

Economics and Management Ekonomika ir vadyba

DIDŽIŲJŲ DUOMENŲ PANAUDOJIMAS SOCIALINĖJE MEDIJOJE

Roberta KARPOVIČIŪTĖ*, Jolanta SABAITYTĖ

Vilniaus Gedimino technikos universitetas, Vilnius, Lietuva

Gauta 2019 m. kovo 15 d.; priimta 2019 m. kovo 25 d.

Santrauka. Ilgą laiką literatūroje buvo pabrėžiama socialinių tinklų ir socialinių medijų kaip komunikacijos priemonių nauda ir panaudojimo galimybės. Tobulėjančios technologijos ir interneto sparta lėmė tai, jog socialinių tinklų populiarumas bei vartotojų kuriamo turinio ir duomenų kiekis sparčiai auga. Susidaro palankios sąlygos įmonėms šiuos duomenis analizuoti bei panaudoti priimant strateginius sprendimus. Šio darbo problemine klausimas yra didžiųjų duomenų, kuriuos sugeneruoja socialinės medijos, panaudojimo galimybės rinkodaroje. Straipsnyje analizuojamos didžiųjų duomenų charakteristikos, socialinių medijų rūšys bei jų generuojamų duomenų panaudojimo galimybės ir rizikos bei analizės metodai, sudaromas socialinės medijos sukurtų didžiųjų duomenų panaudojimo koncepcinis modelis. Straipsnyje taikomi mokslinės literatūros ir kitų informacijos šaltinių sisteminės analizės bei apibendrinimo metodai.

Reikšminiai žodžiai: didieji duomenys, didžiųjų duomenų analizė, socialinės medijos, socialiniai tinklai.

Įvadas

Tobulėjant technologijoms ir didėjant informacijos kiekiui susidarė palankesnės sąlygos didiesiems duomenims kaupti ir panaudoti.

Duomenų generavimo ir augimo greitis didėja dėl judriųjų prietaisų ir kitų prie interneto prijungtų prietaisų paplitimo bei interneto spartos. Interneto vartotojų skaičius pasaulyje jau viršijo 4 milijardus vartotojų, taigi daugiau nei 55 % visos žmonijos populiacijos naudojami interneto ryšiu (Internet World Stats, 2018). Didelę reikšmę didžiųjų duomenų projektams turi ir sparčiai populiarėjantys socialiniai tinklai bei kitos socialinės medijos, nes šių technologijų populiarumas sparčiai auga, o generuojami duomenys pasižymi informacijos gausa. Remiantis Statista (2018), kiekvieną minutę interneto vartotojai peržiūri daugiau nei 4 milijonus *Youtube* vaizdo įrašų, atlieka daugiau nei 3 milijonus paieškų *Google* paieškos sistemoje, pasidalina daugiau nei 2 milijonais nuotraukų socialiniame tinkle *Snapchat* bei atlieka daugybę kitų veiksmų, kurie generuoja didžiulius kiekius duomenų. Be to, 2017 m. buvo prognozuojama, jog 2021 m. socialinių medijų vartotojų skaičius viršys 3 milijardų vartotojų ribą (Statista, 2017). Vis dėlto remiantis *WeAreSocial* (2018) ataskaita, ši riba jau buvo peržengta 2018 m., tad akivaizdu, jog socialinių medijų populiarumas ir generuojamų duomenų kiekis auga vis sparčiau.

Tokių duomenų analizė suteikia galimybę gauti vertingų įžvalgų, nes gali atvaizduoti vartotojų elgsenos tendencijas bei kitą su visuomene susijusią informaciją, leidžiančią vykdyti dar tikslesnę analizę (Anshari, Al-munawar, Lim ir Al-Mudimigh, 2018). Įmonė, kuri pažįsta savo klientą, turi didelį konkurencinį pranašumą, nes remdamasi sukauptais duomenimis ir vykdydama jų analizę gali priimti geresnius sprendimus bei efektyviau vystyti produktą. Tokiai organizacijai lengviau užmegzti kontaktą su esamu ar potencialiu vartotoju, suprasti jo poreikius bei prognozuoti elgseną ateityje. Taigi, verslo organizacijos galėtų efektyviau išnaudoti reklamos biudžetus ir pasiekti geresnių rezultatų lyginant su konkurentais, jeigu skirtų daugiau dėmesio auditorijos analizei, pasitelkdamos didžiuosius duomenis ir jų analizės metodus. Socialinių tinklų, tokių kaip *Twitter*, *Facebook* ar *LinkedIn*, bei *Google* ar kitų paieškos sistemų duomenys ne tik auga savo kiekiu ir įvairove, tačiau taip pat prisijungdamos naujas platformas ir kanalus plečia didžiųjų duomenų ir galimybių ribas. Vis dėlto pastebima, jog stokojama įrankių, kurie leistų suvokti didžiųjų duomenų kuriamą vertę ir sudaryti prielaidas jiems panaudoti.

Šio darbo problemine klausimas yra didžiųjų duomenų, kuriuos sugeneruoja socialinės medijos, panaudojimo galimybės rinkodaroje. Darbo tikslas – suformuoti socialinės medijos sukurtų didžiųjų duomenų

*Autorius susirašinėti. El. paštas roberta.karpoviciute@stud.vgtu.lt

panaudojimo koncepcinį modelį. Tikslui pasiekti iškelti šie uždaviniai:

- išanalizuoti didžiųjų duomenų sampratą ir koncepcijos raidą;
- išanalizuoti socialinių medijų generuojamų duomenų apibrėžimą bei klasifikavimo paradigmą;
- išanalizuoti socialinių medijų generuojamų duomenų panaudojimo galimybes;
- sudaryti didžiųjų socialinės medijos generuojamų duomenų panaudojimo koncepcinį modelį.

Darbe taikomi mokslinės literatūros ir kitų informacijos šaltinių sisteminės analizės bei apibendrinimo metodai.

1. Didžiųjų duomenų charakteristikos ir koncepcijos raida

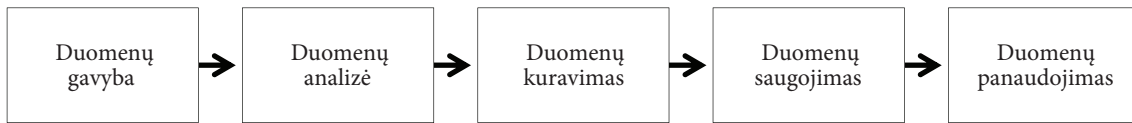
Didėjanti automatizacija, interneto sparta bei išmaniųjų įrenginių paplitimas – visa tai lemia nuolatinį duomenų kiekių augimą. Duomenų kiekiai tapo tokie dideli, jog jiems suvaldyti paprastų duomenų bazių nepakanka, o jiems saugoti ir apdoroti tradiciniai metodai ir programos nebėra tinkami. Tokiems dideliems duomenų kiekiams apibūdinti buvo suformuotas terminas „didieji duomenys“ (angl. „Big Data“).

Teigiama, jog terminas „didieji duomenys“ pirmą kartą buvo paminėtas 1998 m. per įmonės „Silicon Graphics“ atstovo John Mashey pristatymą. Vėliau šis terminas imtas vartoti ir akademinėje literatūroje (Weiss ir Indurkha, 1998). Šiuo metu terminas „didieji duomenys“ yra plačiai paplitęs, tačiau pastebima, kad literatūroje galima rasti įvairių didžiųjų duomenų apibrėžimų, skirtingi autoriai šią sąvoką formuluoja remdamiesi skirtingais aspektais. Pavyzdžiui:

- Oussous, Benjelloun, Ait Lahcen ir Belfkih (2018) didžiuosius duomenis apibrėžia kaip didelį, sparčiai augantį duomenų kiekį, kurie yra sudaryti iš heterogeniškų formatų: struktūruotų, nestruktūruotų ar pusiau struktūruotų duomenų, reikalaujančių galinųjų technologijų ir pažangių algoritmų, nes tradiciniai duomenų tyrybos metodai tokiems duomenų kiekiams nebėra efektyvūs.
- Alharthi, Krotov ir Bowman (2017) didžiuosius duomenis apibrėžia kaip žmonių veiklos sukurtą didelės apimties duomenų kiekį, kurį sudėtinga valdyti, naudojant tradicines analizės priemones.
- Politaitė ir Sabaitytė (2018) teigia, kad didieji duomenys – tai dideli duomenų masyvai, kuriems apdoroti netinka įprastinės duomenų apdorojimo priemonės.
- Yin ir Kaynak (2015) apibrėžia, kad didieji duomenys – tai duomenų rinkiniai, kurių dydis viršija įprastų duomenų bazių programinės įrangos galimybes kaupti, saugoti, valdyti ir analizuoti.
- Sharma (2015) taip pat pabrėžia, jog tai duomenų rinkiniai, kurie yra tokie dideli ir komplikuoti, kad juos tampa sunku apdoroti naudojant tradicines duomenų apdorojimo programas ir įrankius.

Autoriai akcentuoja šių duomenų dydį, įvairialypumą bei tai, jog juos sunku apdoroti naudojant tradicines duomenų analizės priemones. Kita dalis autorių didžiuosius duomenis apibrėžia remdamiesi trimis pagrindinėmis savybėmis, kurios priskiriamos 3V modeliui, tai yra kiekis (angl. *Volume*), įvairovė (angl. *Variety*), greitis (angl. *Velocity*) (Alharthi et al., 2017; Liu, J. Li, W. Li ir Wu, 2016; Russom ir Org, 2011; Tanwar, Duggal ir Khatri, 2015).

