

**TURKU**

Kaupunkitutkimusohjelma

TUTKIMUSKATSAUKSIA

2–2024

Keskustelupalstojen viestidata syventää ymmärrystä asuinalue-mielikuvien muodostumisesta

Asuinalueiden maineen ja asuntojen hinnan yhteys on intuitiivinen tosiasia. Suomi24-keskustelupalstalla on käyty asuinalueiden piirteistä värikästä keskustelua kymmenin tuhansin viestein. Tutkittaessa niiden yhteyttä kyseisten alueiden asuntojen hintoihin selvisi, että säilyttään negatiivisella keskustelulla ei ole yhteyttä asuntojen hintoihin, mutta positiivisilla viesteillä on yhteys hintojen nousuun. Keskustelupalstojen omintakeinen kielimaisema heijastelee julkisuudessa lähiöihin liitettyjä negatiivissävytteisiä keskustelunaiheita ja stereotypioita, kuten päihdeongelmia, ympäristön negatiiviseksi miellettyä esteetiikkaa, köyhyyttä ja maahanmuuton negatiiviseksi miellettyjä lieveilmiöitä.

KESKEISET HAVAINNOT

Keskustelupalstojen viestien tunneanalyysi avaa mahdollisuuksia asuinalueiden mainekehityksen ymmärtämiseen

Säilyttään negatiivinen viestintä on usein peräisin tahoilta, jotka eivät tunne aluetta, ja heijastelee mediajulkisuutta

Negatiivisella keskustelupalsta-viestinnällä ei ole yhteyttä asuinalueiden hintakehitykseen

Keskustelupalstojen viestintä ryvästyy tiettyihin tunnettuihin alueisiin kuten Turun Varissuohon ja Helsingin Kontulaan

KIRJOITTAJAT

Anssi Öörni¹,
Panu Savolainen²,
Tuukka Saarimaa³

¹ Åbo Akademi, tietojärjestelmätiede

² Aalto-yliopisto, arkkitehtuurin historia

³ Aalto-yliopisto, taloustiede

AINEISTOT JA MENETELMÄT

Suomi24-keskustelualustan viestit, asuntojen hinnat postinumeroalueittain. Tunneanalyysi, historiallinen analyysi, regressioanalyysi, laadullinen sisällönanalyysi

TEEMAT

Asuinaluemielikuvat, keskustelupalstat, tunneanalyysi, hintataso

HALUATKO TIETÄÄ...

... mitä on tunneanalyysi **s. 2**

... mikä on asuinalueita koskevan

viestinnän ja hintatason yhteys **s. 5**

... mitkä olivat Suomi24-viestien 30 yleisintä lähiöön liittyvää sanaa tai sanaliittoa **s. 8**

Asuntojen hintatason vaihtelua selittävät toisaalta taloussuhdanteet ja rakentamistalouteen liittyvät ilmiöt ja toisaalta yhteiskunnan eri tasoilla käytävä keskustelu, jossa alueille muodostuu maine ja mielikuva. Julkisessa keskustelussa esiin nousevat tunteet ovat keskeinen osa mielikuvia, jotka puolestaan vaikuttavat siihen, miten ihmiset valitsevat asuinalueensa, ja paljonko he ovat valmiita maksamaan niissä sijaitsevista asunnoista.

Asuntojen hintatason yhteyksiä esimerkiksi tulotasoon, koulutukseen, koulualuearajoihin tai sosioekonomiseen taustaan on tutkittu runsaasti kaupunkitaloustieteessä. Alueiden maineen muotoutumista ja sen yhteyttä hintatasoon on kuitenkin tutkittu vähän, sillä maineen muodostuminen on historiallinen, jopa vuosikymmenten mittainen prosessi, joka ei jätä jälkeensä helposti kvantitatiiviseen analyysiin taipuvaa dataa. Sosiaalisen median julkaisujen tunneanalyysi on kuitenkin yleistynyt viime aikoina tutkimusmenetelmänä. Kaupunkisuunnittelun tutkimukseen liittyvissä tunneanalyseissä on aiemmin käytetty otoksia esimerkiksi Facebookin ja ennen kaikkea Twitterin keskusteluista.

