



# Sotsiaalmeedia läbistumudel

Kuidas hinnata info läbistusvõimet ühiskonnas  
Facebooki postituse avalike mõõdikute alusel?

Detsember 2022

Läbistumudeli koostas Tartu Ülikooli ühiskonnateaduste instituut Riigikantselei tellimisel.

Uuringu autor: Diana Poudel, TÜ infovastupidavuse nooremteadur

Täna kõik oma kolleegid, kes aitavad kaasa selle põneva väljakutse lahendamisele ning suur tänu nendele Tartu Ülikooli tudengitele, kes toetavad sügise heitlike ilmastikuolusid ja aitavad koguda vajalikku sisendi mudeli loomiseks.

Kontaktandmed:

Lossi 36, 51003, Tartu

Telefon: 5309 7003

E-post: [diana.poudel@ut.ee](mailto:diana.poudel@ut.ee)

<https://uhiskond.ut.ee/et>

## Sisukord

Sissejuhatus.....	4
Uuringu läbiviimine .....	6
Esimene etapp: postituste kaardistamine.....	6
Teine etapp: küsitluste läbiviimine .....	6
Kolmas etapp: andmete töötlemine ja analüüs.....	7
Tulemused.....	8
Vastajate profiil.....	8
Postituste läbistus ühiskonnas.....	11
Läbistusmodeli loomine.....	17
Sotsiaalmeedia läbistusmodel.....	20
Mudeli piirangud.....	20
Uuringu kokkuvõte .....	22
Lisad .....	23
Lisa 1. Postituste mõõdikud .....	23

## Sissejuhatus

Eestis on 1,22 miljonit interneti kasutajat ja neist kasutab aktiivselt sotsiaalmeediat 1,05 miljonit. Kõige populaarsemad on YouTube (1,05 miljonit kasutajat) Facebook (685,6 tuhat), LinkedIn (270 tuhat), Twitter (124,5 tuhat).<sup>1</sup> Paljud inimesed kasutavad erinevaid sotsiaalmeediakanaleid info leidmiseks, jagamiseks ja uue info loomiseks. See on muutnud sotsiaalmeediakanalid atraktiivseks kohaks pahatahtlikele tegutsejatele, kes püüavad inimeste inforuumi suunata propagandat, pettuseid ja valeinfot<sup>2</sup>. Uuringud näitavad, et sotsiaalmeedias levib valeinfo kiiremini kui tõene info<sup>34</sup> ning väga oluline on kiiresti sekkuda<sup>5</sup>.

Sekkumiseks on erinevaid võimalusi – faktikontroll, valeinfo eemaldamine, tõese info jagamine kommentaarides jne, kuid need tegevused võivad olla ajamahukad ning puudub ressurss igale desinformatsiooni sisaldavale postitusele reageerida. Otstarbekas oleks reageerida postitustele, millel on suurem oht ühiskonda läbistada, kuid kahjuks ei avalda näiteks Facebook infot postituse leviku kohta isegi postituse teinud kasutajale (kui postitus on tehtud eraisiku kontolt), mistõttu pole võimalik hinnata, kui paljude inimesteni mingi info jõuab ning kui palju viraalsed postitused ühiskonda läbistavad.

Selle uuringu eesmärgiks on selgitada, kas ja kuidas on võimalik hinnata Facebooki postituse läbistusvõimet ühiskonnas selle avalike mõõdikute alusel. Avalikud mõõdikud on postituse reaktsioonid (*erinevad emoji-d*), kommentaarid, jagamised ning postituse teinud kasutaja võrgustiku suurus. Tegelikult võtavad Facebooki algoritmid arvesse rohkem muutujaid (nt millist tüüpi postitused kasutajale meeldivad, mida ta varem meeldivaks märkinud on, millele ta sõbrad reageerinud on, postituse meediumi tüüpi jne)<sup>6</sup>, kuid selliste andmete kaardistamine on keeruline ja osalt ka võimatu ning seetõttu kasutatakse andmeid, mis on kõigile kättesaadavad.

Meetodi valimisel oli algul kaalumisel võrgustikuanalüüs, kuid Facebook ei võimalda kasutajata pirvaatsussätete tõttu kõiki jagamisi kaardistada, lisaks on paljudel aktiivsetel

---

<sup>1</sup> Kemp, S. (2022). Digital 2022: Estonia. Datareportal. <https://datareportal.com/reports/digital-2022-estonia>

<sup>2</sup> Nimmo, B. ja Agranovich, D. (2022). Recapping Our 2022 Coordinated Inauthentic Behavior Enforcements. Meta. <https://about.fb.com/news/2022/12/metas-2022-coordinated-inauthentic-behavior-enforcements/>

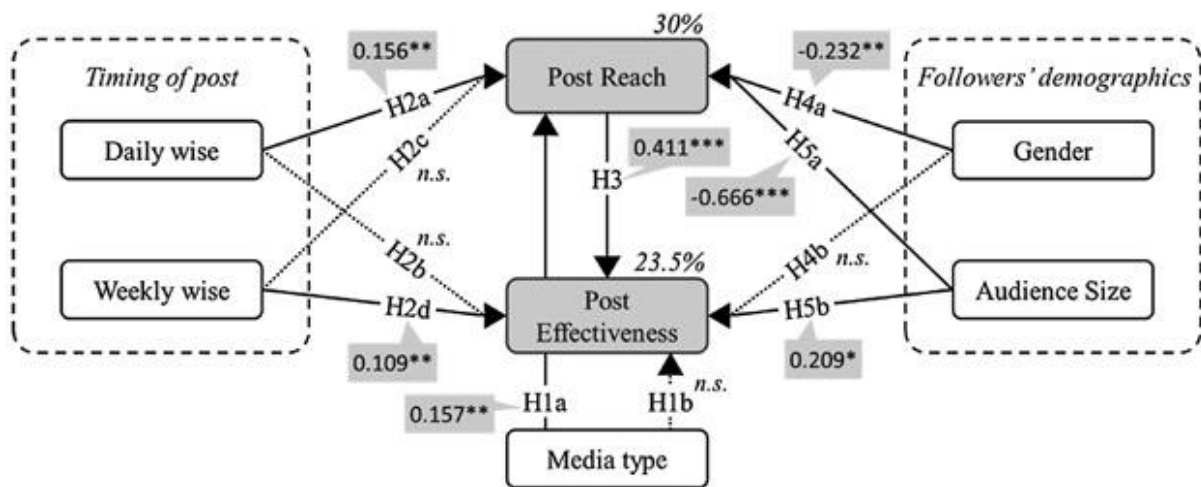
<sup>3</sup> Guess, A.M., Nyhan, B. & Reifler, J. Exposure to untrustworthy websites in the 2016 US election. *Nat Hum Behav* 4, 472–480 (2020)

<sup>4</sup> Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, 359(6380), 1146-1151.

<sup>5</sup> Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, 118(5),

<sup>6</sup> Lada, A., Wang, M. & Yan, T. (2021). How does News Feed predict what you want to see? *Tech at Meta*. <https://tech.facebook.com/engineering/2021/1/news-feed-ranking/>

Facebooki kasutajatel mitu kontot, kuigi see on Facebooki reeglite vastu. Parema idee andis Rassameeroj ja Wu (2019) artikkel<sup>7</sup>, milles kasutati heuristilist lähenemist ja püüti Facebooki algoritmi avada pöördprojekteerimise abil. Nende töö eesmärk oli välja selgitada, kuidas panevad Facebooki algoritmid kasutajaid postitust kommenteerima ning kuidas eelmiste postituste juures reageerimine viib ka uued postitused paremini kasutajate infovoogu. Goncalves jt (2015)<sup>8</sup> löid mudeli (*Joonis 1*), mis näitab kasutajate demograafia, postituse aja ja tüübi seost postituse ulatusega. Siin kasutati sisendina küll Facebooki lehtede sisemist statistikat, kuid seda meetodit on võimalik kohandada ka antud uuringu eesmärkide saavutamisele.