- Kiekis apibrėžia sukauptų duomenų dydį. Didieji duomenys gali būti pateikiami terabaitais, petabaitais, zetabaitais ar eksabaitais. Visgi didžiųjų duomenų dydžio apibrėžimai yra reliatyvūs ir gali skirtis priklausomai nuo duomenų tipo ar naudojimo sektoriaus (Gandomi ir Haider, 2015). Dėl šios priežasties tikslaus apibrėžimo, koks duomenų kiekis yra reikalingas, kad duomenys būtų vertinami kaip didieji, nėra. Be to, dėl vis augančio duomenų kiekio ir technologinės pažangos didžiųjų duomenų dydžio suvokimas kinta, todėl tai, kas šiandien yra laikoma didžiais duomenimis, ateityje gali būti nebetraukojama kaip didieji duomenys. Skaičiuojama, kad 2012 m. kiekvieną dieną buvo sukuriama apie 2.5 eksabaito duomenų ir šie skaičiai per 40 mėnesių išaugo dvigubai (Mcafee ir Brynjolfsson, 2012). *International Data Corporation* duomenimis, 2013 m. šis skaičius siekė 4.4 zetabaitus (Rajaraman, 2016). Prognozuojama, kad 2020 m. duomenų kiekis pasieks 40 zetabaitų (Kune, Konugurthi, Agarwal, Chillarige ir Buyya, 2016).
- Įvairovė apibrėžia sukauptų duomenų heterogeniškumą (Tanwar et al., 2015). Šia charakteristika pabrėžiama tai, jog didieji duomenys gali būti generuojami iš daugybės šaltinių ir formatų bei apimti struktūrizuotus ir nestruktūrizuotus duomenis (Russom ir Org, 2011). Net 95 % duomenų yra nestruktūrizuoti – tai duomenys, kuriuos generuoja socialiniai tinklai, sensoriai, išmanieji įrenginiai. Nors įmonės ir anksčiau kaupė didelius kiekius duomenų iš vidinių ir išorinių šaltinių, tačiau didžiųjų duomenų panaudojimo galimybės pasikeitė atsiradus naujoms technologijoms, kurios naudojamos duomenims apdoroti ir analizuoti (Gandomi ir Haider, 2015). Kitaip tariant, dėl technologinės pažangos įmonės turi geresnes sąlygas rinkti didžiuosius duomenis, juos apdoroti ir analizuoti.
- Greitis žymi duomenų generavimo greitį (kaip dažnai programinė įranga sugeneruoja naujus duomenis) ir greitį, kuriuo duomenys turėtų būti analizuojami ir apdorojami – koku greičiu duomenys yra perduodami analitinių duomenų rinkinyje po to, kai duomenys yra sugeneruoti (Gandomi ir Haider, 2015; Russom ir Org, 2011). Augantis internetu sąveikaujančių įrenginių ir socialinių medijų naudojimas lemia duomenų kaupimo spartą bei didina realaus laiko analizės ir įrodymais pagrįsto planavimo poreikį. Duomenys, kuriuos sugeneruoja išmanieji įrenginiai ir mobiliosios programos, didele sparta generuoja informacijos srautus, kuriuos galima panaudoti analizuojant duomenis realiuoju laiku ir taip kurti individualizuotus pasiūlymus esamiems bei potencialiems klientams (Gandomi ir Haider, 2015).



1 paveikslas. Didžiųjų duomenų vertės kūrimo grandinė (angl. *The chain of big data value creation*) (sudaryta autorių, remiantis Cavanillas et al., 2015)

Figure 1. The chain of big data value creation (created by authors, based on Cavanillas et al., 2015)

Naujesniuose šaltiniuose autoriai praplečia didžiųjų duomenų sąvoką remdamiesi 5V modeliu – tai papildytas 3V modelis, kuriame atsiranda du nauji komponentai – teisingumas (angl. *Veracity*) ir vertė (angl. *Value*) (Sharma, 2015; Yin ir Kaynak, 2015):

- Teisingumas. Didelė dalis duomenų, gaunamų iš interneto ar įvairių jutiklių, yra neteisingi (Rajaraman, 2016). Renkant duomenis turėtų būti taikomi reikalavimai, kurie apibrėžtų kaupiamų duomenų tikslumą ir teisingumą. Todėl svarbu, jog prieš pradėdant analizės procesus, klaidingi duomenys būtų pašalinami. Įmonė turėtų apsibrėžti, kokius duomenis ir koku tikslu kaupia, taip sumažintų klaidingų ar nereikalingų duomenų kiekį (Yin & Kaynak, 2015). Duomenų srautai sparčiai auga, o didelius jų kiekius sunkiau apdoroti, tai reikalauja didesnių resursų ir gali privesti prie klaidingų išvadų (kuo didesnis duomenų kiekis, tuo didesnė klaidingos informacijos tikimybė).
- Vertė. Analizuojant surinktus duomenis aptinkama vertingų įžvalgų, kuriomis remdamosi įmonės gali koreguoti strategiją, tobulinti produkciją, pažinti savo klientą, kurti personalizuotas žinutes ir pelnyti konkurencinį pranašumą. Svarbu suprasti tai, jog patys duomenys nėra vertingi, vertė gaunama juos analizuojant (Rajaraman, 2016). Cavanillas ir kt. minimoje didžiųjų duomenų vertės kūrimo grandinėje vaizduojami tokie procesai kaip duomenų gavyba, analizė, kuravimas, saugojimas ir panaudojimas (Cavanillas, Curry ir Wahlster, 2015) (1 paveikslas). Duomenų analitikas vykdydamas įvairias užklausas ir taip filtruodamas duomenis gali aptikti verslui naudingas tendencijas, todėl vertė gali būti apibūdinama kaip svarbiausia didžiųjų duomenų savybė (Hashem et al., 2015; Sivarajah, Kamal, Irani ir Weerakkody, 2017) ir galutinis produktas.

Analizuojant didžiųjų duomenų sąvoką, buvo pasiūlyti ir 6V (Groves, Kayyali, Knott ir Kuiken, 2013) bei 7V (Gandomi ir Haider, 2015) modeliai, kuriuose atsirado pagrįstumo (angl. *Validity*) ir matomumo (angl. *Visibility*) charakteristikos:

- Pagrįstumas – ši charakteristika apibūdina duomenų logiškumą (Khan et al., 2014). Iš dalies ši charakteristika gali kelti asociacijas su teisingumu, tačiau pagrįstumo charakteristikoje svarbiau, ar duomenys turi jiems priskirtą prasmę, ar atitinka faktus. Pagrįstumo poreikį galima sieti su asmens duomenų apsauga, nes svarbu užtikrinti, kiek skaidrus ir tinkamas tokiems duomenims saugoti yra šaltinis bei kokia yra duomenų kilmė.

- Matomumo charakteristika taip pat svarbi siekiant užtikrinti, jog duomenys yra patikimi ir prieinami. Ši charakteristika suteikia galimybę naudotojui matyti ir analizuoti duomenis, turint prie jų prieigą (Gandomi ir Haider, 2015).

Galima pastebėti, jog V modeliai evoliucionuoja – juos papildo nauji elementai, kurie formuojami remiantis didžiųjų duomenų savybėmis. Atsižvelgiant į tai, jog V modeliai sudaromi remiantis didiesiems duomenims būdingomis charakteristikomis, būtų tikslinga juos vertinti kaip didžiųjų duomenų charakteristikų modelius (2 paveikslas).

Analizuojant sudarytą modelį galima matyti, jog didžiųjų duomenų charakteristikų modeliams priskiriami kiekio, įvairovės, greičio, teisingumo, vertės, pagrįstumo ir matomumo elementai. Vizualizacijoje pilka spalva pavaizduoti elementai, būdingi konkrečiam didžiųjų duomenų charakteristikos modeliui. Evoliucionuojant didžiųjų duomenų charakteristikų modeliams senieji elementai išlieka tokie pat svarbūs, tačiau juos papildo nauji elementai, kurie didžiuosius duomenis apibrėžia ne per jų savybių prizmę, tačiau yra labiau orientuoti į duomenų kokybę.

Apibendrinant galima teigti, jog didieji duomenys – tai didelės įvairovės, sparčiai augantys duomenų kiekiai, kurie reikalauja pažangių saugojimo ir apdorojimo technologijų bei kuriuos analizuojant galima gauti vertingų įžvalgų. Vis dėlto tam, kad iš šių duomenų būtų išgaunama vertė, yra svarbu, jog šie duomenys tenkintų teisingumo, pagrįstumo ir matomumo sąlygas.

| Modelis / Elementas | 3V | 5V | 6V | 7V |
|---------------------|----|----|----|----|
| Kiekis | █ | █ | █ | █ |
| Įvairovė | █ | █ | █ | █ |
| Greitis | █ | █ | █ | █ |
| Teisingumas | █ | █ | █ | █ |
| Vertė | █ | █ | █ | █ |
| Pagrįstumas | █ | █ | █ | █ |
| Matomumas | █ | █ | █ | █ |

2 paveikslas. Didžiųjų duomenų charakteristikų modelių elementai (angl. *The elements of big data characteristics models*) (sudaryta autorių, remiantis Alharthi et al., 2017; Gandomi ir Haider, 2015; Khan et al., 2014; Rai, Meshram ir Gunasekaran, 2018; Russom ir Org, 2011; Sharma, 2015; Tanwar et al., 2015; Yin ir Kaynak, 2015)

Figure 2. The elements of big data characteristics models (created by authors, based on Alharthi et al., 2017; Gandomi & Haider, 2015; Khan et al., 2014; Rai, Meshram, & Gunasekaran, 2018; Russom & Org, 2011; Sharma, 2015; Tanwar et al., 2015; Yin & Kaynak, 2015)

2. Didžiųjų duomenų klasifikacija

Siekiant geriau suprasti didžiųjų duomenų charakteristikas ir įvairovę, svarbu apibrėžti galimus didžiųjų duomenų tipus. Didieji duomenys gali būti skirstomi į kategorijas pagal jiems būdingas ypatybes, tokias kaip šaltinis ar duomenų formatas, turinys.

Autoriai (Boyd ir Crawford, 2012; Cuzzocrea, Song ir Davis, 2011; Dong ir Srivastava, 2013; Gandomi ir Haider, 2015; X. Wu, Zhu, G.-Q. Wu ir Ding, 2003) taip pat įvardija galimus duomenų šaltinius:

- *Socialinės medijos*. Tai informacijos šaltinis, suteikiantis galimybę dalintis ar keisti informacija ir idėjomis virtualiose bendruomenėse, tokiose kaip tinklaraščiai, mikrotinklaraščiai ar socialiniai tinklai (Hashem et al., 2015). Socialinės medijos generuoja tokius duomenis kaip ryšiai socialiniuose tinkluose (Banerjee ir Agarwal, 2012; Chen, Peng ir Lee, 2012), nuotraukos, komentarai po nuotraukomis, nuotraukų antraštės (Y. Wang, S. Wang, Tang, Liu ir Li, 2015), vaizdo įrašai, reakcijos į vartotojų įrašus (paspaudimai „patinka“, pasidalinimai kitų vartotojų įrašais).
- *Išmanieji įrenginiai*. Duomenys gaunami iš išmaniųjų įrenginių (telefonų, kompiuterių ir kt. skaitmeninių įrenginių) yra priskiriami mechanizmų generuojamiems duomenims (angl. *Machine generated data*) – tai duomenys, kurie automatiškai generuojami iš įrangos ar programinės įrangos, pavyzdžiui, kompiuterių, medicinos prietaisų, arba kitų įrenginių, be žmogaus įsikišimo (Hashem et al., 2015).
- *Internetu sąveikaujantys įrenginiai*. Daugybė prie interneto prijungtų išmaniųjų įrenginių, tokių kaip išmanieji telefonai, planšetiniai kompiuteriai, skaitmeninės kameros, sąveikaudami tarpusavyje generuoja didžiulius duomenų kiekius (Hashem et al., 2015). Ši technologija apibūdina fizinio pasaulio daiktų ir žmonių santykį, tarpusavyje sujungus technologijas bei įvairias platformas.
- *Jutikliai*. Jutikliais perduodami signalai, kuriuose gali būti užšifruoti fiziniai duomenys (pvz., fotorezistoriai perduoda duomenis, susijusius su saulės intensyvumu).
- *Transakcijos*. Operacijų duomenys (pvz.: finansinių operacijų duomenys, medicininiai įrašai, elektroninės komercijos svetainėse atliekami veiksmai (užsakymai, mokėjimai)).