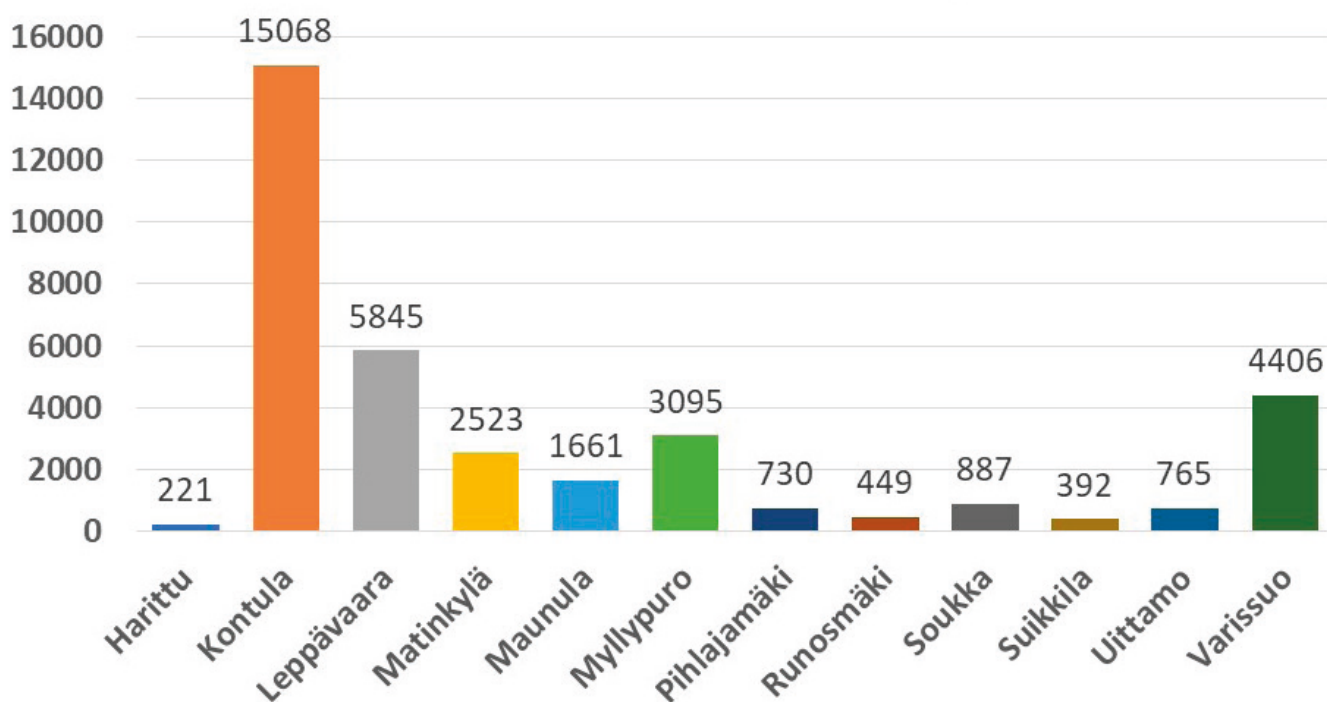
Tutkimuksen tavoite oli selvittää, voiko asuinalueiden hintatason muutoksia tarkastella yhdistelemällä asuinalueita koskevien keskustelupalstojen tunneanalyysiä ja alueiden tapahtumahistoriaa, kuten mediajulkisuutta ja perusparannushankkeita. Hanke pohjattiin poikkitieteelliselle yhteistyölle (tietojärjestelmätiede, taloustiede, arkkitehtuurin historia), jolla saavutettiin aiemmin toteuttamaton kvantitatiivisia ja kvalitatiivisia menetelmiä yhdistelevä tutkimusasetelma.

SUOMI24-KESKUSTELUPALSTAN VIESTIT AINEISTONA

Tutkimuksessa käytettiin aineistona Suomi24-keskustelufoorumille vuosina 2001–2017 kirjoitettuja viestejä, joissa mainittiin vähintään yksi tutkimuksen kannalta mielenkiintoinen asuinalue. Suomi24 oli aikanaan yksi Suomen suosituimpia keskustelualustoja, ja se sisältää pitkältä ajanjaksolta keskusteluja, jotka ovat olleet kaikkien luettavissa. Palvelulla oli enimmillään noin 1,7 miljoonaa käyttäjää.

Suomi24-aineisto on avattu tutkimuskäyttöön tiedeyhteisölle. Yli 81 miljoonaa alustan viestiä on esikäsitelty ja talletettu tietokannaksi CSC Tieteellisen laskenta Oy:n palvelimille. Aineistosta tunnistettiin 36 046 tutkimuksen kannalta oleellista viestiä.

Viestien lkm (yht. 36046)



Kuvio 1. Kahtatoista tutkimukseen valittua asuinalueetta koskevien viestien määrä aineistossa

Koneoppiminen mahdollistaa keskustelupalstan viestien tunneanalyyisin

Tunneanalyyssillä tarkoitetaan suurien tekstiaineistojen käsittelyyn kehitettyjä menetelmiä, joilla voidaan tunnistaa tekstiaineistoon sisältyvät tunnetilat. Tunneanalyyssiin käytetään yleensä koneoppimista: analyyssiin käytettävä ohjelma pyrkii löytämään tekstiaineistosta ne tekstin piirteet, jotka parhaiten erottavat eri tunnetiloja sisältävät tekstit toisistaan, ja luo tietorakenteen, jonka avulla uusia tekstejä voidaan automaattisesti luokitella mahdollisimman tarkasti teksteihin sisältyvien tunnetilojen perusteella.

Tunnetila ilmenee yleensä tekstin osista, lauseista tai yksittäisistä sanoista. Usein tekstissä on monta, joskus vastakkaistakin tunnetilaa. Tunnetilat voivat myös kohdistua useampaan tekstissä käsiteltyyn asiaan tai aiheeseen. Tällöin on mielekästä käyttää ominaisuuspohjaista tunneanalyyssiä, joka yhdistää tunnetilat kohteisiinsa. Tunneanalyyssi on tehokas menetelmä suurten tekstiaineistojen käsittelyyn. Silläkin on kuitenkin rajoituksensa. Esimerkiksi sarkastisten tekstien luokittelu on haasteellista.