*Joonis 1. Postituse ulatuse seos kasutajate demograafia, postituse tegemise aja ja postituse tüübiga (Goncalves jt, 2015)*

Mudeli loomiseks on vaja teada postituste läbistust ühiskonnas. Parim viis selle sisendi saamiseks on läbi viia uuring internetikeskkonnast väljaspool avalikus ruumis, et vältida algoritmide mõju ja kaasata valimisse võimalikult mitmekesisest läbilõiget Eesti ühiskonnast. Täpsemalt on uuringu ettevalmistust ja selle etappe kirjeldatud järgmises peatükis.

<sup>7</sup> I. Rassameeroj and S. F. Wu, "Reverse Engineering of Content Delivery Algorithms for Social Media Systems," 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2019, pp. 196-203 <https://ieeexplore.ieee.org/document/8931859>

<sup>8</sup> Goncalves, J., Liu, Y., Xiao, B., Chaudhry, S., Hosio, S., & Kostakos, V. (2015). Increasing the reach of government social media: A case study in modeling government–citizen interaction on Facebook. *Policy & Internet*, 7(1), 80-102. <https://onlinelibrary.wiley.com/doi/full/10.1002/poi3.81>

# Uuringu läbiviimine

## Esimene etapp: postituste kaardistamine

Esmalt oli plaanis postituste leidmiseks kasutada tehnilisi vahendeid, kuid siin kerkis üles kaks probleemi. Esiteks ei erista enamik tööriistu postituse keelt, mistõttu on väga keeruline just eesti inforuumi läbi otsida. Teiseks on selliste tehniliste vahendite kasutamine Facebooki kasutustingimustega<sup>9</sup> keelatud ning Facebook võib selliseid tehnilisi abivahendeid kasutavate uurijate kontod blokeerida<sup>10</sup>.

Seetõttu otsustasin kasutada võrgustike abi, sest Eesti inforuum on üsna väike ning eeldasin, et jagades üleskutset oma võrgustikes kus kokku umbes 5000 inimest, siis leian üles viimase aja populaarsemad postitused. Postituste kohta saatis infot ligi 50 inimest ja viie päevaga õnnestus leida kümme postitust, mille reaktsioonide, jagamiste ja kommentaaride summa oli üle tuhande. Kuna sarnast uuringut pole varem tehtud siis ei osanud ma täpselt prognoosida millise tegevuste arvu juures oleks postitused juba piisavalt levinud ühiskonnas ning seetõttu oli kõige kindlam võtta võimalikult viraalsed postitused. Tabel postituste mõõdikutega on *Lisas 1*.

## Teine etapp: küsitluste läbiviimine

Uuringu eesmärgiks on luua mudel, mis võimaldaks desinformatsiooniga tegelevatel inimestel paremini hinnata vajadust Facebookis sekkuda lähtudes postituse ja postituse teinud kasutaja avalikest andmetest. Eesmärgi saavutamiseks viisin läbi küsitluse, kus vastajatele (n=631) näidati avalikus ruumis kümmet tänavu oktoobri algul tehtud Facebooki postitust ning uuriti, et kui laialt on need postitused jõudnud ühiskonnas levida.

Uuring viidi läbi avalikus ruumis nelja nädala jooksul perioodil 20.10.2022 kuni 16.11.2022. Andmeid aitasid koguda kümme Tartu Ülikooli tudengit Tartust, Viljandist ja Pärnust. Lisaks neile asukohtadele viidi küsitlusi läbi ka Tallinnas ning ka mitmetes väiksemates omavalitsustes. Eesmärgiks oli luua võimalikult esinduslik valim kvootide järgi (sugu, vanus, haridus) nagu on näidatud *Tabel 1*.

---

<sup>9</sup> What is data scraping and what can I do to protect my information on Facebook? <https://www.facebook.com/help/463983701520800>

<sup>10</sup> Mir, R. & Doctorow, C. (2021). Facebook's Attack on Research is Everyone's Problem. Electronic Frontier Foundation. <https://www.eff.org/deeplinks/2021/08/facebooks-attack-research-everyones-problem>

Tabel 1. Eesmärgiks seatud kvoot uuringule

Kategooria	Rühmad	Soovitud osakaal
Sugu <sup>11</sup>	Mees	47%
	Naine	53%
Vanus	18-30	17%
	31-45	27%
	46-60	24%
	61+	32%
Haridus <sup>12</sup>	Põhiharidus	9%
	Keskharidus	34%
	Kutseharidus	22%
	Kõrgharidus	35%

Iga küsitleja sai endale küsitlusankeedid ja kümme A4 paberile värviliselt printitud Facebooki postitust (ilma reaktsioonide arvuta). Minimaalne eesmärk oli saada 500 vastajat ja ideaalne valim oleks olnud 1000 vastajat.

Küsitlejad sisestasid täidetud ankeedid jooksvalt läbi <https://survey.ut.ee/> loodud vormi ning projekti koordineerimine toimus jooksvalt Tartu Ülikooli Moodle keskkonnas, kuhu pidevalt postitasin ka infot kvootide kohta. Küsitlusperioodi jooksul toimus kolm kohtumist Zoomis (19.10, 2.11 ja 16.11), kus vaatasime koos üle hetke olukorra seoses kvootidega ning panime paika strateegiad soovitud kvootide saavutamiseks.

Kokku vastas küsitlusele 631 inimest..

### **Kolmas etapp: andmete töötlemine ja analüüs**

Kolmandas etapis eksportisin andmed survey.ut.ee lehelt andmetöötlusprogrammi R ning puhastasin ja kodeerisin andmed. Antud aruandes olevate jooniste loomiseks kasutasin pakette ggplot2 ja VennDiagram. Regressioonimudelite loomiseks kasutasin paketti lavaan.

Kõrvutasin andmeid postituse avalike mõõdikutega (reaktsioonid, meeldimised, jagamised) ja postitaja konto näitajatega (sõprade ja jälgijate) arv ning lõin regressioonanalüüsi abil mudeli, mis aitab prognoosida Facebooki postituse võimet läbistada ühiskonda.

---

<sup>11</sup> Rahvastik aasta alguses ja aasta keskmine rahvaarv soo ja vanuse järgi [https://andmed.stat.ee/et/stat/rahvastik\\_rahvastikunaitajad-ja-koosseis\\_rahvaarv-ja-rahvastiku-koosseis/RV0212](https://andmed.stat.ee/et/stat/rahvastik_rahvastikunaitajad-ja-koosseis_rahvaarv-ja-rahvastiku-koosseis/RV0212)

<sup>12</sup> Haridus | Statistikaamet <https://rahvaloendus.ee/et/tulemused/haridus>

# Tulemused

## Vastajate profiil

Küsitlusele vastas 631 inimest, kelle jaotus soo, vanuse ja hariduse järgi on näidatud Tabel 2. Suurim kõrvalekalle kvoodist on noorima ja vanima vanusegrupi puhul. Seda on tulemuste analüüsimisel arvesse võetud.

