinka“, kurį jau yra žinoma, taip pat „Myliu“, „Haha“, „Vau“, „Liūdna“ ir „Pikta“ emocijos. Vis dėlto, pradinuose duomenyse yra pateikiamas tik kintamasis, kuris parodo, kiek žmonių paspaudė mėgti mygtuką. Trūkstant duomenų apie kitas reakcijas negalima tiksliai įvertinti, kaip „Facebook“ vartotojai reagavo į vaizdo įrašus ir kiek susidomėjimo įrašas sulaukė. Šiai problemai spręsti buvo pasitelktas atliktas tyrimas, kurio metu buvo nustatytas vidutinis reakcijų pasiskirstymas vaizdo įrašuose [11]. Buvo nustatyta, jog mėgti reakcija sudaro 77,55 proc. visų reakcijų, paliekant 22,45 proc. visoms kitoms reakcijoms. Remiantis šio tyrimo rezultatais ir proporcijos lygtimis, buvo sukurtas naujas kintamasis, parodantis bendrą vaizdo įrašo surinktų reakcijų skaičių.

Dar vienas kintamasis, kuris buvo sukurtas siekiant analizuoti bendrą vartotojų interakciją su vaizdo įrašų, yra bendras visų interakcijų skaičius. Socialinių tinklų kontekste, interakcija yra bet koks vartotojo atliktas veiksmas, susijęs su vaizdo įrašu. Taigi, šis kintamasis buvo sudarytas sudedant vaizdo įrašo komentarus, pasidalinimus ir bendrą reakcijų skaičių. Taip pat buvo sukurti išvestiniai kintamieji, kurie padėjo atlikti analizę ir parodė vidutinę vaizdo įrašo trukmę minutėmis, įrašo ilgio kategoriją bei kurį metų ketvirtį, savaitės dieną bei valandą buvo publikuotas vaizdo įrašas.

Galiausiai buvo nustatyta, jog duomenyse yra trūkstamų reikšmių. Skirtingi kintamieji turėjo skirtingą kiekį trūkstamų reikšmių, tačiau jos nesudarė daugiau kaip 3,9 proc. visos imties. Kadangi trūkstama informacija sudarė sąlyginai nedidelę dalį visų duomenų, nepilni įrašai buvo panaikinti.

3 Eksperimentinis tyrimas

Siekiant įvertinti puslapiuose keliamo turinio strategiją ir nustatyti jos pasikeitimus laikui bėgant, pirmiausia siekta nustatyti, koks yra keliamas turinys. Dėl šios priežasties buvo atliekama keliamų įrašų temų analizė. Temų sukūrimui ir raktinių žodžių generavimui buvo pasitelkti vaizdo įrašų pavadinimų duomenys.

Pirmiausiai, buvo atliktas tokenizavimas. Tokenizavimas yra eilučių sekos suskaidymas į dalis, tokias kaip žodžiai, raktiniai žodžiai, frazės, simboliai ir kiti elementai, vadinami žetonais (angl. Tokens). Žetonai gali būti atskiri žodžiai, frazės ar net ištisi sakiniai. Taigi, kiekvienas sakinyss buvo suskirstytas į žodžių sąrašą, visiškai pašalinti skyrybos ženklai ir nereikalingi simboliai. Atlikus tokenizaciją buvo atliktas kamienuų sudarymas. Tai yra žodžio redukavimo procesas iki žodžio kamieno, kitaip žinomo kaip lema, kuris pridedamas prie priesagų ir priešdėlių. Tokio proceso tikslas yra sumažinti bendrą žodžių skaičių žodyne ir palikti tik unikalius žodžius. Dėl kamienuų sudarymo temų modeliavimas turėtų pavykti geriau ir turėtų būti sukurtos tikslinės temos. Paskutiniame teksto apdorojimo etape buvo sukurta žodžių dokumento matricą (angl. Document-Word matrix) ir pašalinti angliški stabdomieji žodžiai (angl. stopwords), puslapio ir vaizdo įrašo pavadinimą, paskelbimo laiką, įrašo trukmę, peržiūrų skaičių bei auditorijos skirtingų interakcijų kieki.

