

Ulkoiset datalähteet analytiikassa

© ADVIAN OY

Kiitokset oppaan lataamisesta!

Mistä löytää omalle toiminnalle relevantit avoimet ja kaupalliset datalähteet? Kuinka ulkoinen data tuodaan osaksi nykyistä analytiikkaa ja prosesseja? Entä mitä haasteita ulkoisen datan hyödyntämiseen liittyy?

Vastaukset muun muassa näihin kysymyksiin selviävät tämän oppaan avulla. Oppaasta löydät myös linkkejä erityyppisiin ilmaiseksi saatavilla oleviin avoimen datan lähteisiin.

Toivotamme mukavia lukuhetkiä oppaan parissa!



Vinkki: voit käyttää seuraavan sivun sisällysluetteloä siirtyäksesi helposti haluamaasi kohtaan oppaassa.

/ Sisälllys

1. Mitä on ulkoinen data?

2. Esimerkkejä ulkoisista datalähteistä

3. Miten kartoittaa ulkoisen datan lähteitä?

4. Miten ulkoinen data yhdistetään muuhun dataan?

5. Miten sijaintianalytiikka toimii?

6. Miten pääsen alkuun?

7. Esimerkkejä käyttötapauksista

8. Mitä haasteita ulkoiseen dataan liittyy?

Mitä on ulkoinen data?

Ulkoisessa datassa piilee yksi tiedolla johtamisen suurimmista tutkimattomista kultakaivoksista. Ulkoinen data voidaan määritellä kaikeksi dataksi, joka tuotetaan ja hankitaan oman organisaation ulkopuolelta. Ulkoinen data on myös:

- ✓ Dataa, jota sinulla ei ole heti valmiina saatavilla
- ✓ Dataa, jota säilytetään muualla kuin nykyisissä tietovarastoissasi
- ✓ Dataa, jota et voi hyödyntää heti jokapäiväisistä datalähteistäsi
 - ✓ Dataa, jota lataat tai etsit omaan käyttöösi
 - ✓ Dataa, joka ei ole juuri nyt käytössä

Ulkoiset datalähteet voidaan jaotella useaan eri kategoriaan. Seuraavan sivun taulukossa on lueteltu esimerkkejä eri ulkoisen datan tyypeistä. Olemme keränneet taulukkoon myös linkkejä, joiden takaa löydät esimerkkejä eri tyyppisistä avoimista datalähteistä.

Avoimen datan lisäksi on saatavilla paljon dataa maksullisista kaupallisista lähteistä. Esimerkiksi yritysdataa löytyy usein maksullisista lähteistä ja palveluista (kuten yritystietokanta Finder tai Vainu.io).



LUONNOLLINEN YMPÄRISTÖ



RAKENNETTU YMPÄRISTÖ



HALLINTO



YHTEISÖT



LIIKENNEVIRRAT



STATISTIikka



KUVADATA



SENSORIDATA



Esimerkkejä ulkoisista datalähteistä

Kategoria	Alakategoria ja esimerkkitiedot	Esimerkkejä käyttötapauksista
Luonnollinen ympäristö	<u>Säädata</u> <u>Maankäyttö</u> <u>Korkeuserot</u> <u>Metsä & Luonnonvarat</u>	Tulvariskin mallintaminen Myrskyriskin mallintaminen Valuma-alueanalyysi Ennakoiva kunnossapito
Rakennettu ympäristö	<u>Kiinteistöt</u> <u>Rakennukset</u> <u>Verkot</u> – kuten tiestö, raideverkosto, sähkö-, vesi- ja kaukolämpöverkot	Markkina-analyysit, kiinteistöjen arvonmääritys Toimipaikkasuunnittelu, julkiset palvelut Reititykset, Verkon suorituskyvyn analyysi ja optimointi
Hallinto	<u>Maantieteelliset rajat</u> (postinumerot) <u>Rahoituksen käyttö</u>	Datojen yhdistäminen Toimipaikkasuunnittelu Julkisen tarjoaman kohdentaminen
Yhteisöt	<u>Ihmiset</u> <u>Yritykset</u> Ajoneuvot	Hyperkohdennettu markkinointi Markkinamallit & segmentointi Arvonmääritykset
Liikennevirrat	Ihmiset <u>Tiestö ja raideverkko</u> <u>Rahti- ja laivaliikenne</u> <u>Lentoliikenne</u>	Toimipaikka- ja palvelusuunnittelu Infran kehittäminen ja ylläpito Liikenne-ennusteet Ulkomarkkinointi
Statistiikka	<u>Demografia</u> <u>Talousluvut</u>	Laajan mittakaavan mallinnukset Olemassa olevien rekisterien rikastaminen
Kuvadata	<u>Kaukokartoitus (satelliitit jne.)</u> Valokuvat / Video	Ominaisuuksien tunnistuksen algoritmit Maanliikkeiden analyysi Ihmisvirtojen analyysi
Sensoridata	Ajoneuvot IoT-laitteet	Reaaliaikainen prosessivalvonta Turvallisuus ja valvonta



Miten kartoittaa ulkoisen datan lähteitä?