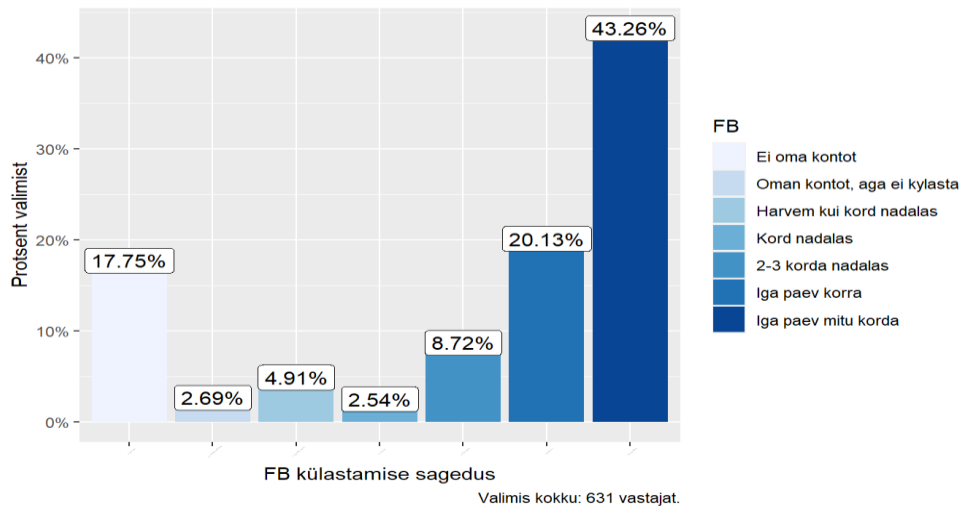
Tabel 2. Vastajate jaotus soo, vanuse ja hariduse järgi

Kategooria	Rühmad	Soovitud osakaal	Arv lõppvalimis	Osakaal lõppvalimis
Sugu	Mees	47%	266	42%
	Naine	53%	365	58%
Vanus	18-30	17%	185	29%
	31-45	27%	148	24%
	46-60	24%	143	23%
	61+	32%	147	23%
	Ei soovinud vastata		8	1%
Haridus	Põhiharidus	9%	36	6%
	Keskharidus	34%	205	32%
	Kutseharidus	22%	119	19%
	Kõrgharidus	35%	265	42%
	Ei soovinud vastata		6	1%

Hariduse poolelt tahan olulise tähelepanekuna esile tuua, et mitmed küsitluste läbiviijad tõid tagasisides esile, et tihti jäid vastajad pikemalt mõtlema hariduse kohta käiva küsimuse juures ning madalama haridustaseme puhul tihti sooviti põhjendada seda. Seega on võimalik, et osad inimesed ei soovi oma õiget haridustaset avaldada, kui nad pole sellega rahul ning pigem antakse vastuseks üks haridustase kõrgem. See võib selgitada, miks hoolimata suurtest jõupingutustest oli väga keeruline leida põhiharidusega vastajaid.

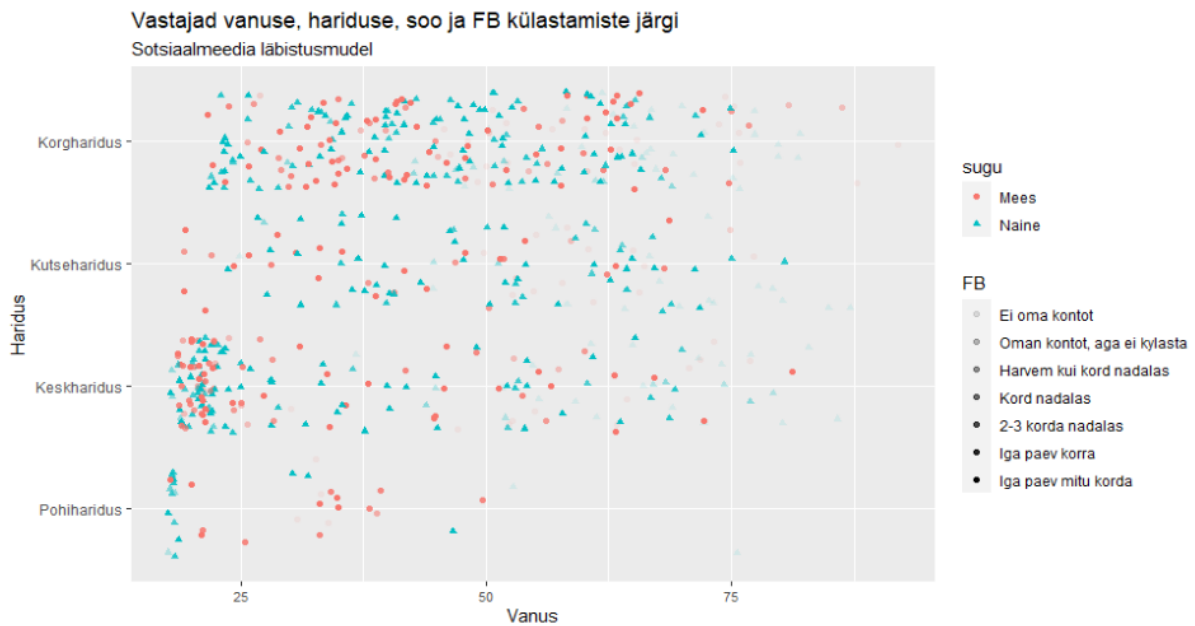
Selleks, et välja selgitada Facebooki olulisust infokanalina Eesti ühiskonnas, küsiti vastajatelt, kui tihti nad platvormi külastavad. Vastajatel paluti mitte arvestada Messengeri rakenduse kasutamist. Tulemused näitavad, et Facebook on Eestis väga populaarne ning 63% vastajatest kasutab seda vähemalt korra päevas ning ainult 18% vastajatest ei oma kontot (vt Joonis 2).





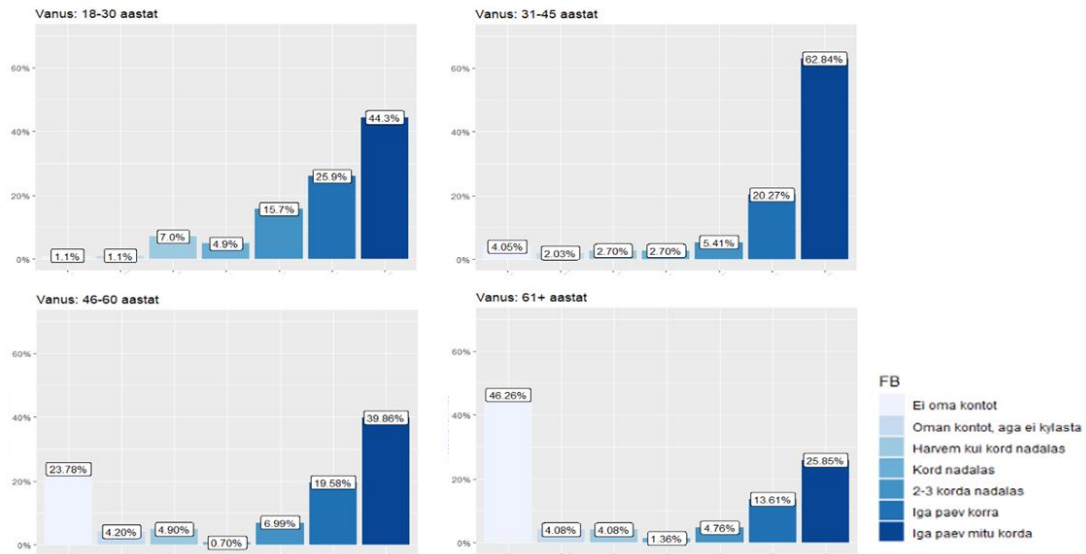
Joonis 2. Vastused küsimusele: „Kui tihti külastatate Facebooki? (Messenger ei loe)“

Vastajate jaotust soo, vanuse, hariduse ja Facebooki kasutamise järgi näitab Joonis 3. Siit on välja jäetud need vastajad, kes ei soovinud anda infot oma vanuse või hariduse kohta (n=14).