Sukūrus žodžių dokumento matricą buvo pasinaudota latentinio Dirichlė paskirstymo modeliu. Natūralios kalbos apdorojimo (angl. Natural language processing) kontekste latentinis Dirichlet paskirstymas (angl. Latent Dirichlet allocation) yra statistinis modelis. Tai yra temų modelio algoritmas, kuris naudojamas tekstus priskirti tam tikroms temoms. Modelio veikimas yra paremtas prielaida, kad dokumentai yra sudaryti iš žodžių, kurie padeda nustatyti temas. Šia prielaida vadovaujantis yra bandoma susieti dokumentus su temų rinkiniu, priskiriant kiekvieną dokumento žodį skirtingoms temoms. Toks priskyrimo procesas yra atliekamas sąlyginių tikimybių įverčių pagalba. Taigi, įvertinus žodžiams tikimybes, yra sudaromas žodžių rinkinys, kuris atvaizduoja tam tikrą temą. Rinkinys yra sudaromas pasirenkant žodžius su didžiausiomis tikimybėmis arba nustatant tikimybinį slenkstį ir pasirenkant tik tuos žodžius, kurių tikimybė yra didesnė arba lygi pasirinktai slenkstinei vertei. Tyrime buvo pasirinkta surasti ir sudaryti 10 temų, nes nustatant kitokį kiekį modelio tikslumas mažėja. Pritaikius latentinio Dirichlė paskirstymo modelį buvo gauta temų žodžių matrica, kuri pateikiama 1 pav.

Iš temų žodžių matricos buvo sudarytos šios temos: „Handicrafts & Food“, „Makeover & Changeover“, „Construction & Water“, „Furniture & Construction Tools“, „Life hacks & Tricks“, „Makeover & Renovation“, „Makeup & Garden decor“, „Wood crafts & Art“, „Design & Decoration“, „Satisfaction & Repair“.

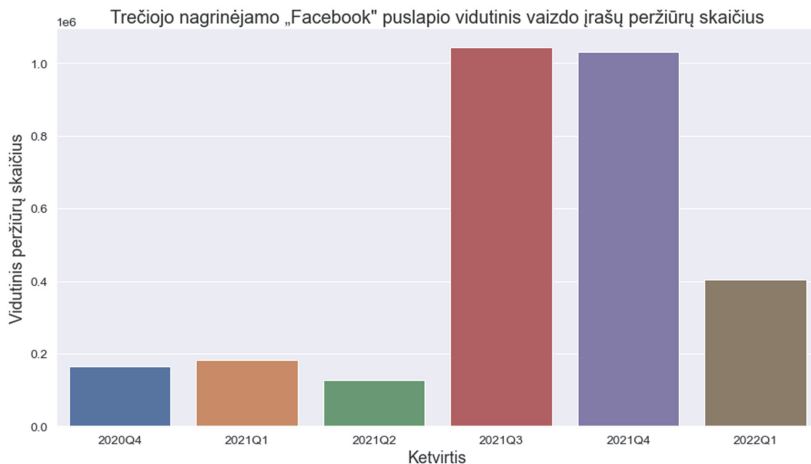
	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10	Word 11	Word 12	Word 13	Word 14
Topic 0	turn	knife	vase	bowl	food	bike	box	wood	steel	shell	coin	piece	building	cast	tank
Topic 1	diy	hair transformation	house	aquarium	brick	shape	axe	soap	concrete	pallet	use	stove	pen	hand	dress
Topic 2	build	transform	house	makeup	brick	backyard	log	water	rebar	wood	bicycle	project	waterfall	concrete	wheel
Topic 3	restoration	table	make	fashion	glass	work	coffee	style	gun	drill	grill	hammer	dam	chain	toy
Topic 4	hack	tip	life	know	try	trick	car	look	learn	weld	fix	repair	leather	clothe	kitchen
Topic 5	make	use	epoxy	makeover	tool	nail	metal	wood	chair	recycle	resin	bend	cake	bolt	time
Topic 6	idea	craft	cement	technique	tip	makeup	hairstyle	shoe	decor	garden	need	wood	plant	welding	check
Topic 7	make	lamp	pipe	woodturne	wood	art	night	ring	cut	cardboard	science	experiment	candy	man	paper
Topic 8	tip	way	restore	candle	create	construction	egg	design	decoration	lock	wire	door	girl	gift	guy
Topic 9	home	make	process	satisfy	beauty	resin	gadget	paint	repair	cleaning	iron	problem	extension	grinder	tile