Dauguma autorių (Gandomi ir Haider, 2015; Hashem et al., 2015; Russom & Org, 2011; Wu et al., 2003), atsižvelgdami į didžiųjų duomenų formatą, išskiria šiuos tipus:

- Struktūrizuotiems duomenims priskiriami duomenys, kurie yra susiję su duomenų bazėmis, turinčiomis struktūrą (Tanwar et al., 2015). Struktūrizuoti duomenys gali būti skaičiai, žodžiai, datos, duomenų lentelės. Tokius duomenis lengva suvesti, saugoti ir analizuoti.
- Nestruktūrizuotiems duomenims priskiriami duomenys, kurie surenkami pasitelkiant šiuolaikines technologijas, tai gali būti nuotraukos, grafiniai piešiniai,

garso failai, komentarai, elektroniniai laiški, blogų įrašai ir kitokio tipo failai. Nestruktūrizuoti duomenys sudaro net 95 % visų duomenų (Gandomi ir Haider, 2015).

- Pusiau struktūrizuotiems duomenims priskiriami duomenys, kurių forma neatitinka griežtų standartų, apimančių visiškai struktūrizuotus ir nestruktūrizuotus duomenis (Tanwar et al., 2015). Pusiau struktūrizuotiems failams priskiriami XML failai.

Taigi, dažniausiai literatūroje didieji duomenys yra skirstomi pagal jų šaltinį arba duomenų struktūrą. Struktūrizuoti ir nestruktūrizuoti duomenys skiriasi apdorojimo metodais. Taip pat svarbu pabrėžti tai, jog nestruktūrizuotus duomenis yra sudėtingiau analizuoti, jie reikalauja pažangesnių technologijų taikymo, tačiau juos apdorojant galima gauti vertingesnių išvalgų, leidžiančių geriau pažinti vartotojo elgeseną ir interesus. Naudojant informacijos gavybos technologijas bei taikant algoritmus iš nestruktūrizuotų duomenų galima išgauti struktūrizuotus duomenis (Gandomi ir Haider, 2015). Kalbant apie socialinių medijų duomenis, juos galima suskirstyti į dar dvi grupes (Gandomi ir Haider, 2015):

- Vartotojo kuriamas turinys (angl. *User-generated content*) – tai duomenys, apimantys vartotojų publikuojamą turinį, pvz.: nuotraukas, vaizdo įrašus, atsiliepimus, komentarus.
- Duomenys, paremti vartotojų ryšiais ir interakcijomis.

Šie aspektai lemia, kokia perspektyva taikoma analizuojant turimus duomenis – vartotojo kuriamo turinio analizėje orientuojamasi į turinio, pvz.: nuotraukų, vaizdo įrašų, atsiliepimų, komentarų analizę, o analizuojant duomenis, kurie paremti vartotojų ryšiais, yra atliekama struktūrinė analizė.

Taigi, klasifikuojant didžiuosius duomenis ir išskiriant jų tipus, galima pastebėti, jog atskiroms duomenų grupėms yra taikomi skirtingi analizės metodai, taip pat planuojant didžiųjų duomenų projektus, galima tiksliau numatyti, kokius duomenis organizacija turėtų pradėti kaupti bei kaip turėtų planuoti tolesnius apdorojimo ir analizės veiksmus.

Iki šiol socialinės medijos buvo vertinamos kaip komunikacijos kanalas, tačiau dabar vis daugiau dėmesio skiriama socialinių medijų kaip duomenų šaltinio analizei bei jų panaudojimo galimybės identifikuoti. Todėl toliau straipsnyje bus analizuojamos būtent iš šio šaltinio gaunamų didžiųjų duomenų panaudojimo galimybės ir galimi analizės metodai.

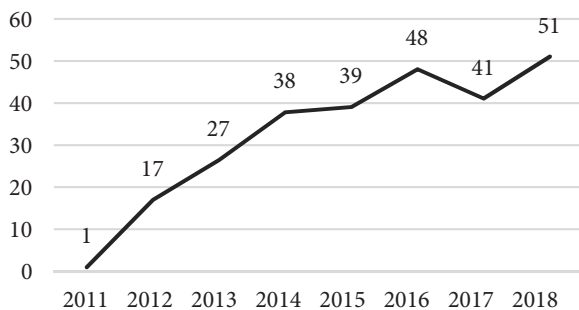
3. Socialinių medijų klasifikavimo koncepcija

Sparčiai augant socialinių medijų populiarumui, auga ir duomenų, kuriuos sugeneruoja socialinės medijos, kiekiai bei panaudojimo galimybės. Nuo 2010 m. iki 2018 m. socialinių medijų vartotojų skaičius išaugo daugiau nei 3 kartus, 2010 m. pasaulyje buvo 0,97 mlrd. socialinių medijų vartotojų (Statista, 2017), 2018 m. šis skaičius viršijo 3,19 mlrd. vartotojų (We Are Social, 2018). Socialinės

medijos ne tik apima informacijos šaltinius, tačiau taip pat savo informacinėse platformose kaupia ir saugo didelius informacijos, susijusias su vartotojų elgsena internete, kiekius (Agichtein, Castillo, Donato, Gionis ir Mishne, 2008). Remiantis komentarais, reakcijomis ir įrašais, naujienomis, kurias vartotojas seka ar kokių vartotojų jis yra sekamas, duomenų analitikai gali daug pasakyti apie vartotojus ir jų elgesį. Naudodami didžiulius duomenis, gaunamus iš socialinių medijų, rinkodaros specialistai gali gauti naudingų įžvalgų ir vykdyti duomenimis paremtos rinkodaros strategijas. Augančią socialinių medijų generuojamų duomenų svarbą taip pat pagrindžia ir didėjantis paieškų šioje srityje skaičius. Remiantis *Google Trends* duomenimis, nuo 2013 m. paieškų, susijusių su socialinių medijų didžiasiais duomenimis, skaičius išaugo beveik dvigubai (3 paveikslas).

Moksliniuose šaltiniuose autoriai socialines medijas išskiria kaip vieną iš pagrindinių didžiųjų duomenų šaltinių (Bello-Organ, Jung ir Camacho, 2016; Kaisler, Armour, Espinosa ir Money, 2014; Olshannikova, Olsson, Huhtamäki ir Kärkkäinen, 2017). Literatūroje galima rasti įvairių socialinių medijų apibrėžimų, pavyzdžiui:

- Gruebner ir kt. (2017) socialines medijas įvardija kaip internetines programas, kuriomis naudodamiesi vartotojai gali dalintis ir naudoti turinį bei bendrauti tarpusavyje įvairiais būdais.
- Constantinides (2014) rašo, kad socialines medijas apibrėžia kaip Web 2.0 programas, leidžiančias kurti, redaguoti ir platinti vartotojų sukurtą turinį.
- Henderson (2010) socialines medijas apibrėžia kaip programų rinkinį, orientuotą į vartotojų dalyvavimą, informacijos dalybas bei bendradarbiavimą.
- Kaplan ir Haenlein (2010) teigimu, socialinės medijos – tai internetu paremtų programų grupė, kuri remiasi Web 2.0 ideologiniais ir technologiniais pagrindais, leidžiančiais kurti ir keistis vartotojų sukurtu turiniu.
- Wildman ir Obar (2015), remdamiesi įvairių autorių apibrėžimais, susintetino informaciją ir išskyrė keturis pagrindinius socialinėms medijoms priskiriamus bruožus, kurie yra svarbūs norint suvokti socialinių medijų veikimo principus: 1) tai internetu paremtos



3 paveikslas. „Big data in Social Media“ paieškos tendencijos 2011–2018 m. (angl. *Trends of „Big data in Social Media“ searches in 2011–2018 m.*) (sudaryta autorių, remiantis *Google Trends* duomenimis)

Figure 3. Trends of „Big data in Social Media“ searches in 2011–2018 m. (created by authors, based on *Google Trends* data)

programos, priskiriamos Web 2.0 saitynui, 2) šių programų varomoji jėga yra vartotojų kuriamas turinys, 3) veikimo principas yra paremtas individualiais vartotojų arba grupių profiliais, 4) šiems profiliams jungiantis į tinklą susiformuoja socialiniai tinklai. Galima matyti, jog anksčiau minėtų autorių apibrėžimai taip pat yra paremti šiais bruožais.

Plačiau gilinantis į socialinių medijų sampratą, tikslinga išskirti socialinių medijų rūšis. Literatūroje išskiriamos pagrindinės socialinių medijų rūšys: socialiniai tinklai, interneto tinklaraščiai, mikrotinklaraščiai, turinio bendruomenės ir kt. (Bello-Organ, Jung ir Camacho, 2016; Constantinides ir Fountain, 2008; Kaplan ir Haenlein, 2011). Siekiant gilesnio socialinių medijų suvokimo, tikslinga atlikti socialinių medijų klasifikavimo paradigmų analizę. Vienu populiariausių socialinės medijos klasifikatorių gali būti laikomas autorių Constantinides ir Fountain (2008), remiantis socialinių medijų naudojimo tipu, siūlomas klasifikatorius. Išskiriamos penkios pagrindinės socialinių medijų kategorijos: tinklaraščiai, socialiniai tinklai, turinio bendruomenės, forumai ir turinio agregatoriai (angl. *Content aggregators*). Kaplan ir Haenlein (2010) socialines medijas klasifikuoja remdamiesi informacijos, kurią galima perduoti per tam tikrą laiką, kiekiu ir savęs atskleidimo galimybėmis – jų klasifikacija anksčiau minėtas socialinių medijų formas papildė virtualių socialinių pasaulių ir virtualių žaidimų pasaulių socialinių medijų tipais. Vis dėlto, atsižvelgiant į šio darbo tikslą bei orientuojantis į socialinių medijų duomenų panaudojimą rinkodaros ir komunikacijos tikslais, pagrindinėmis socialinių medijų rūšimis išlieka: socialiniai tinklai, naujienų programėlės, turinio dalinimosi programos, tinklaraščiai ir mikrotinklaraščiai, forumai ir turinio agregavimo sistemos.