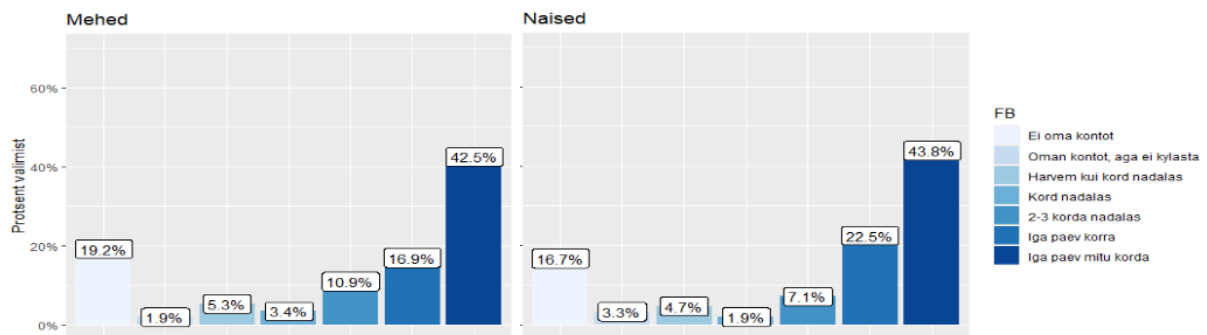
Joonis 3. Vastajate jaotus soo, vanuse, hariduse ja Facebooki kasutamise sageduse järgi.

Jooniselt on näha, et vanemate inimeste seas on rohkem neid, kes Facebooki ei kasuta. Kui suur on erinevus vanusegruppide lõikes täpsemalt näitab Joonis 4.



Joonis 4. Facebooki kasutamine vanuste lõikes

Kõige aktiivsem vanusegrupp Facebooki kasutamisel on vanuses 31-45 aastat, kus 83% vastajatest kasutab Facebooki igapäevaselt. Kõige nooremas rühmas kasutavad Facebooki igapäevaselt 70% vastajatest, kuid huvitav on see, et selles vanusegrupis on päris suur nende vastajate osakaal, kes kasutavad Facebooki vähem kui kord nädalas (7%). Vanemate kui 61-aastaste inimeste seas ei ole Facebook väga populaarne ja pooled vastajad ütlesid, et neil pole kontot või on konto küll olemas, kuid ei kasutata lehte.



Joonis 5. Facebooki kasutamise sagedus meeste ja naiste lõikes

Joonis 5 on toodud Facebooki kasutamise sagedus meeste ja naiste lõikes ning on näha, et igapäevaselt kasutab Facebooki 66% naistest ning 59% meestest. Ehk siis sugude lõikes suurt erinevust ei ole.

## Postituste läbistus ühiskonnas

Et hinnata postituste levikut ühiskonnas, paluti igal vastajal vaadata postitusi ning vastata iga postituse puhul küsimusele, et kas on seda postitust Facebookis näinud. Vastuse variandid olid „Jah“, „Võibolla“, „Ei“ ja „Ei oska öelda“. Vastused kodeeriti järgnevalt: „Jah“ vastus andis postitusele 1 punkti, „Võibolla“ 0,5 punkti ning „Ei“ ning „Ei oska öelda“ 0 punkti ning iga postituse tulemus liideti kokku ning jagati vastajate arvuga (n=631) ning seda loeti postituse läbistusnäitajaks (Tabel 3).

Tabel 3. Postituste läbistusnäitajad

Postituse nr	1	2	3	4	5	6	7	8	9	10
Läbistuvus	10%	27%	13%	11%	10%	8%	14%	11%	25%	15%
Võrgustik	31000	26329	19266	9092	5713	1659	24500	22500	1310	17738
Tegevused	2828	2415	1471	3029	1173	1071	1545	1360	1406	1259

Kõige tugevamalt olid ühiskonda läbistanud 2. ja 9. postitus, mida oli näinud põhimõtteliselt iga neljas vastaja. Kõige kehvema läbistuvusega oli kuues postitus, mis oli jõudnud ainult 8% vastajateni. Tabel 3 näitab ka postituse kohta, kui suur on selle autori võrgustik (sõbrad ja jälgijad kokku) ning kui palju erinevaid tegevusi oli postituse juures 19.10 seisuga. Tegevuste all peetakse silmas postituse kommentaare, reaktsioone postitusele ning jagamisi. Kontrollisin ka üle postitusega seotud tegevused peale küsitlusperioodi lõppu, kuid kuna 9. postitus oli vahepealsel perioodil raporteerimiste tõttu eemaldatud ning ülejäänud postituste juures polnud tegevuste arv märkimisväärselt muutunud (samuti ei olnud muutunud võrgustiku suurus), siis võrreldavuse huvides kasutan enne küsitluseperioodi algust fikseeritud näitajaid.

Tabel 3 näitab, et võrgustiku suurus ei määra postituse levikut. Näiteks üheksanda postituse autori võrgustik on väga väike, kuid postitus levis hästi. Samas esimese postitaja võrgustik on enam kui 20 korda suurem, kuid postituse levis märkimisväärselt vähem. Samuti ei ole otsest seost postituse reaktsioonide ja postituse leviku vahel. Näiteks neljanda postituse reaktsioonide arv on kõige suurem, kuid postitus ise pole nii laialt levinud kui mitmed teised.

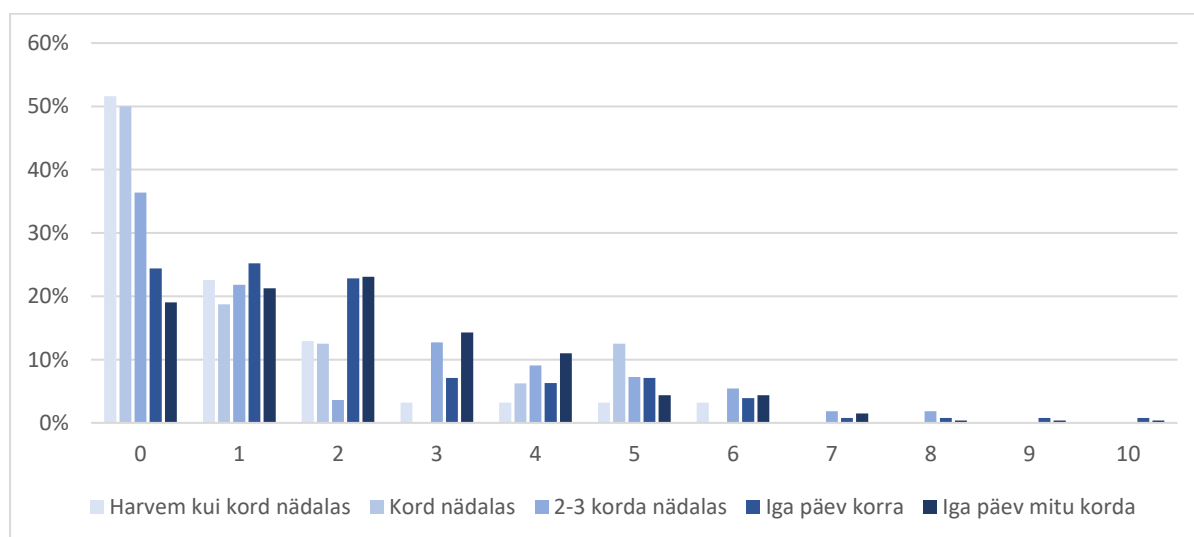
Põnev küsimus on ka see, et kuidas mõjutab Facebooki küllastuste sagedus viraalsete postituste jõudmist inimese uudisvoogu. Tabel 4 on toodud vastajate arv nähtud postituste ja nende Facebooki küllastamise sageduse järgi. Siin tabelis pole neid vastajaid, kes Facebooki

kontot ei oma või omavad kontot, kuid Facebooki ei kasuta. Väga palju vastajaid ei olnud näinud ühtegi valimis olevat postitust, ning ainult kaks vastajat olid näinud kõiki 10 postitust Facebookis.