1 pav. Žodžių matrica.

4 Keliamo turinio strategijos analizė skirtingais pjūviais

Turinio strategija buvo nagrinėjama pagal: vidutinį vaizdo įrašo peržiūrų skaičių kiekvieną ketvirtį, kiekvieno ketvirčio savaitės dienomis Vaizdo įrašų publikavimo pasiskirstymą, vaizdo įrašų publikavimo laiką kiekvieną ketvirtį, vidutinį vaizdo įrašų peržiūrų skaičių pagal vaizdo įrašų trukmę, keliamų vaizdo įrašų kategorijų pasiskirstymą kiekvieną ketvirtį, vidutinį įrašų interakcijų kiekį kiekvieną ketvirtį, vidutinę vaizdo įrašų trukmę kiekvieną ketvirtį, kiekvieną ketvirtį įkeliamų vaizdo įrašų temų pasiskirstymą, vidutinį vaizdo įrašų peržiūrų skaičių pagal temas.

Ištyrus „Facebook“ konkurentų kanalus išsiaiškinta, jog vidutinis vaizdo įrašų peržiūrų skaičius kiekvieną ketvirtį pirmame kanale nebuvo pastovus, tačiau ženkliai nesikeitė. 2 pav. yra pateikiamos trečiojo nagrinėjamo „Facebook“ puslapio vidutinės vaizdo įrašo peržiūros. Taip pat pastebėta, jog antro ir trečio kanalų grafikai buvo labai panašūs. Įdomu tai, jog 2021 m. trečio ir ketvirto ketvirčio vidutinis peržiūrų skaičius šiuose kanaluose buvo nepaprastai didelis. Tokio teigiamo rezultato priežastys gali būti sėkmingos marketinginės kampanijos, labai kokybiškų ir sudominančių vaizdo įrašų sukūrimas.



2 pav. Trečiojo nagrinėjamo „Facebook“ puslapio vidutinis vaizdo įrašo peržiūrų skaičius.

Žvelgiant į kiekvieno ketvirčio savaitės dienų vaizdo įrašų publikavimo pasiskirstymą aiškiai matyti, jog pirmojo nagrinėjamo puslapio visų ketvirčių keliamų įrašų strategija buvo išlaikyta tokia pati. Visomis dienomis buvo stengiamasi kelti po panašų kiekį vaizdo įrašų. Antrasis nagrinėjamas puslapis 2020 m. ketvirtame ketvirtyje didesnę kiekį įrašų stengėsi kelti antradieniais, trečiadieniais ir ketvirtadieniais, o trečiasis – pusę keliamų įrašų publikuodavo ketvirtadieniais, likusią dalį po lygiai antradieniais ir sekmadieniais. Tačiau galima pastebėti, jog tokia strategija nepasitvirtino, nes likusiais ketvirčiais įrašai buvo keliami apytiksliai po lygiai kiekvieną dieną.