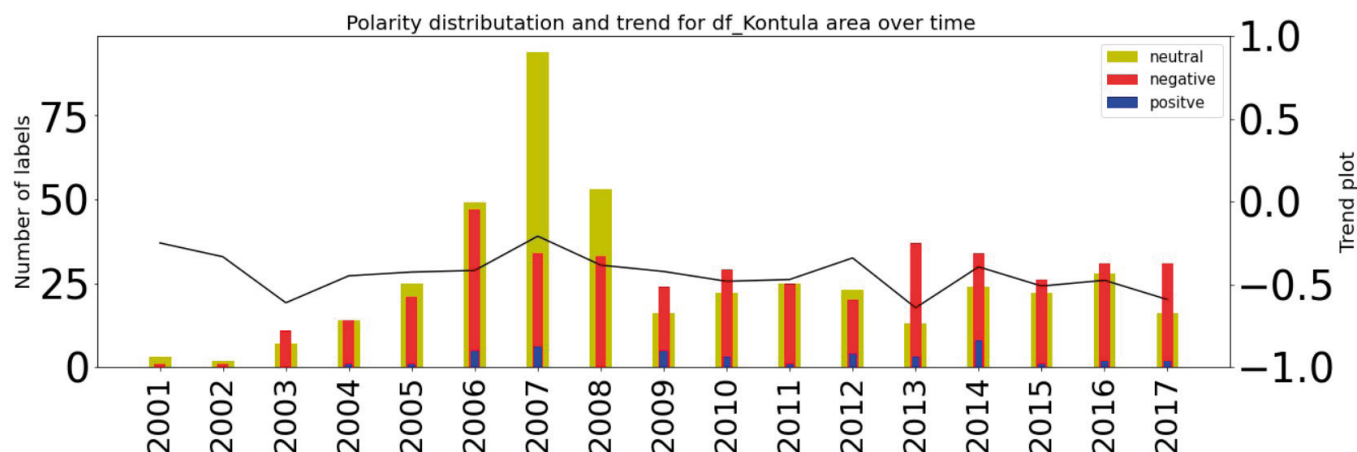
Käytimme tässä tutkimuksessa ominaisuuspohjaista tunneanalyyssiä. Jo Suomi24-viestien alustavassa tarkastelussa kävi ilmi, että usein yhdessä viestissä käsiteltiin useita mielenkiinnon kohteena olleita asuinalueita, ja samassa viestissä käsiteltyihin asuinalueisiin liittyi tyypillisesti erilaisia tunnetiloja. Tutkimusaineistosta poimittiin painotetulla satunnaisotannalla 3166 viestiä hermoverkon opettamista varten. Kaksi tutkijaa merkitsi itse näisessä viesteissä mainittuihin asuinalueisiin liittyvän polaarisen tunnetilan (myönteinen – neutraali – kielteinen). Myös viesteissä käytettyjen 33 248 sanan polaarinen tunnetila merkittiin.

Opetettu hermoverkko tunnisti lopulta 83 prosenttia opetukseen käytettyjen viestien tunnetiloista oikein. Koko tietojoukon viesteihin sisältyvistä tunnetiloista hermoverkko tunnisti oikein 72 prosenttia. Tätä ennustetarkkuutta voi pitää kohtuullisena, kun ottaa huomioon aineiston heterogeenisyyden ja aineiston arviointiin tyypillisesti liittyvän epätarkkuuden.

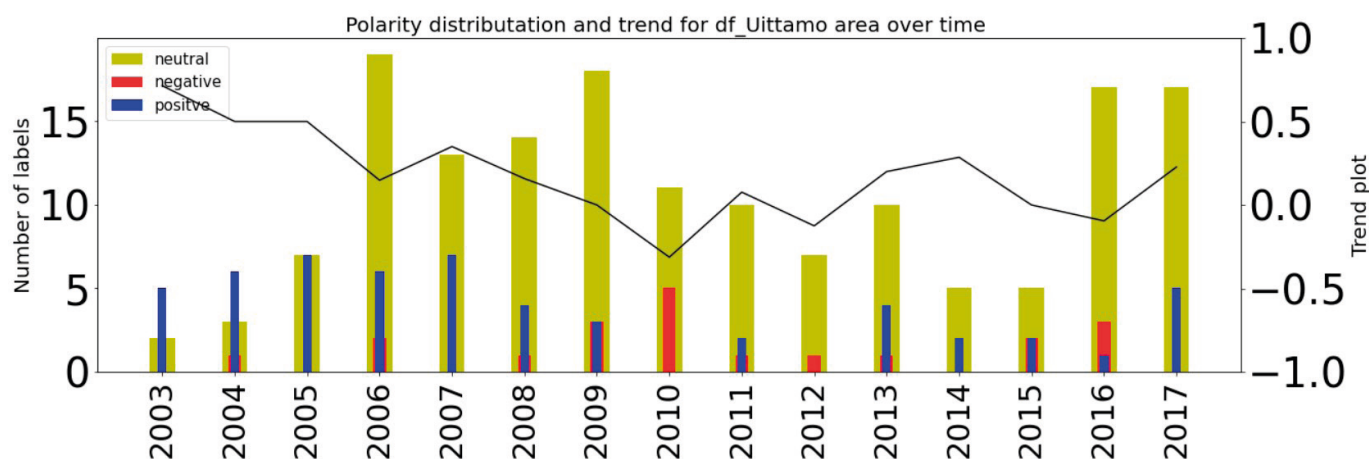
Keskustelupalstojen lähiöitä koskevasta viestinnästä (kuviot 2–4) paljastuu useita kiinnostavia piirteitä. Ensinnäkin viestimäärässä näkyy josain määrin sosiaalisen median läpimurron vaikutus 2007 alkaen. Huomatavan paljon keskustelupalstoilla esiintyvien Varissuon ja Kontulan kohdalla korostuu negatiivissävytteinen viestintä, joskin vuosittainen vaihtelu neutraalien ja negatiivisten viestien välillä on yllättävän suurta. Keskiluokkaisen ja melko hyvämaineisen Uittamon kohdalla viestejä on vähän ja ne ovat lähes poikkeuksetta positiivisia tai neutraaleja, ja vuoden 2010 negatiivisten viestien kertymä selittyi maankäyttöhankkeen aiheuttamilla reaktioilla.