Tabel 4. Nähtud postituste arv Facebooki külastamise sageduse alusel

		Mitu FB postitust on vastaja näinud											
		0	1	2	3	4	5	6	7	8	9	10	Kokku
FB külastamise sagedus	Harvem kui kord nädalas	16	7	4	1	1	1	1	0	0	0	0	31
	Kord nädalas	8	3	2	0	1	2	0	0	0	0	0	16
	2-3 korda nädalas	20	12	2	7	5	4	3	1	1	0	0	55
	Iga päev korra	31	32	29	9	8	9	5	1	1	1	1	127
	Iga päev mitu korda	52	58	63	39	30	12	12	4	1	1	1	273
		127	112	100	56	45	28	21	6	3	2	2	502

Joonis 6 on antud info esitatud vastajate osakaal postituse nähtavuse osas vastavalt nende FB külastamise sagedusele. Näiteks nende vastajate seas, kes külastavad Facebooki kord nädalas või harvem ei olnud ligi 50% näinud mitte ühtegi postitust. Samas nende vastajate hulgas, kes külastavad Facebooki korra päevas või sagedamini oli enamvähem võrdset inimesi, kes polnud näinud ühtegi postitust või olid näinud 1-2 postitust. Neid inimesi, kes olid näinud seitset või enam postitust oli vastajate seas väga vähe.



Joonis 6. Vastajate osakaal postituste nähtavuse osas nende Facebooki kasutamise sageduse lõikes

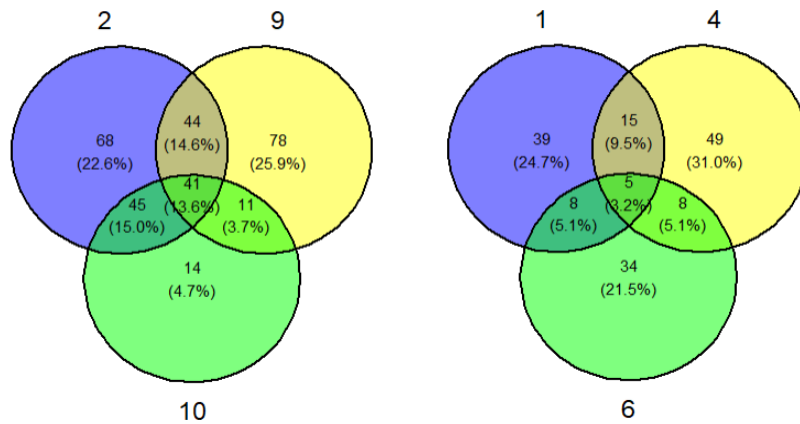
Tihti räägitakse sotsiaalmeediaga seoses filtrimullist, ning seetõttu otsustasin analüüsida kui suur on kattuvus erinevate postituste auditooriumitel. Tabel 5 toob välja auditooriumite kattuvuse erinevate postituspaaride puhul. Tumesinisel taustal on iga konkreetse postituse auditooriumi suurus (vastajate arv, kes ütles, et on kindlalt või võibolla seda postitust näinud). Rohelisel taustal on suurima auditooriumiga postitusepaarid ja punase taustaga on näidatud postitusepaarid, mille auditooriumite kattuvus on kõige väiksem.

Tabel 5. Postituste auditooriumite kattuvus

	1	2	3	4	5	6	7	8	9	10
1	67	40	26	20	19	13	29	22	36	26
2	40	198	53	43	35	30	62	34	85	86
3	26	53	97	28	30	15	51	23	53	33
4	20	43	28	77	20	13	25	17	37	35
5	19	35	30	20	77	15	31	30	37	25
6	13	30	15	13	15	55	14	20	33	20
7	29	62	51	25	31	14	108	37	68	33
8	22	34	23	17	30	20	37	80	39	20
9	36	85	53	37	37	33	68	39	174	52
10	26	86	33	35	25	20	33	20	52	111

Kui teha Venni diagramm (vt Joonis 7) kõige suurema ühisosaga postituste kohta (2, 8 ja 10) siis selgub, et kõiki kolme postitust olid näinud 41 inimest (6,5% vastajatest). Eriti suur on ühisosa 2. ja 9. postituse vahel, kus mõlemat postitust oli näinud 85 vastajat (13,5%). Huvitav on siin ka see, et ainult 10. postitust on näinud 14 inimest (2,2%) kuigi postitus ise on jõudnud 111 inimeseni (17,6%). Ehk siis selle postituse auditoorium kattub väga tugevalt 2. ja 9. postituste auditooriumitega, kuid vastupidist trendi nii tugevalt esile ei tule. Väga huvitav oleks teada saada, kas see on postituste sisust tingitud kattuvus (Ukraina sõda ja päevapoliitika) või on nende kontode vahel ka üldiselt suur kattuvus jälgijate hulgas.

Kõige väiksem oli ühisosa postitustel 1, 4 ja 6 ja kõiki kolme postitust olid näinud vaid viis vastajat (0,8%). Tegu on väiksema läbistusega postitusteha, mille sisu kõnetab väga erinevaid auditooriume. Aga filtrimullid on täiesti olemas ka Eesti inforuumis.



Joonis 7. Venni diagramm kolme kõige suurema auditooriumi kattuvusega ja kolme kõige väiksema auditooriumi kattuvusega postituse kohta.

Kontrollisime ka seda, kas postituste nähtavus sõltub vastajate demograafilistest näitajatest. Tabel 6 näitab iga postituse nähtavust eri demograafilistes gruppides. Tabelis on rohelisega kuvatud need lahtrid, mille väärtus on üle keskmise ja punasega on markeeritud need, mille väärtus on alla postituse üldise keskmise.

Tabel 6. Postituse nähtavus demograafiliste näitajate põhjal

			1	2	3	4	5	6	7	8	9	10
Sugu	Mehed	266	9%	33%	14%	12%	10%	6%	16%	6%	21%	19%
	Naised	365	11%	24%	13%	10%	11%	9%	14%	15%	29%	12%
Haridus	Põhiharidus	36	17%	17%	13%	8%	17%	8%	13%	13%	19%	8%
	Keskharidus	205	10%	26%	15%	8%	14%	7%	21%	13%	28%	14%
	Kutseharidus	119	16%	28%	16%	10%	11%	9%	16%	16%	22%	14%
	Kõrgharidus	265	6%	31%	11%	14%	6%	9%	9%	7%	27%	18%
Vanus	18-30	185	6%	23%	12%	7%	14%	11%	20%	11%	32%	11%
	31-45	148	10%	29%	17%	13%	8%	13%	17%	10%	32%	18%
	46-60	143	10%	36%	11%	17%	9%	4%	9%	12%	18%	20%
	61+	147	13%	26%	15%	8%	9%	4%	13%	10%	19%	14%
<b>Kokku</b>			<b>10%</b>	<b>28%</b>	<b>13%</b>	<b>11%</b>	<b>11%</b>	<b>8%</b>	<b>15%</b>	<b>11%</b>	<b>26%</b>	<b>15%</b>

Kui Venni diagramm näitas, et postitustel 2, 9 ja 10 on väga suur auditooriumi kattuvus, siis üllatavalt on neist andmetest näha, et postitus 9 levis palju paremini naiste seas ja postitused 2 ja 10 levisid paremini meeste seas. Meeste seas oli kõige populaarsem postitus 2 mida oli näinud iga kolmas meessoost vastaja. Naiste seas oli kõige populaarsem postitus 9, mida oli näinud samuti ligi iga kolmas naine vastajate hulgast.

Kui võrrelda neid kolme suure ühisosaga postituse hariduse lõikes, siis on näha, et need levisid kehvemini põhiharidusega vastajate seas, kuid postitused 2 ja 10 olid jõudnud kõige paremini kõrgharidusega vastajateni samal ajal kui 9 postitus kõnetas väga ka keskharidusega

vastajaid. Huvitav tähelepanek on siin, et postituste levik igas vanusegrupis erinev ning lineaarseid trende eriti ei esine (erandiks postitus 1, mille populaarsus oli seotud vanusega).

Vanuse puhul on nendest kolmest postitusest kõnetanud kõige paremini nooremaid (18-30 aastat) 9. postitus, kuid samas erinevalt teisest kahest ei levinud ta hästi vanemate (61+ aastat) inimesteni. Postitus 6 on samuti kõnetanud eelkõige kahte nooremat vanusegruppi ning selle postituse levik vanemate vastajate seas oli väga väike. Ehk siis isegi suure auditoriumi kattuvusega postituste puhul võivad need levida erinevate sihtrühmade seas.