Kiitos viime vuosina heräilleen ja yhä voimakkaammaksi kasvaneen Avoin data -liikkeen, aineistoja löytyy yhä helpommin ja yhä useammin myös täysin ilmaisena hyödynnettäväksi.

Kuten edellisen sivun taulukosta varmasti jo huomasi, ulkoinen data voidaan jakaa useaan eri kategoriaan ja saatavilla on hyvin paljon erilaista dataa. Mitä kaikkea sitten on? On meitä ympäröivä luonnollinen ympäristö metsineen, vesistöineen, korkeuseroineen ja säätiloineen sekä toisaalta meidän luomamme rakennettu ympäristö kiinteistöineen, rakennuksineen ja tiestöineen.

Yllä mainituissa ympäristöissä liikkuu virtoja: vettä, ajoneuvoliikennettä ja tietoliikennettä. Virroissa taas kulkee useita liikkuvia ja ahkeroivia toimijoita: ihmisiä, yrityksiä ja ajoneuvoja. Tästä kaikesta on saatavilla kokonaisvaltaista kuvaa luovaa kaukokartoitusaineistoa satelliiteista ja lentokoneista. Maan tasolla taas löytyy tätä kaikkea konemuodossa jatkuvasti mittaavaa laitelähtöistä tietoa, niin sanottua sensoridataa. Lisäksi saatavilla on tietoa tätä koko näytöstä ohjaavien tahojen eli hallinnon julkaisemien aineistojen muodossa. Lopulta löytyy koko kokonaisuuden yhteen sitovaa, valmiiksi pureskeltua ja niin sanottua perinteistä, rakenteellista tilastoaineistoa.

**Kuinka selvittää, millainen data on juuri minulle tarpeellista?
Datalähteitä kartoittaessa kannattaa pohtia muun muassa:**

- ✓ Kuinka saavutettavaa himoitsemani data on? Onko se maksumuurin tai rekisteröitymisen takana, vaiko täysin vapaata riistaa?
- ✓ Millä resoluutiolla, ajallisella tai alueellisella, tarvitsen tietoa?
- ✓ Pitääkö minun tietää, kuinka monta ihmistä tietyn liikkeen sisäänkäynnillä kulkee tuntitasolla, vai riittäisikö esimerkiksi vuosittain päivittyvä postinumeroaluetason tilastotieto alueen väestöpohjasta?

Miten ulkoinen data yhdistetään muuhun dataan?

Miten ulkoinen data tuodaan osaksi olemassa olevia prosesseja?

Yleisin haaste soveltuvien datalähteiden löytämisen jälkeen lienee kyseisten aineistojen liittäminen olemassa oleviin varantoihin. Useita datalähteitä työstettäessä ei aina ole käytettävissä eri kokonaisuuksia yhdistävää avainta (vaikkapa nimeä tai postinumeroa), ja mitä useampaa aineistolähdettä käytetään, sitä monimutkaisemmiksi työvaiheet käyvät. Mikä siis avuksi?

Lähes kaikki data on suhteutettavissa **sijaintiin** tavalla tai toisella. Mobiilidata tarjoaa suuntaa-antavia tietoja ihmisten liikkeistä, rakennukset ja niissä sijaitsevat toimijat voi sijoittaa kartalle osoitteen perusteella ja tilastotieto taas luodaan usein jonkin alueellisen jaon pohjalta. Näin ollen sijaintitieto voi tarjota vahvan työkalun eri datalähteiden yhdistelyssä ja suhteuttamisessa toisiinsa.

Tietystä ympäristössä vallitsevan säätilan voi esimerkiksi tuoda reittisuunnittelun tueksi läheisimmän sääaseman koordinaattitietojen perusteella ja toimitilan X läheisyydessä liikkuvia kuluttajavirtoja voi kartoittaa vertaamalla toimipaikan osoitetietoa ja mobiilidatasta irrotettuja koordinaattipareja.