TUNNEANALYYSI

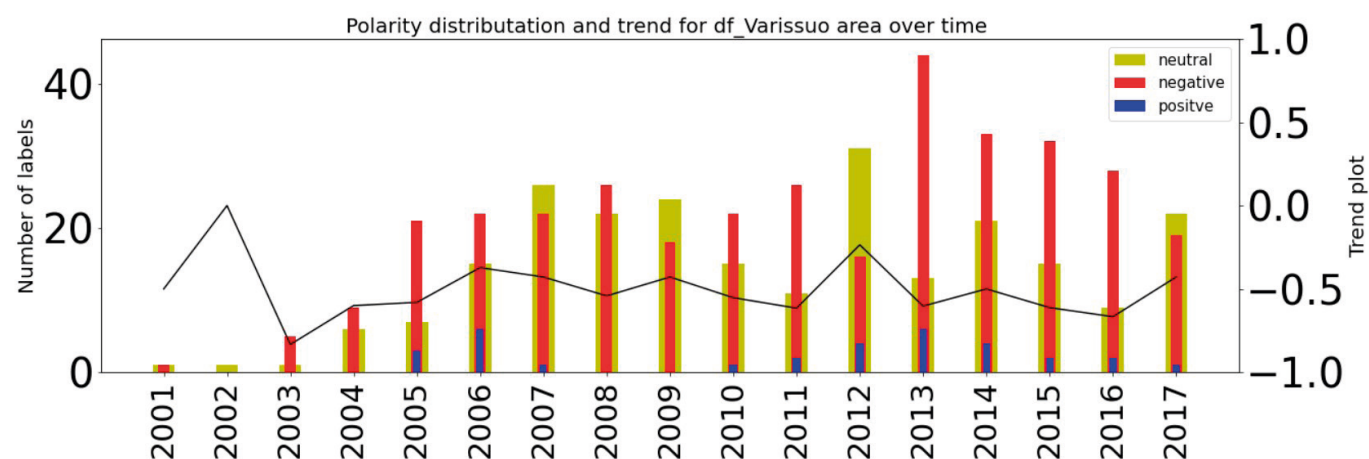
Yksi tekstianalyyssin muoto, jonka tavoitteena on tunnistaa tekstin piirteet, jotka parhaiten erottelevat eri tunnetilaa ilmaisevia tekstejä toisistaan. Erotteleva piirre voi olla esimerkiksi sana, lauseen osa tai kokonainen lause. Tyypillisesti erottelevia piirteitä on isossa tekstimassassa paljon. Tunneanalyyssissä käytetään yleensä ohjattua oppimista, jossa käytettyyn ohjelmaan syötetään tekstiaineiston osajoukko, jonka sisältämät tunnetilat on etukäteen merkitty. Ohjelma tunnistaa itse tunnetiloja erottavat tekstin piirteet, ja muodostaa niistä tietorakenteen, usein hermoverkon, jota voi käyttää koko aineiston tekstien automaattiseen luokitteluun.