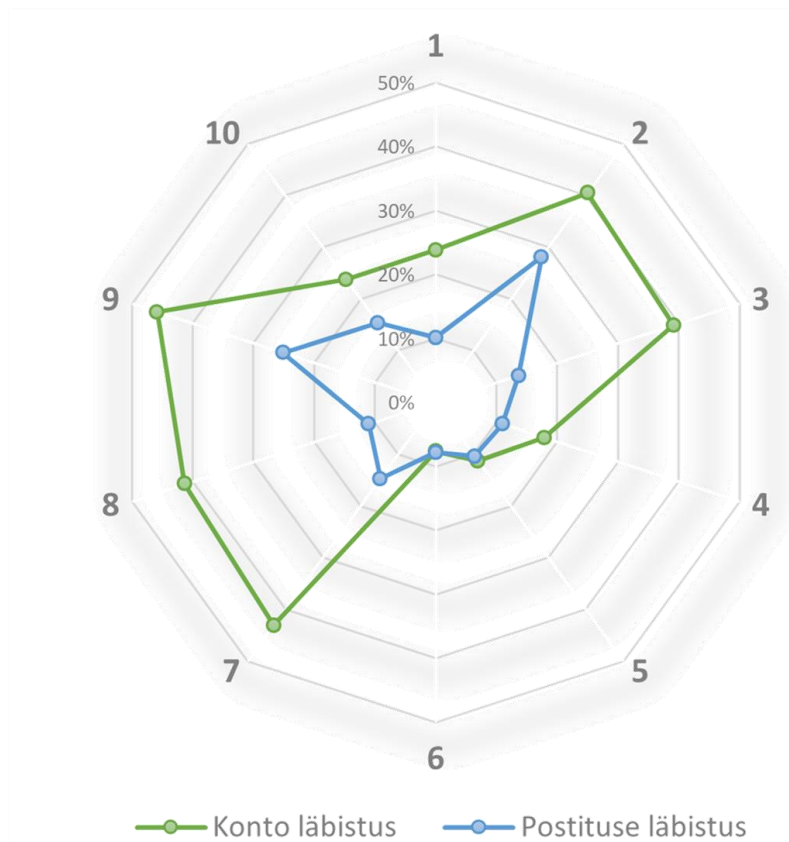
Nüüd on ka küsimus selles, et kui tugevalt on seotud konto nähtavus ja postituste nähtavus. Palusime vastata ka küsimusele „Kas olete seda kontot näinud?“ ja vastusevariandid olid „Jälgin kontot“, „Keegi on jaganud“, „Ei tea seda kontot“ ja „Ei oska öelda“. Kodeerimisel arvestasime, et nii konto jälgimine kui selle nägemine ajajoonel andis ühe punkti ja vastus „Ei tea seda kontot“ ja „Ei oska vastata“ kodeeriti nulliks. Selle põhjal on tehtud Tabel 7, kus värvikoodid on sama süsteemi järgi kui eelmises tabelis.

Tabel 7. Konto nähtavus demograafiliste näitajate põhjal

			1	2	3	4	5	6	7	8	9	10
Sugu	Mehed	266	17%	47%	45%	21%	10%	6%	45%	29%	42%	27%
	Naised	365	29%	36%	37%	16%	12%	9%	43%	52%	51%	22%
Haridus	Põhiharidus	36	18%	34%	42%	15%	12%	9%	47%	46%	61%	19%
	Keskharidus	205	22%	36%	50%	12%	15%	8%	51%	47%	55%	23%
	Kutseharidus	119	31%	31%	36%	16%	11%	6%	43%	49%	39%	18%
	Kõrgharidus	265	23%	48%	33%	23%	8%	9%	37%	35%	43%	29%
Vanus	18-30	185	17%	33%	55%	13%	13%	11%	55%	55%	67%	24%
	31-45	148	27%	50%	48%	24%	12%	13%	54%	49%	55%	25%
	46-60	143	27%	44%	29%	23%	11%	3%	31%	39%	33%	30%
	61+	147	24%	37%	24%	13%	9%	4%	31%	23%	28%	17%
<b>Kokku</b>			<b>24%</b>	<b>40%</b>	<b>40%</b>	<b>18%</b>	<b>11%</b>	<b>8%</b>	<b>44%</b>	<b>42%</b>	<b>47%</b>	<b>24%</b>

Kui võrrelda kahe tabeli läbistusnäitajat, siis enamike kontode puhul oli konto nähtavus suurem, kui postituse nähtavus, mis on ka loogiline. Anomaaliad olid 5. ja 6. konto, kus kontot olid eelkõige näinud need inimesed, kes olid ka postitust näinud. Need olid ka võrreldes enamike teistega väiksema jälgijate arvuga kontod (5. kontol 5713 inimest võrgustikus ning 6. 1659 inimest võrgustikus), kuid sellisel juhul tekib küsimus, et kuidas on 9. kontol nii hea nähtavus Facebookis, kuigi sellel on ainult 1310 sõpra ning jälgimisvõimalust kontol ei ole. Taustainfoks võib küll öelda, et tegu on ühiskonnas tuntud inimesega ning võiks ju eeldada, et inimestel on tunne, et nad on kontot näinud. Selle vastu räägib asjaolu, et 26% inimestest

oli näinud ka postitust. Kõige suurem erinevus konto ja postituse läbistuse osas oli 8. konto puhul (3,81 korda rohkem inimesi oli näinud konto teisi postitusi).



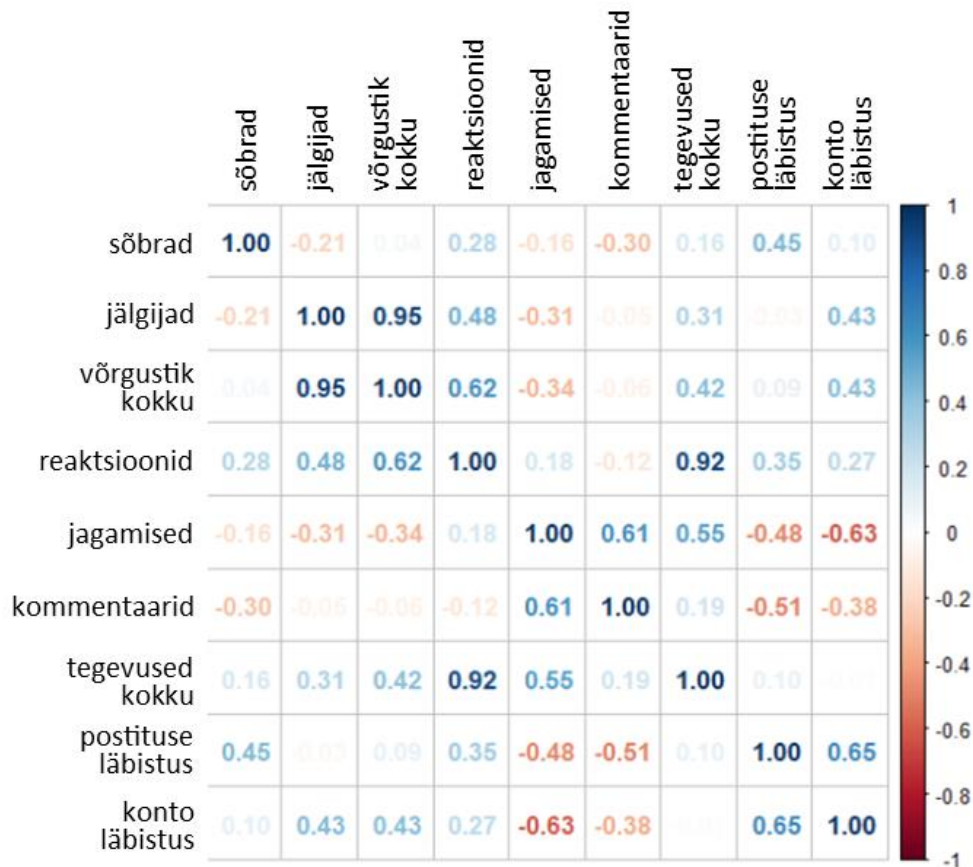
Joonis 8. Kui palju erinevad postituse läbistus ühiskonnas ja konto läbistus ühiskonnas?