Skirtingi socialinės medijos tipai skiriasi savo paskirtimi, funkcionalumu ir generuojamais duomenimis:

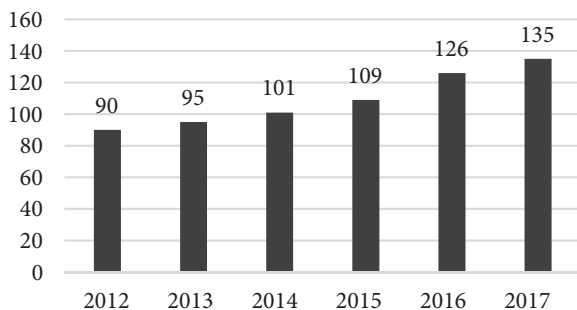
- *Interneto tinklaraščiai* – tai socialinių medijų forma, primenanti dienoraštį (Thevenot, 2007). Tinklaraščiai apibrėžiami kaip paprasto naudojimo asmeninės internetinės svetainės, kuriose vartotojai laisvai dalinasi savo patirtimi, istorijomis, nuomone, kelia turinį (tekstus, nuotraukas), komentuoja kitų vartotojų įrašus. Atsiradus *Twitter* programai, tinklaraščių mediją praplėtė mikrotinklaraščių formatas.
- *Mikrotinklaraščiai* leidžia publikuoti nedidelės apimties tekstinius pranešimus, matomus visiems skaitytojams arba vartotojo pasirinktai jų grupei (Kaplan ir Haenlein, 2010). *Twitter* duomenų analizei gali būti taikoma teksto analizė (He, Zha ir Li, 2013).
- *Socialinius tinklus* galima apibrėžti kaip rinkinį paslaugų, suteikiančių galimybes individualiems vartotojams kurti asmeninius profilius bei juose pateikti įvairių informaciją apie save, publikuoti nuotraukas, rašyti viešus pranešimus, taip pat ieškoti kitų vartotojų bei užmegzti naujus ryšius, keistis kontaktais (Čičevič, Samčović ir Nešić, 2016). Socialinių tinklų populiarumas tarp vartotojų sparčiai auga, remiantis Statista (2017) duomenimis, 2012 m. vidutinis

socialiniuose tinkluose praleidžiamas laikas per dieną buvo 90 minučių, 2017 m. šis skaičius siekė 135 minutes (4 paveikslas).

- *Turinio bendruomenės* – tai socialinių medijų erdvės, kuriose vartotojai dalijasi tam tikros rūšies informacija. Populiariausios turinio bendruomenės yra nuotraukų (*Flickr, Picasa*) ir vaizdo įrašų (*YouTube, Vimeo*).
- *Forumai* – tai virtualių diskusijų erdvės, dažnai susijusios su specifinėmis temomis ar pomėgiais (Constantinides ir Fountain, 2008). Forumuose sukuriamas vartotojų turinys yra laisvai randamas, naudojant paieškos sistemas (pvz., *Google*), bei pasiekiamas ir matomas visiems interneto vartotojams.
- *Turinio agregavimo sistemos* – pirmoji turinio agregavimo sistemų kategorija apima programas, leidžiančias vartotojams lengvai pasiekti visiškai pritaiktą, sindikuotą interneto turinį, naudojant socialinio žymėjimo arba RSS priemones, antroje kategorijoje turinio agregatoriai apima programas, pagrįstas turiniu, surinktu iš įvairių šaltinių, kuriant naujus, dažnai pritaikomus produktus ar paslaugas (pvz., *Google Maps*) (Constantinides, 2014).

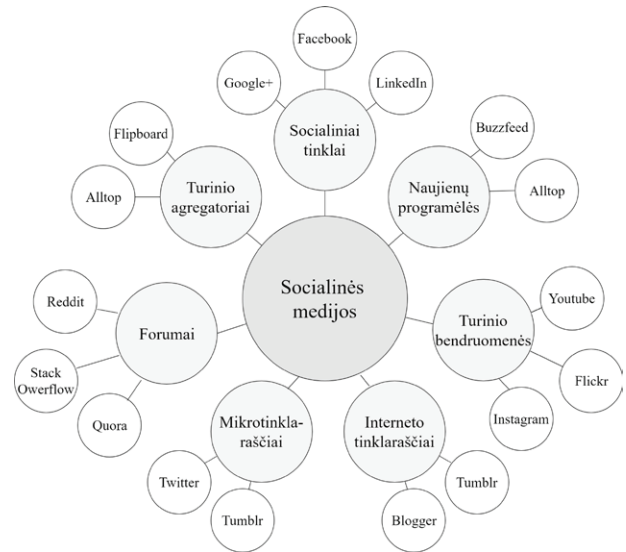
Smulkiau gilinantis į socialinių medijų tipus galima pastebėti, jog šie socialinių medijų tipai skirstomi į smulkesnius vienetus – kanalus, arba kitaip – platformas. Pavyzdžiui, remiantis *SimilarWeb* platformos duomenimis, galima išskirti šiuos populiariausius kanalus: tinklaraščiams yra priskiriamos tokios platformos *Tumblr, Blogger*, mikrotinklaraščiams – *Twitter, Tumblr* (platformoje yra galimybė tiek pateikti trumpus pranešimus, tiek rašyti ilgus tekstus), socialiniams tinklams – *Facebook, LinkedIn, Google+*, turinio bendruomenės – *Youtube, Flickr, Instagram*, forumai – *Quora, Stack Overflow*, turinio agregavimo programos – *Flipboard, Alltop* ir kt. Svarbu pabrėžti, kad tai tik maža dalis socialinėms medijoms priskiriamų kanalų.

Zhu ir Chen (2015) socialines medijas klasifikuoja, remdamiesi ryšio pobūdžiu (tai gali būti programos, paremtos vartotojo profiliu arba paremtos turiniu) ir pranešimų pritaikymo lygiu (koku mastu paslauga yra pri-



4 paveikslas. Vidutinis vartotojo laikas (min), praleidžiamas socialiniuose tinkluose 2012–2017 m. (angl. Daily time spend on social networks per day worldwide 2012–2017 m) (sudaryta autorių, remiantis Statista (2017))

Figure 4. Daily time spend on social networks per day worldwide 2012–2017 m. (created by authors, based on Statista, 2017)



5 paveikslas. Socialinių medijų klasifikacija (angl. *Social media classification*) (sudaryta autorių, remiantis Constantinides ir Fountain, 2008; Kaplan ir Haenlein, 2010)

Figure 5. Social media classification (created by authors, based on Constantinides & Fountain, 2008; Kaplan & Haenlein, 2010)

taikyta tam, kad patenkintų tam tikrus individo pageidavimus). Ryšiu paremtoms socialinėms medijoms gali būti priskiriami socialiniai tinklai (pvz.: *Facebook, Twitter, Whatsapp*), turiniu paremtoms socialinėms medijoms priskiriamos tokios turinio bendruomenės kaip *Instagram, Flickr, Pinterest, Youtube*. Remiantis autorių pateikiama socialinių medijų rūšimis bei joms priskiriamais kanalais, sudaryta socialinių medijų klasifikacijos, paremtos socialinių medijų tipais ir kanalais, schema (5 paveikslas).

Apibendrinant galima teigti, jog socialinės medijos – tai Web 2.0 saitynu paremtos programos, suteikiančios galimybę vartotojams kurti profilius, jungtis į bendruomenes bei keistis informacija. Galima pastebėti, jog socialinės medijos pasižymi skirtinga vartotojų įtraukimo forma, kanalo specifika bei generuojamų duomenų tipu, apimančiu tiek tekstinius duomenis (pvz., žinutes, komentarus, atsiliepimus), tiek sudėtingus, kompleksinius duomenis (pvz., vaizdo įrašus, nuotraukas), interakcijas ir operacijas. Sparčiai augant socialinių medijų populiarumui ir vartotojų įsitraukimui, duomenų kiekiai taip pat auga. Susidaro palankios sąlygos socialinių medijų duomenų analizei pritaikyti verslo aplinkoje.

4. Didžiųjų duomenų analizės metodai

Svarbu pabrėžti tai, jog iš duomenų vertė gaunama tada, kai yra atliekama sukaupytų duomenų analizė. Kalbėdamas apie socialinių medijų duomenų analizę, Aggarwal (2011) įvardijo dvi galimas socialinių medijų duomenų analizės kryptis – ryšių bei struktūros analizę ir turiniu paremtą analizę. Ryšiais ir struktūra paremta analizė suteikia galimybę analizuoti tinklo mazgus, ryšius ir bendruomenes bei nustatyti besivystančius regionus – tokia analizė suteikia gerą pagrindinio tinklo elgesio apžvalgą, gali būti taikoma vartotojui pažinti ar tam tikrų grupių

elgsenai pažinti, prognozuoti. Taip pat, kadangi socialinės medijos pasižymi įvairaus turinio gausa, yra palanki terpė pasitelkti tekstinių ir vaizdinių duomenų analizę, kurią taikant kartu su ryšių ir struktūros analize galima gauti vertingų išvalgų.

Kadangi, kaip jau buvo apžvelgta, didieji duomenys pasižymi sudėtinga struktūra ir tradicinių analizės metodų jiems nepakanka, svarbu apžvelgti galimus mašininio mokymosi analizės metodus, kurie gali būti taikomi analizuojant didžiuosius duomenis. Remiantis atlikta mokslinės ir metodinės literatūros analize, galima teigti, kad bendri mašininio mokymosi duomenų analizės metodai apima klasifikaciją, klasterizaciją ir regresiją (Batrinca ir Treleaven, 2015). Šie metodai skirstomi į dvi mašininio mokymosi (angl. *Machine learning*) metodų grupes – prižiūrimus (angl. *Supervised*) ir neprižiūrimus (angl. *Unsupervised*) metodus. Prižiūrimiems mašininio mokymosi metodams priskiriami klasifikacijos ir regresijos metodai, neprižiūrimiems – klasterizacijos metodas.