Kuvio 2. Kontulaa (Helsinki) koskevien viestien tunnetilojen polariteettijakauma



Kuvio 3. Uittamoa (Turku) koskevien viestien polariteettijakauma

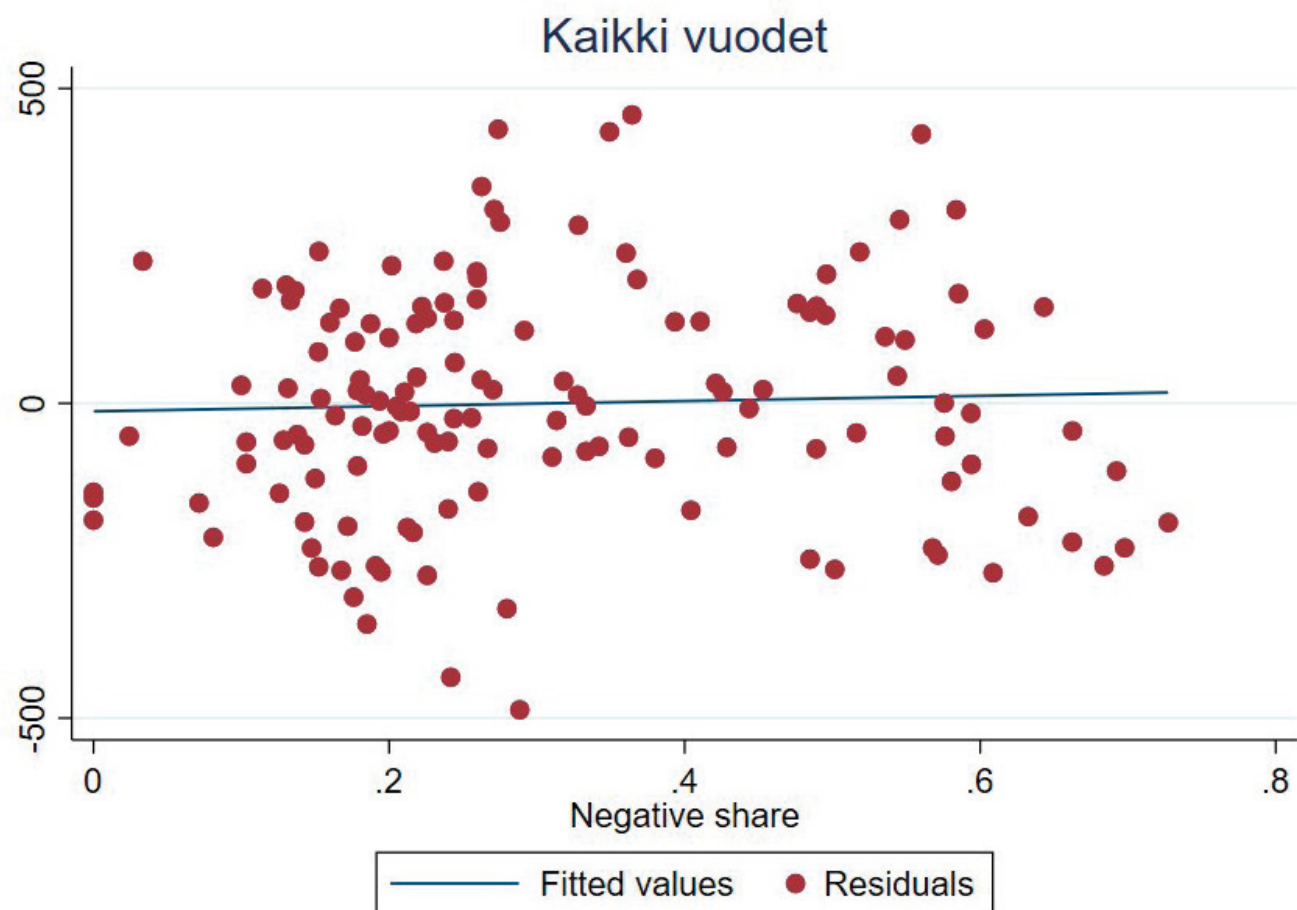


Kuvio 4. Varissuota (Turku) koskevien viestien polariteettijakauma

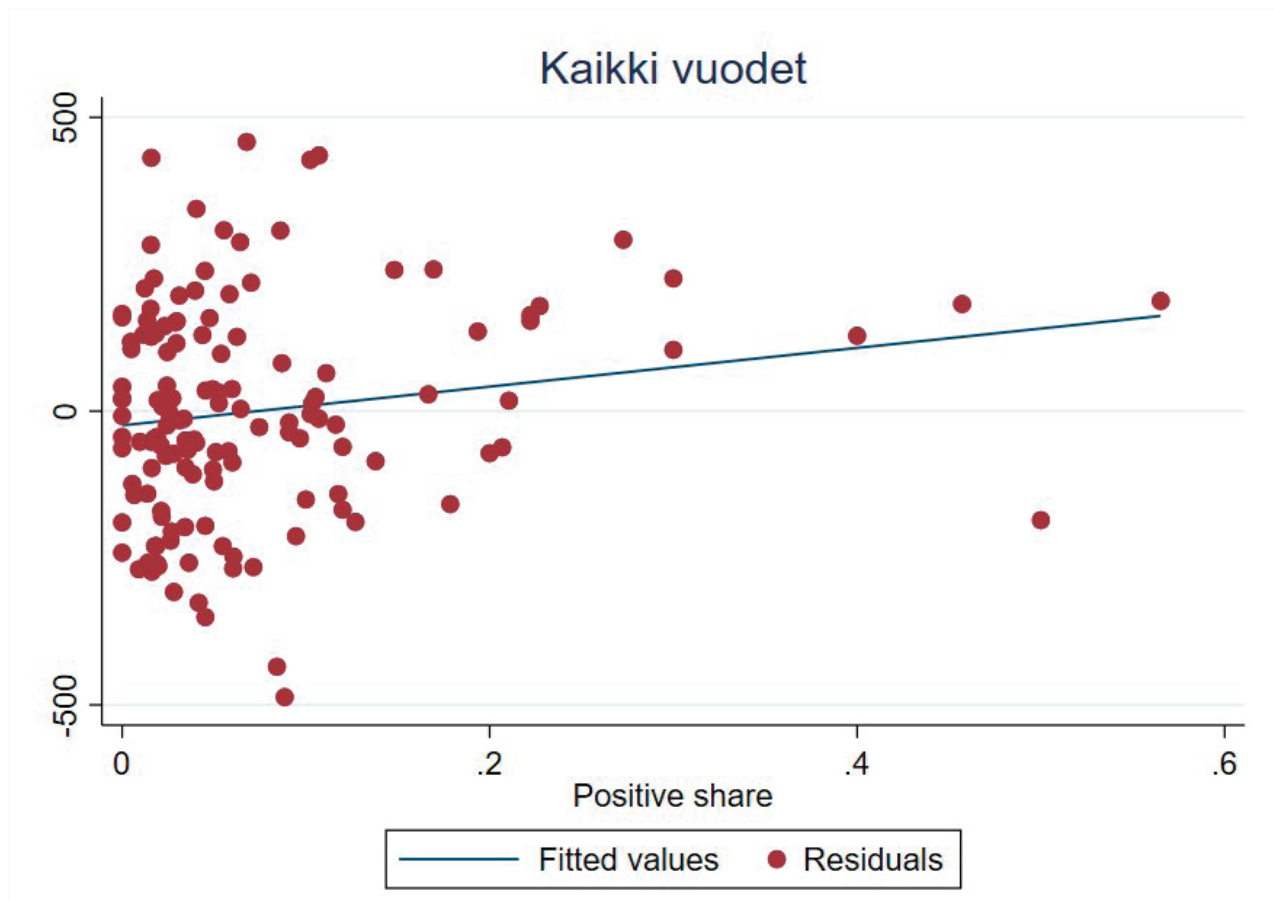
Viestien tunnetilojen yhteydet hintatasoon