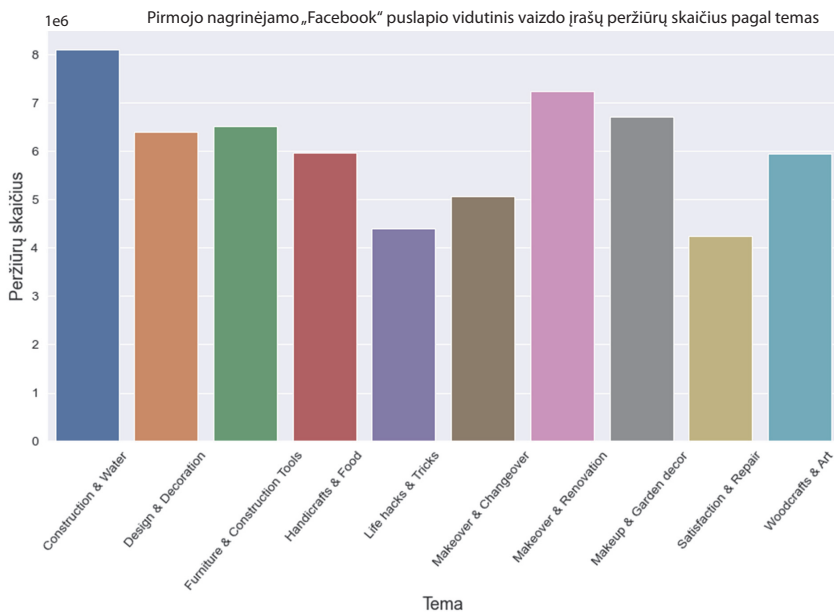
Analizuojant vidutinį vaizdo įrašų peržiūrų skaičių pagal vaizdo įrašų trukmę pastebėta, jog pirmuose dvejuose kanaluose didžiausio susidomėjimo sulaukė vidutinės trukmės vaizdo įrašai. Paskutiniojo puslapio lankytojams labiausiai patiko stebėti trumpesnius vaizdo įrašus. Svarbu paminėti tai, kad visuose kanaluose labiausiai dominuoja labai trumpi vaizdo įrašai. Vis dėlto galima daryti prielaidą, kad kanalų moderatoriai atsižvelgė į žiūrovų susidomėjimą ilgesniais įrašais, nes vėlesniuose ketvirčiuose buvo pradedami publikuoti ir ilgesnės trukmės filmukai.

Pradėjus nagrinėti kanalų vidutinį įrašų interakcijų kiekį iš karto pastebėtas pirmojo puslapio publikuojamų įrašų prastėjantis rezultatas. Matoma, kad auditorijos sąveikavimas su vaizdo įrašais palaipsniui mažėja. Tai gali indikuoti žiūrovų nepasitenkinimą turinio kokybe. Nors šiame kanale nėra stebimas vaizdo įrašų peržiūrų skaičiaus mažėjimas, interakcijų smukimas atskleidžia, jog ateityje šio kanalo populiarumas ir peržiūrų skaičius taip pat gali smukti. Tuo tarpu kitų puslapių pirmais trimis ketvirčiais sąveikų su vaizdo įrašais buvo labai mažai. Tačiau tokio rezultato galima buvo tikėtis, nes šiais ketvirčiais ir vaizdo įrašai atitinkamai sulaukdavo mažai peržiūrų.

Peržvelgus vidutinę publikuojamų vaizdo įrašų trukmę kiekvieną ketvirtį nepastebėta jokių stebinančių rezultatų. Visuose kanaluose vidutinė trukmė auga. Taip yra todėl, nes kaip jau buvo pasakyta, vėlesniuose ketvirčiuose kanalų savininkai atsižvelgė į faktą, jog didžiausio žiūrovų susidomėjimo sulaukia trumpi ir vidutinės trukmės vaizdo įrašai.