Dažniausiai duomenų tyryboje pasitelkiamas klasifikacijos metodas, kurį taikant iš anksto suklasifikuoti pavyzdžiai naudojami klasifikacijos modeliui ar algoritmui parengti (Ahmed ir Elaraby, 2014). Šio metodo pagrindinis uždavinys yra pagal pateiktus objekto duomenis sukurti duomenų žymėjimo algoritmą, kuriuo remiantis duomenys bus skirstomi į klases. Duomenų klasifikavimas yra dviejų žingsnių metodas (Agarwal, 2013):

1. Mokymasis – turimų duomenų klasifikavimas pagal klasifikavimo algoritmą. Šiame etape nustatomas klasifikavimo modelis ar algoritmas, kuris analizuoja didelius duomenų kiekius, remdamasis pateiktomis duomenų bazėmis ir jų suderinamumo klasėmis. Šis etapas laikomas mokymosi etapu, nes modelis yra sukuriamas remiantis duomenimis, iš kurių algoritmas mokosi.
2. Metodo taikymas – šiame etape tikrinamas modelio tikslumas ir nuoseklumas. Analizuojamos testinių duomenų grupės ir taip tikrinamas algoritmo veikimo tikslumas – jeigu jis yra tikslus ir klasifikaciją vykdo teisingai, pradedama nežinomų ir žymų neturinčių duomenų analizė.

Populiariausias ir paprasčiausias klasifikacijos metodas yra sprendimų medžio metodas, tačiau taip pat galima išskirti palaikymo vektoriaus, neuroninio tinklo, Naive Bayes klasifikatoriaus (vykdomas remiantis Bajeso teorema), asociacijomis grįstos klasifikacijos ir artimiausių kaimynų metodus. Kalbant apie socialinių medijų duomenų tyrybą, klasterizacija gali būti taikoma duomenų *valymo* etape. Kadangi duomenys renkami iš daugelio vartotojų ir šaltinių, nemaža dalis gali būti klaidingi arba pertekliniai (t. y. turėti pašalinių duomenų, kurie yra nereikalingi ar klaidinantys), todėl prieš pradėdant duomenų analizę yra būtina juos apdoroti ir paruošti. Tai gali būti atliekama taikant taisyklėmis pagrįstą teksto klasifikavimą arba mokymusi paremtą klasifikatorių, kuris veiktų remdamasis pažymėtų duomenų analize (Weiguo ir Gordon, 2014).

Dar vienas mašininio mokymosi metodas, taikomas duomenų tyryboje, yra klasterizacija. Taikant šį metodą tarp objektų yra ieškoma panašumų, o juos radus – panašūs ir skirtingi objektai skirstomi į atskiras grupes (Liu, Li, Xiong, Gao ir Wu, 2010). Tai savarankiškai vykdoma, neprižiūrima duomenų tyrybos technika, kuri iš pirmo žvilgsnio primena klasifikaciją. Nors šie metodai atrodo panašūs, tačiau iš tiesų jie veikia skirtingais principais. Taikant klasifikacijos metodą duomenys turi priskirtas žymas, pagal kurias jiems priskiriamos klasės. Taikant klasterizacijos metodą, duomenys yra grupuojami pagal panašumus ir skirtumus duomenų grupės viduje, o ne remiantis bendrais grupei būdingais bruožais (žymomis) (Agarwal, 2013).

Pokyčių ir tendencijų prognozėms yra taikomas mašininio mokymosi regresijos metodas. Regresija taip pat gali būti taikoma modeliuojant santykius tarp vieno ar kelių kintamųjų bei aptinkant pokyčių tendencijas (Politaitė ir Sabaitytė, 2018). Išskiriami šie regresijos tipai: linijinė regresija, nelinijinė regresija, daugiamatė tiesinė regresija, daugiamatė netiesinė regresija. Tendencijų analizės taikymas apima klientų ar pardavimų augimo prognozavimą, numatant reklamos kampanijų efektyvumą, likusius vartotojų elgsenos pokyčius ir t. t. (Weiguo ir Gordon, 2014).

Apibendrinant galima teigti, jog visos duomenų tyrybos technikos yra skirtingos – vienos iš jų reikalauja žmogaus įsikišimo, kitos gali veikti savarankiškai. Technikos skiriasi ir tuo, kokį rezultatą jos gali grąžinti atlikus duomenų analizę. Duomenų tyrybos technikos turėtų būti pasirenkamos atsižvelgiant į tai, kokio rezultato tikimasi bei ko norima pasiekti – ar įmonė siekia nustatyti bendrus, tarp vartotojų pasikartojančius interesus bei skirtumus, ar norima nustatyti pokyčius. Įmonė turėtų pasirinkti tokį metodą, kuris būtų tinkamas jų duomenų analizei bei padėtų įgyvendinti išsikeltus tikslus.

5. Socialinių medijų generuojamų didžiųjų duomenų panaudojimo galimybės

Remiantis atlikta literatūros analize, išskiriamos socialinių medijų generuojamų didžiųjų duomenų panaudojimo galimybės:

Auditorijos skirstymo galimybės. Siekiant vykdyti efektyvią rinkodarą yra svarbu nustatyti auditoriją, kurią norima pasiekti ir paveikti. Socialinių tinklų duomenys leidžia nustatyti ryšius tarp vartotojų bei nustatyti skirtingas klientų, kurie gali dalytis panašiais interesais, bendruomenes. Remiantis tokiais duomenimis įmonės gali priimti sprendimus, kaip koreguoti savo strategiją, koreguoti rinkodaros strategijas bei analizuojant vartotojų elgseną prognozuoti, kaip tai gali paveikti įmonės veiklą ateityje. Taip pat tiriant ryšius tarp vartotojų, atsiranda galimybė nustatyti nuomonės formuotojus (angl. *Influencers*), kurie daro įtaką bendruomenei, bei naudoti juos taikant komunikacijos strategiją. Vartotojo elgsenos internete analizė, internetinės reklamos auditorijos segmentavimas – didieji duomenys suteikia galimybę stebėti vartotojo elgseną – ką

jis žiūrėjo, kokiuose puslapiuose lankėsi, kokie jo interesai, bei priskiriant jam žymą susikurti panašaus profilio vartotojų interesų grupes. Tai reklamos pardavėjams leido kurti ir pasiūlyti iš anksto sukurtus interesų segmentus reklamuotojams bei taip rodyti reklamą tik tam tikrus interesus turinčioms vartotojų grupėms (Couldry ir Turow, 2014). Taip atsiranda galimybė pasiekti vartotojus, remiantis ne tik jų demografinė informacija (šalimi, miestu, amžiumi, lytimi), tačiau taip pat, remiantis jo elgsena ar lankytinomis vietomis, priskirti jam tam tikrus bruožus, pvz., tokius kaip socialinė ir ekonominė klasė, šeimyninis statusas, interesai. Internetinių filmų kompanija *Netflix*, analizuodama vartotojo elgseną ir profilį, pateikia jo interesus atspindinčias reklamines žinutes, o tai gerina vartotojų patirtį bei didina įsitraukimą (Adweek, 2018).

Mikroorientacija. Tai auditorijos pasirinkimo strategija, taikoma labai konkrečioms tikslams skirtiems skelbimams kurti. Ši strategija gali būti taikoma rinkimų, politinėms kampanijoms (Boerman, 2016). Apie vartotojus kaip rinkėjus yra renkami dideli informacijos kiekiai, kuriuos analizuojant nustatomi rinkėjų įpročiai – kokias partijas jie palaiko, kaip dažnai balsuoja ir t. t. Remiantis surinktais duomenimis galima prognozuoti galimus rinkimų rezultatus, taip pat nustatyti geografines zonas, kuriose kandidatai turėtų organizuoti susitikimus su potencialiais rinkėjais ar suaktyvinti reklamines kampanijas, personalizuoti pranešimus, orientuodamiesi į problemas, kurios aktualios potencialių rinkėjų grupėms (Canick, 2016; MIT Technology Review, 2012). Mikroorientacija taip pat gali būti taikoma ir verslo aplinkoje, siekiant pasiekti konkrečią, siaurą, nišinę auditoriją. Organizacijos, kurdamos reklamines kampanijas *Facebook* platformoje, gali skirstyti auditoriją į labai siauras nišas ir taip susiaurinti savo auditoriją iki keleto ar vieno žmogaus (Barbu, 2014).

Vartotojų profiliavimas. Socialiniai tinklai taip pat suteikia galimybę nustatyti kliento profilį. Klientų profilis yra kliento modelis, pagal kurį įmonės formuoja strategiją, kuri patenkintų tokio vartotojo poreikius. Remiantis surinktais duomenimis, sudaromi vartotojų profiliai, pagal kuriuos identifikuojamos vartotojų savybės ir elgsenos bruožai. Vartotojų profiliai naudojami taikant pardavimo ir rinkodaros strategijas (Khade, 2016). Socialinių tinklų generuojami duomenys gali būti naudojami sudarant vartotojų profilius, nes suteikia galimybę sekti vartotojo viešai reiškiamą nuomonę bei elgseną internetinėje erdvėje (Pencacchiotti ir Popescu, 2011). Remdamasi vartotojų profiliumi, įmonė gali personalizuoti reklaminius pranešimus potencialiems klientams (Anshari et al., 2018). Vartotojų profiliai taip pat gali būti naudojami reklaminiams žinutėms pritaikyti (angl. *Customization*). Vykdamas vartotojų segmentavimą arba vartotojų profiliavimą yra taikomi klasterizacijos ir klasifikacijos metodai. Elgsenos klasifikavimas ir sprendimų medis suteikia galimybę nustatyti, kokius produktus galima vartotojui pateikti kaip rekomendacijas, kokios reakcijos galima tikėtis (Khade, 2016). Remiantis vartotojų elgsena ir interesais galima sudaryti interesų grupes ir taip analizuojant vartotojų elgseną pateikti pasiūlymus ir žinutes kitiems, panašios elgsenos

vartotojams. Taip pat remiantis vartotojo profiliumi pateikti jam produktų ar paslaugų rekomendacijas, kaip tai daro elektroninės komercijos bendrovė *Asos* ar internetinės televizijos bendrovė *Netflix*.

Reputacijos valdymas. Analizuodamos tekstinius duomenis programos realiuoju laiku gali grupuoti duomenis bei taip suteikti galimybę įmonei stebėti reputacijos rodiklius ir tendencijas realiuoju laiku bei priimti su jais susijusius sprendimus. Vienas iš tokios programinės įrangos pavyzdžių yra *Brand Fibres*, realiuoju laiku analizuojanti tekstinius socialinių medijų duomenis. Programa, aptikusi pranešimus, susijusius su prekės ženklu, geba priskirti juos atitinkamoms kategorijoms remdamasi duomenų šaltiniu (pvz., nurodyti socialinį tinklą, kuriame buvo paliktas komentaras, bei pateikti įrašo nuorodą) bei vizualiai atvaizduoti prekės ženklo reputacijos pokyčius bei rodiklius. Klasterizacijos ir asociacijų analizės metodai yra vieni dažniausiai taikomų reputacijos valdymo programinės įrangos palaikymo metodų (Politaitė ir Sabaitytė, 2018).