Tunneanalyysin tuloksia tarkasteltiin tilastokeskuksen postinumeroalueen vanhojen kerrostalojen keskihintoihin nähden. Menetelmänä käytettiin regressioanalyysiä, ja hintoja koskeva aineisto vakioitiin huomioiden kaikkiin alueisiin ajassa vaikuttavat samanlaiset tekijät kuten suhdannevaihtelut tai korkotaso. Lisäksi huomioitiin eri alueisiin vaikuttavat vakiona pysyvät tekijät, kuten raideliikenne. Tuloksia tarkasteltiin seuraavan vuoden asuntojen hintakehitystä vasten.

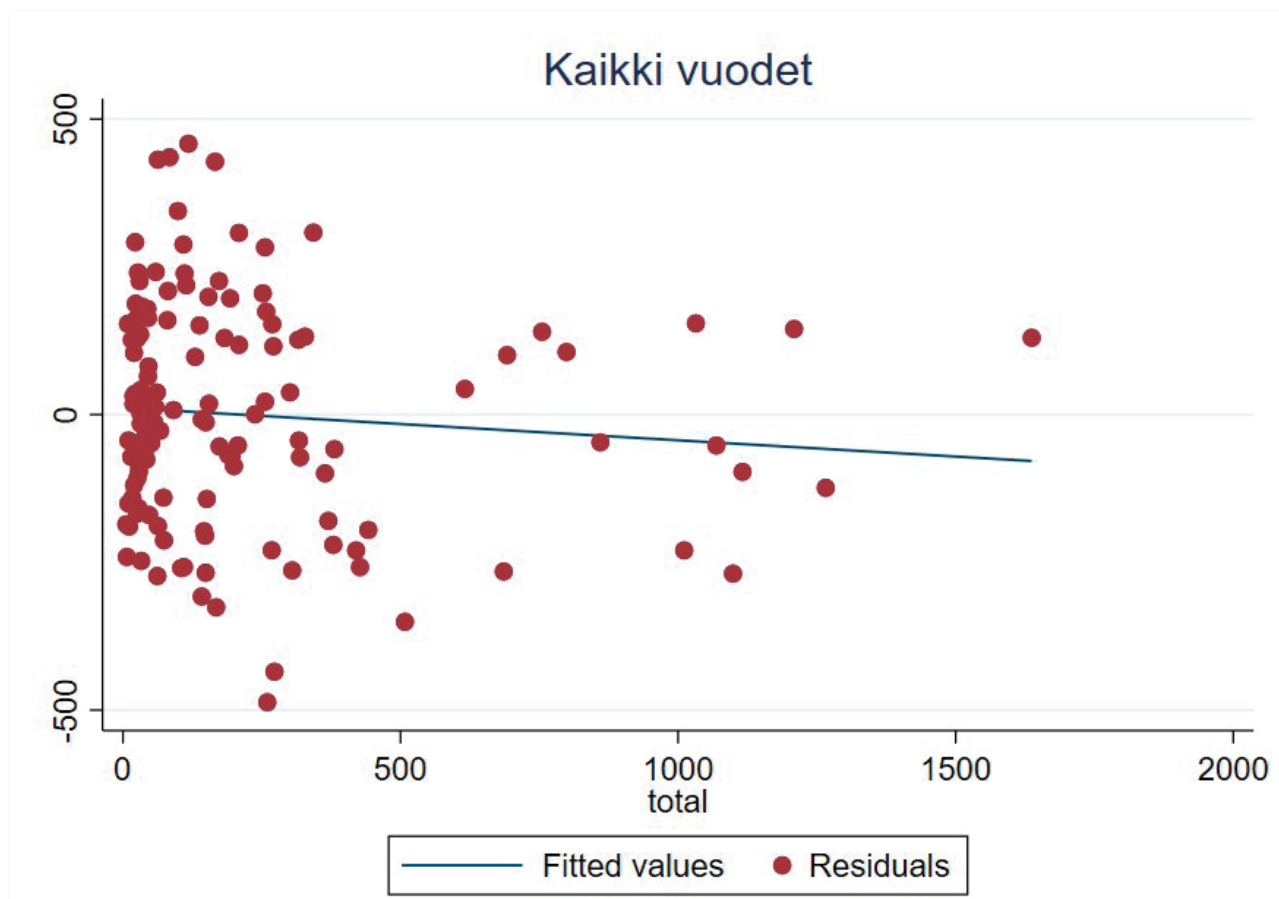
Kuvioissa 5–7 on esitettyä jokaisen asuinalueen osalta vuosittain vuoden sisällä tapahtuneen hintakehityksen ja aiemman vuoden viestien tunneanalyysin (negatiivinen, neutraali, positiivinen) korrelaatio. Kuvioista näkyy viestien määrän ryvästymisen tiettyihin lähiöihin. Huolimatta suurista eroista eri lähiöitä koskevien viestien määrässä mielikuvien ja hintatason yhteydessä, näkyy asuntojen hintojen ja positiivisten viestien osuuden välinen tilastollinen yhteys yhden vuoden aikaskaalalla.