Paskutiniame etape buvo analizuojami duomenys pagal įkeliamų vaizdo įrašų temų pasiskirstymą ir vidutinį vaizdo įrašų peržiūrų skaičių pagal šias sudarytas temas. 3 pav. atskleidžia, jog pirmame kanale vaizdo įrašai keliami apytiksliai po lygiai visomis temomis. Vis dėlto, labiausiai žiūrimos temos buvo šios: „Construction & water“ bei „Makeover & revonation“. Nors antrame puslapyje su lyg kiekvienu ketvirčiu pradėjo pasirodyti vis įvairesnis turi-

nys, tačiau labiausiai publikuojantys vaizdo įrašai buvo „Makeup & Garden decor“ bei „Life hacks & Tricks“ temomis. Įdomu tai, jog kanalo savininkų daugiausiai keliamo turinio temos išsiskyrė su žiūrovams labiausiai patikusiomis temomis. Pasirodo, jog auditorijai labiausiai patiko žiūrėti „Satisfaction & repair“, „Construction & water“ bei „Design & decoration“. Tuo tarpu paskutiniame puslapyje pirmąjį ketvirtį dominavo tik „Satisfaction & repair“, „Design & decoration“ bei „Woodcrafts & art“ keliamų vaizdo įrašų tematikos. Vėliau turinys buvo publikuojamas įvairesnis. Vis dėlto lankytojų mėgstamiausios temos buvo: „Makeup & garden decor“, taip pat „Makeover & renovation“.



3 pav. Pirmojo nagrinėjamo „Facebook“ puslapio vidutinis vaizdo įrašų peržiūrų skaičius pagal temas.

5 Išvados

Tyrimo metu buvo ištirtos trijų, socialinio tinklo „Facebook“, puslapių keliamo turinio strategijos, nustatytas požymių rinkinys su papildomais išvestiniais parametrais, kuris leido įvertinti tiriamų strategijų reikšmingu-

mą. Strategijas sudarė tokie požymiai, kaip publikuojamo turinio laikas bei diena, kuriamų vaizdo įrašų trukmė ir turinio tematikos. Pasinaudojus latentiniu Dirichlet paskirstymo algoritmu buvo sudarytos 10 temų. Vėliau, šiomis temomis buvo pasinaudota analizuojant keliamą turinį. Pastebėta, jog 2021 m. trečias ir ketvirtas ketvirčiai buvo sėkmingi nagrinėjamiems puslapiams. Lyginant su kitais ketvirčiais, šiuo laikotarpiu vidutiniškai buvo surenkama ženkliai didesnis vaizdo įrašų peržiūrų kiekis. Taip pat buvo atskleista, jog stebimu laikotarpiu „Facebook“ puslapiai pradėjo kurti ir publikuoti ilgesnės trukmės vaizdo įrašus. Galiausiai pastebėta, jog puslapiuose pradėjo pasirodyti vis įvairesnis turinys, nors temų populiarumai ir žiūrovų susidomėjimas jomis skyrėsi.

Literatūra

- [1] McAuliffe, Jon, and David Blei. Supervised topic models. *Advances in neural information processing systems*, 20, 2007.
- [2] Li, D., He, B., Ding, Y., Tang, J., Sugimoto, C., Qin, Z., Yan, E., Li, J. and Dong, T. Community-based topic modeling for social tagging. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, 1565-1568, 2010.
- [3] Daud, A. Using time topic modeling for semantics-based dynamic research interest finding. *Knowledge-Based Systems*, 26: 154-163, 2012.
- [4] Lim, K.W., Chen, C. and Buntine, W. Twitter-network topic model: A full Bayesian treatment for social network and text modeling, arXiv preprint arXiv:1609.06791, 2016.
- [5] Weng, J., Lim, E.P., Jiang, J. and He, Q. Twiterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, 261-270, 2010.
- [6] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3: 993-1022, 2003.
- [7] Kelly Kennedy. Facebook Page vs Profile: Everything You Need To Know. <https://yourblogworks.com/facebook-page-vs-facebook-profile/>.
- [8] Kristina Halvorson. The Discipline of Content Strategy. <http://alistapart.com/article/thedisciplineofcontentstrategy/>.
- [9] Tubular. <https://tubularlabs.com/>.
- [10] Sammi Krug. Reactions Now Available Globally. <https://about.fb.com/news/2016/02/reactions-now-available-globally/>.
- [11] Olga Rabo. We Analyzed 2,810 Pages to Calculate Average Facebook Engagement Rate. <https://blog.iconosquare.com/average-facebook-engagement-rate/>.