Taigi, didieji duomenys suteikia galimybę pažinti klientą realiuoju laiku bei taip priimti geresnius sprendimus, tobulinti savo veiklą ir efektyviau vykdyti komunikaciją. Galimybė parinkti tinkamą žinutę tinkamam vartotojui ir tinkamu metu suteikia didžiulę vertę įmonei ir konkurencinį pranašumą. Žinučių pritaikymas yra svarbus todėl, nes padeda užmegzti glaudesnę ryšį tarp organizacijos ir vartotojo, o tai didina vartotojo lojalumą, prekės ženklo atpažįstamumą bei padeda auginti pajamas ir verslo galimybes (Anshari et al., 2018). Taip pat įmonės, kurios pažįsta savo vartotoją, gali efektyviau išnaudoti reklaminių biudžetą, nes duomenų analizės rezultatai leidžia įmonei susitelkti į potencialių klientų segmentą, o ne atsitiktines auditorijas ir taip sumažinti reklamos išlaidas bei vykdyti efektyvesnę komunikaciją. Be to, tai padeda tobulinti produktą ar paslaugą atsižvelgiant į vartotojų patirtį ir taip užmegzti geresnius ryšius su vartotoju.

6. Didžiųjų duomenų panaudojimo rizikos

Didieji duomenys yra įvardijami kaip veiksnys, galintis suteikti įmonei konkurencinį pranašumą bei pridėtinę vertę. Nepaisant didiesiems duomenims priskiriamų privalumų, 2015 m. Gartner teigė, jog didžiųjų duomenų projektų žlugimas siekia 60 % (2015). Prieš pradėdant didžiųjų duomenų projektus svarbu suprasti, kokie veiksniai lemia projekto žlugimą bei kokios yra potencialios rizikos ir sunkumai (angl. *Challenges*).

Didiesiems duomenims būdingi tokie požymiai kaip dydis, sparta ir duomenų įvairovė – tai daro didžiuosius duomenis unikalius ir vertingus, tačiau šios ypatybės taip pat yra ir problemų, susijusių su duomenų kaupimu, saugojimu, analizavimu ir vizualizavimu, priežastis (Kaisler et al., 2014; Khade, 2016). Wamba, Akter, Edwards, Chopin ir Gnanzou (2015) identifikuoja penkis galimus didžiųjų duomenų projektų sunkumus: duomenų apsaugos teisiniai aspektai, technologiniai ir techniniai sprendimai, organizacijos kultūra (žmogiškieji išteklių), prieiga prie duomenų ir sunkumai, susiję su pramonės šakos struktūra.

Sparčiai augantys duomenys reikalauja technologinių ir techninių resursų kaupti, tvarkyti bei galėti efektyviai prieiti prie didelio duomenų kiekio. Nors dabartinės technologijos jau gali saugoti zetabaitus duomenų – tokio kiekio negana ir yra reikalinga saugyklos architektūra, kurioje galėtų tilpti didesnis duomenų kiekis, todėl vienas iš išteklių, kurių gali trūkti įmonei ar organizacijai, yra technologiniai ištekliai (Sharma, 2015). Įmonė, prieš pradėdama projektus, susijusius su didžiųjų duomenų rinkimu ir saugojimu, privalo numatyti, kur bus kaupiami duomenys – dėl didelės duomenų augimo spartos svarbu, kad būtų galimybės didinti duomenų bazės talpą bei turėti finansinių išteklių, kurie leistų tai padaryti, nes pradėtas projektas gali žlugti dėl susidariusių per didelių išlaidų ir finansinių išteklių trūkumo. Didelių finansinių išteklių reikalauja ne tik duomenų saugyklos, tačiau taip pat ir programinė įranga, kuri naudojama duomenims analizuoti.

Didelę svarbą turi ne tik techniniai ir finansiniai, bet iš žmogiškieji ištekliai. Svarbu užtikrinti, jog darbuotojai, kurie dirba su dideliais duomenų kiekiais, yra pakankamai kompetentingi, nes klaidingos išvalgos gali lemti viso projekto žlugimą. Taigi, galima išskirti dar vieną riziką, kuri susijusi su žmogiškąja veikla, – tai darbuotojų kompetencija. Pastaraisiais metais didelis dėmesys skiriamas asmens duomenų ir privatumo apsaugai. Įmonės, kurios pradeda projektus su didžiaisiais duomenimis, turėtų numatyti, kur ir kokį laiko tarpą duomenys bus saugomi, kaip bus užtikrinamas duomenų saugumas, kad duomenys nebūtų nutekinti bei būtų užtikrinamas asmens privatumas ir saugumas, o asmeniniai duomenys nepasiektų trečiųjų šalių (Sharma, 2015).

Sivarajah et al. (2017) išskiria tris sunkumų, susijusių su didžiaisiais duomenimis, kategorijas, remdamiesi duomenų gyvavimo ciklu: duomenimis, procesu ir valdymu. Remiantis įvardijamomis kategorijomis, sunkumai gali būti susiję su duomenimis ir jų charakteristikomis (duomenų dydžiu, įvairove, tikslumu, aptikimu, atvaizdavimu, loginiu pagrindu, faktų atitikimu), proceso eiga ir technikomis (kaip duomenys bus surenkami, kur saugojami, klausimai, susiję su integracija, analizės metodo parinkimu bei rezultatų pateikimu). Tai sunkumai, su kuriais susiduriama viso proceso metu (saugumo, privatumo, teisiniai ir etniniai aspektai, valdymas, išteklių paskirstymas).

Didžiųjų duomenų kaupimas, apdorojimas ir analizavimas yra sudėtingas procesas, reikalaujantis daugybės išteklių (finansinių, techninių, žmogiškųjų). Prieš pradėdama projektus, susijusius su didžiųjų duomenų kaupimu ir analizavimu, įmonė turėtų atsižvelgti į savo galimybes bei kritiškai pažvelgti, ar jai tikrai reikia vykdyti procesus, susijusius su didžiaisiais duomenimis.

7. Konceptinis didžiųjų duomenų panaudojimo socialinėje medijoje modelis

Remiantis mokslinės literatūros analize, sudarytas teorinis socialinių medijų generuojamų didžiųjų duomenų panaudojimo modelis (6 paveikslas). Modelis sudarytas,

remiantis ciklinio modelio principu, nes baigus vykdyti vieną projektą, galima vykdyti pakartotinus projektus, orientuotus į kitas socialinių medijų duomenų panaudojimo galimybes. Modelyje naudojamos vientisos linijos žymi tiesioginį ryšį tarp modelio elementų bei veiksmų seką. Punktyrinės linijos žymi galimą, tačiau neprivalomą veiksmų seką.

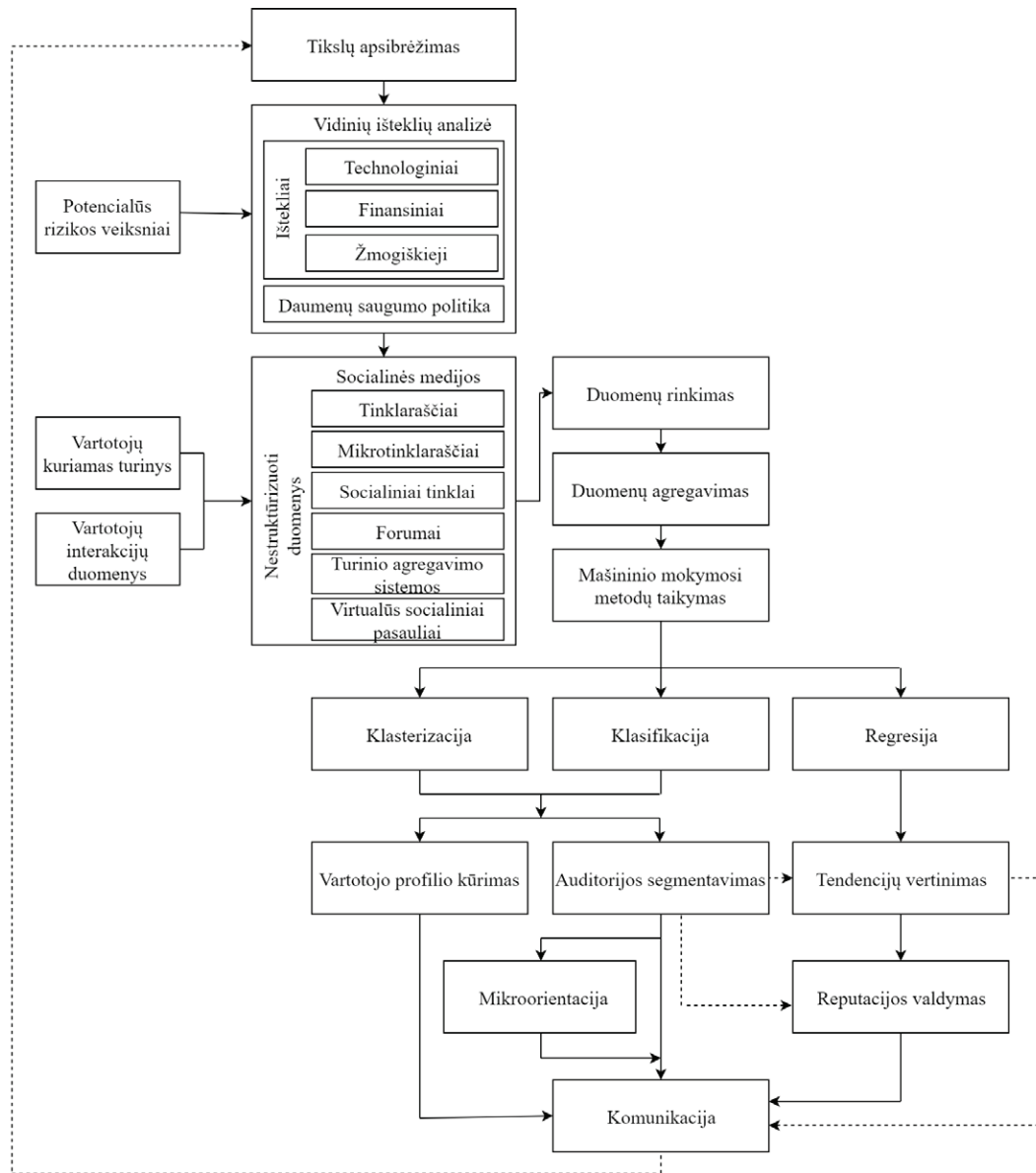
Pasirengimas vystyti didžiųjų duomenų projektą reikalauja kompleksinio planavimo, apimančio tokius procesus kaip duomenų rinkimas, agregavimas, analizės metodai bei tolesnės komunikacijos vykdymas. Prieš pradėdama projektą organizacija turėtų apsibrėžti projekto tikslą bei išsigryninti, ko bus siekiama atliekant duomenų analizę. Šis žingsnis svarbus todėl, nes taip susiaurinamas duomenų laukas, eliminuojant duomenis, kurie nėra reikalingi, ir taip sumažinama klaidingų išvalgų galimybė.