Kuvio 5. Asuntojen hintojen ja negatiivisten viestien osuuden välinen tilastollinen yhteys



Kuvio 6. Asuntojen hintojen ja positiivisten viestien osuuden välinen tilastollinen yhteys



Kuvio 7. Asuntojen hintojen ja viestien kokonaismäärän välinen tilastollinen yhteys

TULOKSET

1. Negatiivisten mainintojen osuudella ei ole tilastollista yhteyttä alueen neliöhintoihin
2. Positiivisten mainintojen osuudella on positiivinen yhteys alueen neliöhintoihin
3. Viestien kokonaismäärällä on heikko negatiivinen yhteys alueen neliöhintoihin

On korostettava, että tutkimuksessa havaitut korrelaatiot eivät todennäköisesti ole kausaalivaikutuksia. Positiivisten viestin osuus ei siis aiheuta hintojen nousua. Hintojen nousu voi aiheuttaa positiivisen keskustelun tai molempia voi ajaa jokin yhteinen havaitsematon tekijä, kuten alueen palveluiden tai saavutettavuuden paraneminen. Esimerkiksi ensin alueen hinnat nousevat ja sitten syitä spekuloidaan keskustelussa positiiviseen sävyyn. On kuitenkin huomionarvoista, että negatiivissävytteisellä puheella ei tutkimuksen perusteella ole vaikutusta hintatasoon.

Negatiivissävytteiset viestit heijastelevat julkista keskustelua

Keskustelupalstadatan laadulliseen sisällönanalyysiin valittiin Turun Varisuo ja Helsingin Kontula. Tunneanalyyseissä havaittua vuosittaista negatiivisten ja positiivisten viestien välistä vaihtelua (kuviot 2-4), joka koski erityisesti Helsingin Kontulaa, ei pystytty selittämään alueen mediajulkisuuden muutoksilla tai lähiössä sattuneilla positiivisilla tai negatiivisilla tapahtumilla. Kummankin asuinalueen kohdalla oli huomionarvoista, että fyysistä ympäristöä käsiteltiin erittäin harvoissa viesteissä. Lähtöoletusten vastaisesti lähiöt rakennettuna ympäristönä tai lähiöiden luonto lähes loistivat poissaolollaan viesteistä, jopa "rumuus" oli esillä melko harvoin, pois lukien käsitteeseen "betonilähiö" sisällytettynä. Kummankin asuinalueen kohdalla keskustelussa korostui leimaava puhe sekä alueiden sosiaalisen tai sosioekonomisen statuksen käsittely. Suomi24 oli näin ollen eräänlainen mediajulkisuuden peili tiettyjen alueiden ja lähiöiden maineenrakentamiseen.

SANA	MÄÄRÄ
lähiö	33080
nukkumalähiö	1392
betoni lähiö	1373
lähiö pubi	1270
lähiö baari	1230
lähiö kuppila	846
lähiö kapakka	827
kerros talo lähiö	743
asuma lähiö	653
lähiö räkö	565
ongelma lähiö	525
vuokra lähiö	342
lähiö valas	341
lähiö alue	339
asuin lähiö	330
maahanmuuttaja lähiö	310
naapuri lähiö	297
kaupunki lähiö	242
lähiö itä	229
lähiö lehmä	213
lähiö elämä	180
lähiö kerros talo	170
mamulähiöissä	169
slummi lähiö	168
ghetto lähiö	131
lähiö kauppa	123
lähiö asukas	120
lähiö ravintola	118
koti lähiö	113
paska lähiö	112