Kui vaadata konto nähtavust demograafiliste näitajate alusel, siis näitab Tabel 7, et soolises võrdluses on eriti suur erinevus 1. ja 8. kontol, mille auditooriumis on märgatavalt enam naisi. Põhiharidusega ja keskharidusega vastajad teadsid kõige enam 9. kontot, kutseharidusega vastajate seas teati kõige paremini 8. kontot ning kõrgharidusega vastajate seas tundub populaarseim olevat 2. konto. Vanusegruppide puhul näeme, et tugevaim seos vanuse ja huvi vastu konto osas tundub olevat 5., 8. ja 9. kontol, kus vanuse kasvades on vähem neid, kes selle konto postitusi on üldse näinud oma uudisvoos.



## Läbistusmudeli loomine

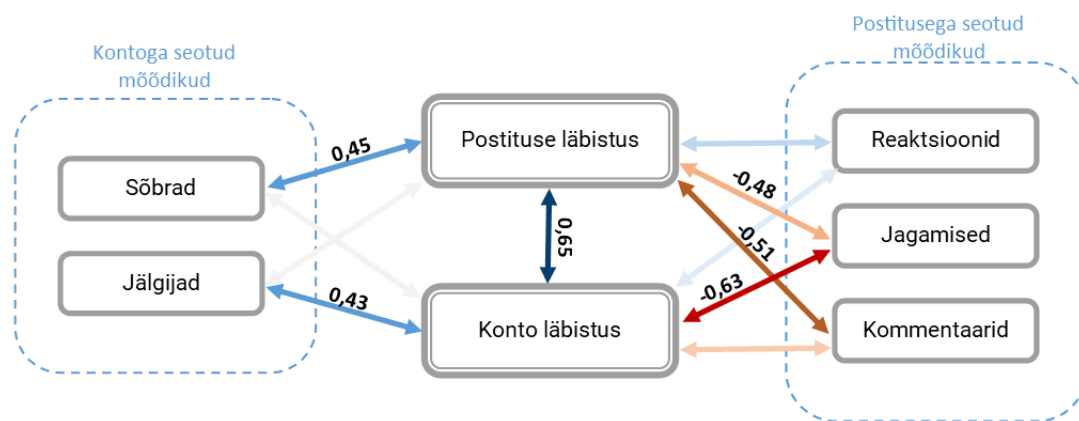
Kuna selle uuringu eesmärgiks on mudeli loomine, mis aitaks hinnata erinevate Facebooki postituste läbistusvõimet ühiskonnas, siis esmalt koostati korrelatsioonimaatriks (Joonis 9), et kontrollida seoseid erinevate mõõdikute vahel.



Joonis 9. Korrelatsioonimaatriks postituste mõõdikute vahel

Kuna Facebookis on kontaktide ülempiir 5000 sõpra, kuid kontol võib piiramatu arv jälgijaid, siis on loogiline, et võrgustiku suurusel on tugev korrelatsioon järgijate arvuga, kuid mitte sõprade arvuga. Üllatav on aga see, et võrgustiku suurus on pöördvõrdelises seoses jagamiste ja kommentaaride arvuga. Selle kohta head selgitust teaduskirjandusest ei leidnud, kuid ühe hüpoteesina pakuks välja, et mida suurem on inimese võrgustik, seda rohkem on seal erinevate huvidega inimesi ning seetõttu üksik postitus kõnetab vähemat osa tema jälgijatest. Konto läbistusel on tugev korrelatsioon võrgustikuga, mis tundub seda hüpoteesi ka kinnitavat.

Huvitav on ka see, et postituste läbistus on negatiivses korrelatsioonis postituse jagamise ja kommentaaridega, kuid positiivses korrelatsioonis postituse reaktsioonidega (Joonis 10). Kindlasti tuleb arvesse võtta, et valimisse sattusid ainult postitused millel oli tegevuste arv suur ning seetõttu ei saa järeldada, et postituse jagamine või kommenteerimine vähendab postituse levikut. Lihtsalt nende tegevuste mõju postituse laiemale levikule ei pruugi nii suur olla, kui tihti eeldatakse.



Joonis 10. Kõige tugevamad korrelatsiooniseosed joonisena

Regressioonianalüüsi jaoks arvasin ümber ka postituse läbistuse kvootide põhjal. Tabel 8 näitab postituste läbistust peale demograafiliste näitajate kaalude arvesse võtmist. Enamike postituste puhul on erinevus väga väike, kuid võimalikult täpse mudeli saamiseks kasutan edaspidi kaalutud läbistusnäitajat.

Tabel 8. Postituste läbistus arvatatud demograafiliste kaalude põhjal

	1	2	3	4	5	6	7	8	9	10
<i>Sugu</i>	9,6%	28,4%	13,4%	11,0%	10,4%	7,9%	15,0%	10,7%	25,4%	15,6%
<i>Haridus</i>	10,5%	27,4%	13,7%	10,5%	11,1%	8,1%	15,1%	11,5%	25,5%	14,8%
<i>Vanus</i>	10,4%	28,7%	13,9%	11,3%	9,6%	7,4%	14,1%	10,8%	24,4%	15,8%
<b><i>Kaalutud</i></b>	<b>10,2%</b>	<b>28,2%</b>	<b>13,7%</b>	<b>10,9%</b>	<b>10,4%</b>	<b>7,8%</b>	<b>14,7%</b>	<b>11,0%</b>	<b>25,1%</b>	<b>15,4%</b>
<i>Algne</i>	9,7%	28,0%	13,3%	10,9%	10,5%	8,1%	14,9%	11,1%	25,8%	15,3%

Täpsemaks analüüsiks tegin üheksa erinevat regressioonimudelt, et võrrelda nende kirjeldusvõimet postituse läbistuvuse hindamiseks. Tulemused on toodud Tabel 9.

Tabel 9. Regressioonimudelite võrdlus

	Tunnused mudelis	$R^2$	$p$
Mudel 1	jagamised	0,1374	0,1574
Mudel 2	kommentaarisid	0,1662	0,1331
Mudel 3	reaktsioonid	0,0161	0,3153
Mudel 4	jagamised, kommentaarisid, reaktsioonid	0,1780	0,2752
Mudel 5	sõbrad, jälgijad	-0,0256	0,4533
Mudel 6	võrgustik kokku	-0,1151	0,0797
Mudel 7	jagamised, kommentaarisid, reaktsioonid, võrgustik kokku	<b>0,7443</b>	<b>0,02392</b>
Mudel 8	jagamised, kommentaarisid, reaktsioonid, sõbrad, jälgijad	0,4549	0,1975
Mudel 9	tegevused kokku, võrgustik kokku	-0,2792	0,9558

Kõige paremate näitajatega on Mudel 7, mis on mitmene lineaarne regressioonimudel, mille determinatsioonikordaja on 75% ning p-väärtus 0,02. Mudeli kordajad on näidatud Tabel 10.

Tabel 10. Läbistuvuse hindamise regressioonimudeil (Mudel 7) jagamiste, kommentaaride, reaktsioonide ja võrgustiku suuruse kaudu.