Antrasis žingsnis yra vidinių išteklių analizė. Atlikus literatūros analizę, buvo pastebėta, kad technologinių, finansinių, žmogiškųjų išteklių trūkumas gali nulemti projekto žlugimą ir yra potencialus rizikos veiksnys, todėl siekiant sklandžiai vykdyti projekto veiklą bei apsaugoti nuo galimų nesėkmių, turėtų būti užtikrinta, jog organizacija yra apsirūpinusi reikiamais ištekliais. Dar vienas svarbus veiksnys, kuris turėtų būti įvertintas projekto pradžioje, yra duomenų saugumo užtikrinimas bei organizacijos politika šiuo klausimu.

Socialinių medijų duomenys yra sudaryti iš vartotojų kuriamo turinio bei vartotojų interakcijų – šie duomenys gali būti gaunami iš tokių socialinių medijų platformų ir programinės įrangos kaip tinklaraščiai, mikrotinklaraščiai, socialiniai tinklai, forumai, turinio agregavimo sistemos ir virtualūs socialiniai pasauliai. Šiuose šaltiniuose gali būti aptinkami tokie nestruktūrizuoti duomenys kaip nuotraukos, vaizdo įrašai, atsiliepimai, komentarai, įvairios operacijos ir kt.

Duomenų analizės procesą galima suskirstyti į tris etapus: duomenų rinkimas, agregavimas ir analizavimas. Kadangi didieji duomenys pasižymi dideliu informacijos kiekiu ir įvairove, tokių duomenų analizei yra reikalingi pažangūs analizės metodai, tokie kaip mašininio mokymosi metodai, apimantys klasifikaciją, klasterizaciją ir regresiją. Taikant klasterizacijos ir klasifikacijos metodus yra sudaromas vartotojo profilis. Taip pat šie metodai gali būti taikomi atliekant auditorijos segmentaciją. Išskyrus tam tikrus auditorijos segmentus, apimančius tiek plačias, tiek nišines auditorijas, organizacija gali vykdyti į auditoriją orientuotą komunikaciją. Tiek analizuojant projekto pradžioje surinktus duomenis, tiek vykdant projektą atlikus auditorijos segmentaciją, galima aptikti vertingų išvalgų ir tendencijų, galinčių nulemti tolesnę komunikacijos strategiją ar reikalaujančių su organizacijos reputacijos valdymu susijusių veiksmų. Taip pat tai gali būti tinkamas įrankis vertinant organizacijos poziciją rinkoje.

Paskutinis šio modelio etapas yra komunikacija, kuri gali būti vykdoma taikant suasmenintus pranešimus ar individualizuotus pasiūlymus. Komunikacijos procesuose vartotojų reakcijos į pranešimus gali grįžti duomenų,



6 paveikslas. Didžiųjų socialinės medijos generuojamų duomenų panaudojimo koncepcinis modelis
(angl. *Conceptual framework of big data generated by social media use process*)
Figure 6. Conceptual framework of big data generated by social media use process

tokių kaip komentarai ar atsiliepimai, pavidalu, lemiančiu tai, jos procesas gali tapti cikliškas ir būti kartojamas nuo pradžių.

Sudarytas modelis yra teorinio pobūdžio, todėl reikalauja empirinio patvirtinimo. Modelis iliustruoja organizacijos kelią nuo tikslo apibrėžimo iki socialinių medijų duomenų analize paremtos komunikacijos vykdymo. Kadangi literatūroje trūksta modelių, kokius analizės metodus ir kaip įmonės turėtų pasirinkti, norėdamos vykdyti socialinių medijų analizę, šis modelis galėtų būti taikomas sudarant veiksmų planą prieš pradėdant didžiųjų duomenų analize paremtos komunikacijos projektą. Taikant tolesnius tyrimus ir literatūros analizę, modelis gali būti papildytas technologinėmis analitikos bei duomenų gavybos priemonėmis taip sukuriant dar didesnę vertę.

Išvados

Atlikus mokslinės literatūros analizę nustatyta, jog vieno tikslaus didžiųjų duomenų apibrėžimo nėra. Vis dėlto dauguma autorių didžiuosius duomenims apibrėžia pagal jiems priskiriamas charakteristikas, tokias kaip dydis, sparta, įvairovė ir kt. Apibendrinant didžiuosius duomenis galima apibūdinti kaip didelės apimties ir spartos informacijos kiekius, sugeneruojamus žmonių arba įrenginių ir reikalaujančius inovatyvių bei tobulėjančių technologijų, galinčių juos rinkti, kaupti ir apdoroti. Didieji duomenys pasižymi įvairove, todėl juos galima klasifikuoti remiantis skirtingomis ypatybėmis, tokiomis kaip šaltinis, turinys, rinkimo, saugojimo ir apdorojimo būdai. Dažniausiai literatūroje išskiriamos didžiųjų duomenų kategorijos yra struktūrizuoti,

nestruktūrizuoti ir pusiau struktūrizuoti duomenys – tokia klasifikacija paremta duomenų turiniu ir formatu.

Augant socialinių medijų ir socialinių tinklų vartotojų skaičiui, didėja ir duomenų, kuriuos sugeneruoja socialinės medijos, kiekiai. Tai yra naudinga įmonėms, nes kaupdamos ir analizuodamos tokius duomenis jos gali geriau pažinti savo klientus bei aptikti naudingų išvalgų, kuriomis remiantis gali būti koreguojamos strategijos, atliekami vartotojų profiliavimo, auditorijos segmentavimo ar reputacijos valdymo procesai ar nišinė orientacija, leidžianti susitelkti į siauresnes auditorijas.

Duomenys, gaunami iš didžiųjų duomenų, gali būti analizuojami realiuoju laiku – tam pasitelkiami mašininio mokymosi metodai. Analizuojant socialinių medijų duomenis dažniausiai taikomi klasifikacijos ir klasterizacijos metodai. Planuodama didžiųjų duomenų projektus įmonė turėtų apsibrėžti projekto tikslus ir pasirinkti analizės metodą, kuris būtų efektyviausias, analizuojant sukauptus duomenis.

Didelė dalis projektų, orientuotų į didžiuosius duomenis, žlunga dėl nepakankamų technologinių, finansinių ir žmogiškųjų išteklių. Duomenų analizė yra sudėtingas procesas – didelius duomenų kiekius yra sudėtinga suvaldyti, tai reikalauja didelių išlaidų bei kompetencijos, kad iš surinktų duomenų būtų galima gauti vertingų išvalgų. Prieš imdamos didžiųjų duomenų projektą, įmonė turėtų įsivertinti savo poziciją, galimybes bei apsibrėžti projekto tikslus – taip sumažintų projekto riziką bei eliminuotų nereikalingų ir nepanaudojamų duomenų kaupimą. Neįvertinus galimų rizikų gauti rezultatai gali būti netikslūs, o tai gali lemti projekto žlugimą.

Remiantis literatūros analize parengtas koncepcinis socialinių medijų didžiųjų duomenų panaudojimo modelis. Modelyje atvaizduojamos socialinių medijų ir gaunamų duomenų rūšys. Taip pat vaizduojama proceso, vykstant didžiųjų duomenų analizės projektą, eiga, kurią sudaro duomenų rinkimas, agregavimas, analizė bei panaudojimo galimybės. Sudarytas modelis turėtų padėti planuojant didžiųjų duomenų panaudojimą taikant rinkodaros ir komunikacijos strategijas bei padėti identifikuojant potencialias rizikas ir galimybes.

Literatūroje trūksta modelių, kokius analizės metodus ir kaip įmonė turėtų pasirinkti bei taikyti skirtingų tipų duomenims. Šioje srityje būtina išsamesnė socialinių medijų duomenų panaudojimo galimybių analizė ir tyrimai, kurie padėtų nustatyti tokių duomenų analizės metodų taikymą rinkoje.