Taulukko 1. Aineiston 30 yleisintä lähiöön liittyvää sanaa tai sanaliittoa

Lisäksi merkittävä havainto oli, että negatiivissävytteiset viestit painoutuivat tahoille, jotka olivat muodostaneet kuvansa alueesta julkisen maineen pohjalta. Lähiöitä koskeva 30 yleisimmän sanan tai sanaliiton luettelo heijastelee vanhasti yhteiskunnassa esillä olevia negatiivissävytteisiä lähiöihin liittyviä keskustelunaiheita, kuten päihdeongelmia, ympäristön negatiiviseksi miellettyä estetiikkaa, köyhyyttä ja maahanmuuton negatiiviseksi miellettyjä lieveilmiöitä. Alueiden asukkaat osallistuivat keskusteluun vähän, ja tavallisesti asuinalueitaan puolustavalla ja positiivisia puolia esiin nostavilla viesteillä.

Poikkitieteellinen hanke osoitti, että kuvattu tutkimusasetelma voisi toimia laajemmilla ja toisenlaisilla aineistoilla, kuten sanomalehtiaineistoilla. Suomi24-keskustelupalstan luonne ja sisältö osoittautui kuitenkin haastavaksi tutkimusasetelman tarkoituksiin, sillä ympäristöä kommentoitiin viesteissä hyvin vähän, lukuun ottamatta betonin ja teollisen rakennustuotannon rumuuteen viittaavia vertauskuvallisia ilmaisuja.

**Asukkaiden viestit
olivat alueita
puolustavia**

LUE LISÄÄ

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). "More than a Feeling: Accuracy and Application of Sentiment Analysis". *International Journal of Research in Marketing* **40**(1): 75-87.

Kovacs-Gyori, A., Ristea, A., Kolcsár, R., Resch, B., Crivellari, A., & Blaschke, T. (2018). "Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data". *International Journal of Geo-Information* **7**: 378.

Resch, B., Summa, A., Zeile, P., & Strube, M. (2016). "Citizen-centric urban planning through extracting emotion information from Twitter in an interdisciplinary space-time linguistic algorithm". *Urban Planning* **1**(2), 114–127.

Roberts, H., Resch, B., Sadler, J., Chapman, L., Petutschnig, A., & Zimmer, S. (2018). "Investigating the Emotional Responses of Individuals to Urban Green Space Using Twitter Data: A Critical Comparison of Three Different Methods of Sentiment Analysis". *Urban Planning* **3**(1): 21.

Savolainen, P., Berger, L., Hautamäki, R., Kanner, A., Kouri, K., Lähde, E., Piirainen, P., Ruoppila, S., Taipale, A., Teräsvuori, H., Tolonen, A., Turtiainen, P., Verma, I. (2023). *Keskustelunavaus ja politiikkasuositukset: Lähiöpride. Lähiöympäristöt ja arkkitehtuuri identiteettien raaka-aineena*. Aalto-yliopisto.

Sim, J., Miller, P., & Swarup, S. (2020). "Tweeting the High Line Life: A Social Media Lens on Urban Green Spaces". *Sustainability* **12**(21).

NÄIN TUTKIMUS TEHTIIN:

Tutkimuksessa käytettiin aineistona Suomi24-keskustelufoorumille vuosina 2001–2017 kirjoitettuja viestejä, joissa mainittiin vähintään yksi tutkimuksen kannalta mielenkiintoinen asuinalue. Alueita koskevat noin 30 tuhatta viestiä analysoitiin asuinalueisiin liitetyn polaari-sen tunnetilan (myönteinen – neutraali – kielteinen) pohjalta opetulla hermoverkolla. Asuinalueisiin liittyvien tunnetilojen muutoksia verrattiin tämän jälkeen asuinalueiden vuosittaiseen hintatasomuutokseen. Lopuksi tarkasteltiin, minkälaiset laadulliset tekijät esimerkiksi alueiden tapahtumahistoriassa voisivat selittää kvalitatiivisessa analyysissä havaittuja mielikuva- ja hintatasomuutoksia. Tutkimuksen rahoitti Länsi-Suomen yleishyödyllinen asuntosäätiö osana Turun kaupunkitutkimusohjelmaa.

OTA YHTEYTTÄ:

Anssi Öörni

Tietojärjestelmätieteen professori
Åbo akademi
anssi.oorni@abo.fi

Tuukka Saarimaa

Kaupunkitaloustieteen professori
Aalto-yliopisto
tuukka.saarimaa@aalto.fi

Panu Savolainen

Arkkitehtuurin historian ja
restauroinnin professori
Aalto-yliopisto
panu.savolainen@aalto.fi

Tutkimuskatsauksia-sarjassa

julkaistaan tiiviisti Turun kaupunkitutkimusohjelmaan kuuluvien tutkimusten tuloksia. Kirjoittajat ovat tutkijoita.

TOIMITTAJA

Sampo Ruoppila, tutkimusjohtaja
sampo.ruoppila@turku.fi

JULKAISIJA

Turun kaupungin konsernihallinto
PL 355 (Yliopistonkatu 27a), 20101 Turku
www.turku.fi/kaupunkitutkimus