	Regressioonikordaja	Regressioonikordaja standardhälve	t	Olulisuse tõenäosus p
<b>Jagamised (j)</b>	-0,041	0,010	-4,196	0,00852**
<b>Kommentaarisid (k)</b>	0,049	0,027	1,728	0,14455
<b>Reaktsioonid (r)</b>	0,016	0,003	4,406	0,00698**
<b>Võrgustik (v)</b>	-0,0008	0,0002	-3,931	0,01106**
<b>Vabaliige</b>	12,778	3,203	4,114	0,00923**

## Sotsiaalmeedia läbistusmudel

Elmises alapeatükis loodud mudelite põhjal selgus, et postituse läbistavust saab hinnata mudeliga, mida kirjeldab valem

$$12,778 - 0,041 * j + 0,049 * k + 0,016 * r - 0,0008 * v,$$

kus  $j$  jagamisete arv,

$k$  kommentaaride arv,

$r$  reaktsioonide arv ja

$v$  võrgustikus olevaid inimesi (sõbrad ja jälgijad kokku).

### Mudeli piirangud

Mudel on koostatud kümne postituse baasil, mis tehti 2022. aasta oktoobri alguses. Regressioonimudeli näitajad on head, kuid kindlasti tuleb arvesse võtta piiratud näidete hulka, seega antud mudel on hinnanguline ning mudeli koostaja ei saa garanteerida mudeli täpsust iga postituse puhul. Mudeli eesmärgiks on aidata inimestel ja organisatsioonidel hinnata mingi info leviku ulatust, kuid kindlasti tuleks lisaks info levikule võtta arvesse ka info eesmärki ja selle negatiivse mõju suurust. Samuti on tihti nii, et infot ei jagata algallikast, vaid tehakse uusi postitusi ning postituse sisu läbistus ühiskonnas võib olla märgatavalt suurem üksiku postituse läbistusest.

Oluline on arvesse võtta ka asjaolu, et Facebook muudab pidevalt oma algoritme<sup>13</sup> ning mudel muutub ajapikku ebatäpsemaks. Juhul kui seda mudelit kasutatakse igapäevase töövahendina, siis on tungivalt soovituslik viia aasta pärast läbi kordusuuring, et kontrollida mudeli täpsust.

---

<sup>13</sup> Newberry, C. (2022). How the Facebook Algorithm Works in 2023 and How to Make it Work for You. Hootsuite blog. <https://blog.hootsuite.com/facebook-algorithm/>

Antud mudel on koostatud postituste baasil, mille kommentaaride, jagamiste ja reaktsioonide summa jäi 1071 ja 3029 vahele. Mudel sobivust sellest väikemate ja suuremate postitusega seotud tegevuste summa puhul tuleks eraldi testida.

Samuti ei sobi mudel nende postituste läbistuse hindamiseks, kus on kasutatud postituse leviku tõstmiseks mingeid väliseid motivaatoreid – näiteks „Jaga ja võida“ tüüpi postitused.

### Näide 1

Postitus, millel on 400 jagamist, 500 kommentaari, 600 reaktsiooni ning postituse teinud kasutajal on 2000 sõpra ja 4000 jälgijat, siis arvutuskäik on järgmine:

$$12,778 - 0,041 * 400 + 0,049 * 500 + 0,016 * 600 - 0,0008 * (2000 + 4000) = 25,7\%$$

Ehk siis mudeli järgi oleks seda hüpoteetilist postitust näinud iga neljas Eesti inimene.

### Näide 2

Postitusel on 600 jagamist, 150 kommentaari ja 300 reaktsiooni ning postituse teinud kasutajal on 15 000 jälgijat, siis arvutuskäik on järgmine:

$$12,778 - 0,041 * 600 + 0,049 * 150 + 0,016 * 300 - 0,0008 * 15\,000 = -11,7\%$$

**Tulemus ei tule loogiline seetõttu, et väga harva on Facebookis postitusi, mille jagamiste arv on suurem kui postituste reaktsioonid ning tihti on põhjuseks auhinnamängud või mõni muu stiimul. Selliseid postitusi selle mudeli alusel hinnata ei saa.**

### Näide 3

Postitusel on 100 jagamist, 50 kommentaari ja 200 reaktsiooni ning postituse teinud kasutajal on võrgustikus 12 000 sõpra.

$$12,778 - 0,041 * 100 + 0,049 * 50 + 0,016 * 200 - 0,0008 * 12\,000 = 4,7\%$$

**Seda tulemust ei saa usaldada, sest postituse reaktsioonide + kommentaaride + jagamiste arv on alla tuhande ning selliste postitustega pole selle mudeli toimivust kontrollitud.**

## Uuringu kokkuvõte

Antud uuringu põhjal saab väita, et viraalsete postituste läbistusvõimet ühiskonnas on teatud piirini võimalik prognoosida avalike andmete baasil. Kindlasti võiks tulevikus täpsemalt ka uurida varasemate postituste mõju postituse läbistavusele. Olulised tähelepanekud uuringu käigus on järgmised:

- Eestis on Facebook väga populaarne platvorm. Eriti populaarne on see vanusegrupis 31-45 aastat, kus 83% vastajatest külastab Facebooki igapäevaselt. Seega on oluline hoida Facebookis toimuval silma peal, sest kui desinfo hakkab liikuma läbi Facebooki siis jõuab see väga kiirelt suure auditooriumini.
- Postituste auditooriumid võivad olla väga erinevad ning seetõttu ei pruugi inimestel isegi väga aktiivse Facebooki kasutamise korral olla täit ülevaadet toimuvast. Selle uuringu käigus kasutati erinevate võrgustike abi viraalsete postituste üles leidmisel. Võrgustike kaasamine tundubki olevat parim viis sotsiaalmeedias sisu monitoorimiseks.
- Suur jälgijaskond ei garanteeri, et postitus muutub viraalseks. Pigem võib väga suur võrgustik takistada viraalsust, kui jälgijate huvid on erinevad ning ainult väike osa jälgijatest reageerib postitusele.
- Kõige paremini aitavad postituse laiemale levikule ühiskonnas kaasa kommentaarid ja reaktsioonid. Seetõttu tuleks vältida selliste postituste kommenteerimist ja nendele reageerimist (sh ka negatiivse reaktsiooniga), mille levikule kaasa aidata ei soovita. Kui soovitakse jagada postitust negatiivse näitena, siis on soovitav seda teha kuvatõmmise ja mitte lingina.
- Valimis olnud postituste keskmine reaktsioonide arv oli 1387, jagamiste arv 277, kommentaaride arv 91 ning postitajate võrgustiku suurus ca 16 000 ( sõbrad ja jälgijad kokku). Keskmiselt jõudis postitus 14% inimesteni.
- Valimis olnud postituste läbistusvõime ühiskonnas oli suurem, kui võis eeldada. Selgus, et enamik sotsiaalmeedia kasutajaid tarbib infot passiivselt ning ei reageeri postitustele aktiivselt.

# Lisad

## Lisa 1. Postituste mõdikud

<i>kood</i>	<i>Kuup<span>ä</span>ev</i>	<i>S<span>õ</span>brad</i>	<i>J<span>ä</span>lgijad</i>	<i>V<span>õ</span>rgustik kokku</i>	<i>Reaktsioonid</i>	<i>Jagamised</i>	<i>Kommenta<span>ar</span>id</i>	<i>Tegevused kokku</i>	<i>Postituse l<span>ä</span>bistus</i>
1	12.10.22	0	31000	31000	2380	375	73	2828	10%
2	16.10.22	5000	21329	26329	2250	126	40	2415	27%
3	16.10.22	5000	14266	19266	1340	82	49	1471	13%
4	11.10.22	2981	6111	9092	1985	855	189	3029	11%
5	13.10.22	956	4757	5713	689	371	113	1173	10%
6	07.10.22	1279	380	1659	508	430	133	1071	8%
7	14.10.22	0	24500	24500	1160	172	213	1545	14%
8	13.10.22	0	31000	22500	1135	175	50	1360	11%
9	13.10.22	1310	0	1310	1260	146	0	1406	25%
10	13.10.22	3450	14288	17738	1160	54	45	1259	15%
<i>Keskmine</i>		<b>1998</b>	<b>14763</b>	<b>15911</b>	<b>1387</b>	<b>277</b>	<b>91</b>	<b>1756</b>	<b>14%</b>