Literatūra

- Adweek. (2018). *How marketing teams can unleash personalized creative at scale*. Retrieved from <https://www.adweek.com/digital/how-marketing-teams-can-unleash-personalized-creative-at-scale/>
- Agarwal, S. (2013). Data mining: data mining concepts and techniques. In *Machine Intelligence and Research Advancement (ICMIRA), 2013 International Conference*. Katra, India. <https://doi.org/10.1109/ICMIRA.2013.45>
- Aggarwal, C. C. (2011). *An introduction to social network data analytics. Social network data analytics*. Springer, Boston, MA. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.673.3407&rep=rep1&type=pdf>
- Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. In *Proceedings of the 2008 international conference on web search and data mining* (pp. 183–193). Palo Alto, California. <https://doi.org/10.1145/1341531.1341557>
- Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data mining: a prediction for student's performance using classification method. *World Journal of Computer Application and Technology*, 2(2), 43–47. <https://doi.org/10.13189/wjcat.2014.020203>
- Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, 60(3), 285–292. <https://doi.org/10.1016/j.bushor.2017.01.002>
- Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2018). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics* (In Press, corrected proof). <https://doi.org/10.1016/j.aci.2018.05.004>
- Banerjee, S., & Agarwal, N. (2012). Analyzing collective behavior from blogs using swarm intelligence. *Knowledge and Information Systems*, 33, 523–547. <https://doi.org/10.1007/s10115-012-0512-y>
- Barbu, O. (2014). Advertising, microtargeting and social media. *Procedia-Social and Behavioral Sciences*, 163, 44–49. <https://doi.org/10.1016/j.sbspro.2014.12.284>
- Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *AI & Soc*, 30, 89–116. <https://doi.org/10.1007/s00146-014-0549-4>
- Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: recent achievements and new challenges. *Information Fusion*, 28, 45–59. <https://doi.org/10.1016/j.inffus.2015.08.005>
- Boerman, S. C. (2016). Political microtargeting: relationship between personalized advertising on Facebook and voters' responses. *Cyberpsychology, Behavior, and Social Networking*, 19(6), 367–372. <https://doi.org/10.1089/cyber.2015.0652>
- Boyd, D., & Crawford, K. (2012). Critical question for Big Data. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Canick, H. (2016). *How social media, microtargeting and big data revolutionized political marketing*. Retrieved from https://www.ama.org/publications/MarketingNews/Pages/social-media-big-data-microtargeting-revolutionized-political-marketing.aspx?utm_content=buffer519bf&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer
- Cavanillas, J. M., Curry, E., & Wahlster, W. (Eds.). (2015). *New horizons for a data-driven economy – a roadmap for Big Data in Europe*. Springer, Cham. <https://doi.org/10.1007/978-3-319-21569-3>
- Chen, Y.-C., Peng, W.-C., & Lee, S.-Y. (2012). Efficient algorithms for influence maximization in social networks. *Knowledge and Information Systems*, 33, 577–601. <https://doi.org/10.1007/s10115-012-0540-7>
- Čičević, S., Samčović, A., & Nešić, M. (2016). Exploring college students' generational differences in Facebook usage. *Computers in Human Behavior*, 56, 83–92. <https://doi.org/10.1016/j.chb.2015.11.034>
- Constantinides, E. (2014). Foundations of social media marketing. *Procedia-Social and Behavioral Sciences*, 148, 40–57. <https://doi.org/10.1016/j.sbspro.2014.07.016>
- Constantinides, E., & Fountain, S. J. (2008). Special issue papers Web 2.0: Conceptual foundations and marketing issues. *Journal of Direct, Data and Digital Marketing Practice*, 9, 231–244. <https://doi.org/10.1057/palgrave.ddmp.4350098>

- Couldry, N., & Turow, J. (2014). Advertising, big data and the clearance of the public realm: marketers' new approaches to the content subsidy. *International Journal of Communication*, 8, 1710-1726.
- Cuzzocrea, A., Song, I.-Y., & Davis, K. C. (2011). Analytics over large-scale multidimensional data: The Big Data revolution! In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP* (pp. 101-104). ACM. <https://doi.org/10.1145/2064676.2064695>
- Dong, X. L., & Srivastava, D. (2013). Big data integration. In *Data Engineering (ICDE), 2013 IEEE 29th International Conference on* (pp. 1245-1248). IEEE. <https://doi.org/10.1109/ICDE.2013.6544914>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gartner. (2015). *Gartner says business intelligence and analytics leaders must focus on mindsets and culture to kick start advanced analytics*. Retrieved from <https://www.gartner.com/newsroom/id/3130017>
- Groves, P., Kayyali, B., Knott, D., Kuiken ir S. Van. (2013). The „big data“ revolution in healthcare. *McKinsey Quarterly*, 2(3).
- Gruebner, O., Sykora, M., Lowe, S. R., Shankardass, K., Galea, S., & Subramanian, S. V. (2017). Big data opportunities for social behavioral and mental health research. *Social Science and Medicine*, 189, 167-169. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/26030/3/gruebner_et_al_big_data_v-def.pdf
- Hashem, I. A. T., Yaqoob, I., Anuar, B., Mokhtar, S., Gani, A. ir Khan, U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, (47), 98-115. <https://doi.org/10.1016/j.is.2014.07.006>
- He, W., Zha, S. ir Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464-472. <https://doi.org/10.1016/j.ijinfomgt.2013.01.001>
- Henderson, A. (2010). Authentic dialogue? The role of “friendship” in a social media recruitment campaign. *Article in Journal of Communication Management*, 14(3). <https://doi.org/10.1108/13632541011064517>
- Internet World Stats. (2018). *World internet usage and population statistics*. Retrieved from <https://www.internetworldstats.com/stats.htm>
- Kaisler, S., Armour, F., Espinosa, J. A. ir Money, W. (2014). Big data: Issues and challenges moving forward. *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 995-1004). Wailea, USA. <https://doi.org/10.1109/HICSS.2013.645>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53, 59-68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Kaplan, A. M., & Haenlein, M. (2011). Two hearts in three-quarter time: how to waltz the social media/viral marketing dance. *Business Horizons*, 54(253-263). <https://doi.org/10.1016/j.bushor.2011.01.006>
- Khade, A. A. (2016). Performing customer behavior analysis using big data analytics. *Procedia Computer Science*, 79, 986-992. <https://doi.org/10.1016/j.procs.2016.03.125>
- Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., ... Gani, A. (2014, July 17). Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal*, 2014. Hindawi. <https://doi.org/10.1155/2014/712826>
- Kune, R., Konugurthi, P. K., Agarwal, A., Chillarige, R. R., & Buyya, R. (2016). The anatomy of big data computing. *Software: Practice and Experience*, 46(1), 79-105. <https://doi.org/10.1002/spe.2374>
- Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: a review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 134-142. <https://doi.org/10.1016/j.isprsjprs.2015.11.006>
- Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on* (pp. 911-916). IEEE. Sydney, Australia. <https://doi.org/10.1109/ICDM.2010.35>
- Mcafee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-68.
- MIT Technology Review. (2012). How Obama's team used big data to rally voters. Retrieved from <https://www.technology-review.com/s/509026/how-obamas-team-used-big-data-to-rally-voters/>
- Olshannikova, E., Olsson, T., Huhtamäki, J., & Kärkkäinen, H. (2017). Conceptualizing Big social data. *Journal of Big Data*, 4(1), 0-19. <https://doi.org/10.1186/s40537-017-0063-x>
- Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University – Computer and Information Sciences*, 30(4), 431-448. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- Pennacchiotti, M., & Popescu, A.-M. (2011). A machine learning approach to Twitter user classification. *Icwsn*, 11(2), 281-288.
- Politaite, S., & Sabaitytė, J. (2018). Didžiųjų duomenų naudojimas klientui pažinti. *Mokslas – Lietuvos Ateitis / Science – Future of Lithuania*, 10, 1-10. <https://doi.org/10.3846/mla.2018.932>
- Rajaraman, V. (2016). Big data analytics. *Resonance*, 21(8), 695-716. <https://doi.org/10.1007/s12045-016-0376-7>
- Russom, P., & Org, T. (2011). *Big Data Analytics*.
- Senthilkumar, A. A., Rai, B. K., Meshram, A. A., & Gunasekaran, A. (2018). Big Data in healthcare management: a review of literature. *American Journal of Theoretical and Applied Business*, 4(2), 57-69. <https://doi.org/10.11648/j.ajtab.20180402.14>
- Sharma, S. (2015). Rise of Big Data and related issues. In *2015 Annual IEEE India Conference (INDICON)* (pp. 1-6). IEEE. New Delhi, India. <https://doi.org/10.1109/INDICON.2015.7443346>
- SimilarWeb. (2019). *Website traffic statistics & market intelligence*. Retrieved from <https://www.similarweb.com/>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Statista. (2017). *Number of social media users worldwide 2010–2021*. Retrieved from <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- Statista. (2018). *User-generated internet content per minute 2018*. Retrieved from <https://www.statista.com/statistics/195140/new-user-generated-content-uploaded-by-users-per-minute/>
- Tanwar, M., Duggal, R., & Khatri, S. K. (2015). Unravelling unstructured data: A wealth of information in big data. In *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICRITO.2015.7359270>
- Thevenot, G. (2007). Blogging as a social media. *Tourism and Hospitality Research*, 7(3-4), 287-289. <https://doi.org/10.1057/palgrave.thr.6050062>

- Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How “big data” can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234-246. <https://doi.org/10.1016/j.ijpe.2014.12.031>
- Wang, Y., Wang, S., Tang, J., Liu, H., & Li, B. (2015). Unsupervised sentiment analysis for social media images. In *IJCAI International Joint Conference on Artificial Intelligence, 2015*, 2378-2379. <https://doi.org/10.1109/IJCAI.2015.142>
- We Are Social. (2018). *Global Digital Report 2018*. Retrieved from <https://digitalreport.wearesocial.com/>
- Weiguo, F., & Gordon, M. D. (2014). Unveiling the power of social media analytics. *Communications of the ACM*, 57(6). <https://doi.org/10.1145/2602574>
- Weiss, S. M., & Indurkha, N. (1998). *Predictive data mining: a practical guide*. Morgan Kaufmann Publishers.
- Wildman, S., & Obar, J. A. (2015). Social media definition and the governance challenge: An introduction to the special issue. *Telecommunications Policy*, 39(9), 745-750. <https://doi.org/10.2139/ssrn.2637879>
- Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2003). Data mining with big data Xindong. *IEEE Transactions on Knowledge and Data Engineering*, 15(2), 353-367. <https://doi.org/10.1109/TKDE.2013.109>
- Yin, S., & Kaynak, O. (2015). Big data for modern industry: challenges and trends [point of view]. *Proceedings of the IEEE*, 103(2), 143-146. <https://doi.org/10.1109/JPROC.2015.2388958>
- Zhu, Y.-Q., & Chen, H.-G. (2015). Social media and human need satisfaction: Implications for social media marketing. *Business Horizons*, 58, 335-345. <https://doi.org/10.1016/j.bushor.2015.01.006>

THE BIG DATA USE IN SOCIAL MEDIA

R. Karpovičiūtė, J. Sabaitytė

Abstract

The digital revolution and the communication platforms provided by the web 2.0 virtual space era, such as social media, social networks, other tools and channels, create new opportunities for better marketing decisions based on user-generated data analysis. Every day customers of social media and other virtual tools are creating huge amounts of their actions caused data, and business lack management tools for the support of this process, which could create knowledge in the area of customer profiles and preferences deeper cognition. Growing numbers of social media users indicate the popularity of these communication tools among the information society, but science today lacks a deeper knowledge of social media generated data and other algorithms for this data usage. Therefore, the purpose of the article is defined as the development of the conceptual model of big data generated by social media usage in business. The formation of the conceptual model is based on the analysis of big data assumptions and application possibilities, social media classification peculiarities and different channel specifics, identification of big data analysis methods and analysis of large data applications generated by social media. The conceptual model creates preconditions for deeper knowledge of user-generated big data in nowadays widely used communication platforms, as well as creation of the decision support tool for marketing specialists in order to use big data from social media in deeper customer profile and preferences cognition. Methods employed in this research are: literature and other references analysis, synthesis and logical analysis of information, comparison of information, systemization and visualization.

Keywords: big data, big data analytics, social media, social networks.