Lähes kaikkeen dataan voidaan liittää sijainti



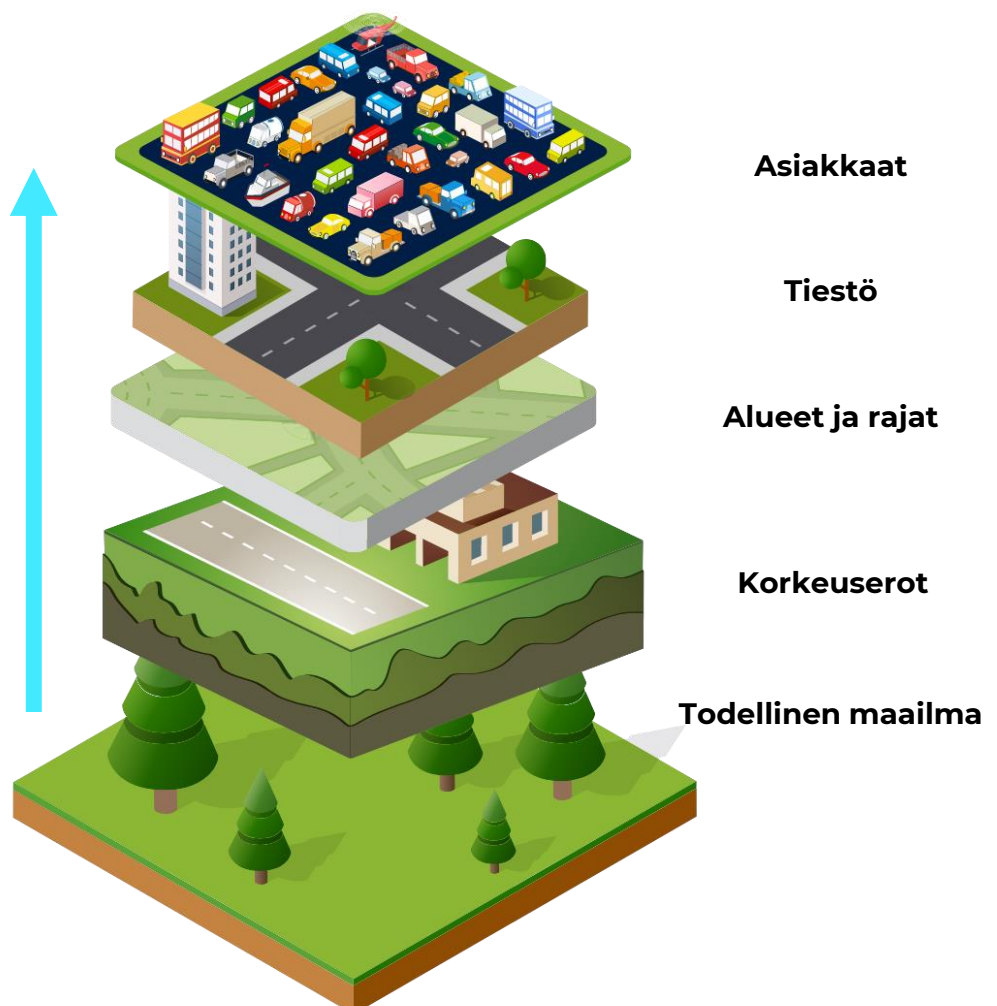
- ✓ Ihmisten sijainti selviää puhelimen paikannuksen avulla
- ✓ Rakennusten sijainti osoitteen perusteella
- ✓ Valokuvien ottopaikka selviää metadatatista
- ✓ Ajoneuvon kulkureitti saadaan selville seuraamalla sijainnin koordinaatteja eri ajan hetkillä

Sijaintianalytiikka: Miten datat yhdistetään?



Sijaintianalytiikka toimii datasiilojen yhdistäjänä. Jos kaksi erillistä datasettiä halutaan yhdistää tai olemassa olevaa dataa rikastaa ulkoisella datalla, on ratkaisu tähän sijaintitieto. Sijaintitieto on teippi, joka liimaa yrityksen oman datan ulkoiseen dataan. Sijaintitietoja vertailemalla kaksi erillistä tietolähdettä voidaan yhdistää myös ilman yhteisiä muuttujia. [Lue lisää täältä.](#)

Sekä nykyisellä datalla että ulkoisen datan kerroksilla on sijaintiulottuvuus. Sijaintianalytiikkaa hyödyntämällä olemassa olevaa informaatiota voidaan rikastaa ulkoisen datan kerroksilla.





Miten pääsen alkuun?

Ulkoisen datan hyödyntämisessä kannattaa lähteä liikkeelle käyttötapausten määrittelystä ja tietoaukkojen tunnistamisesta.

Määrittele käyttötapaus ja -tarve

- ✓ Mikä on se kysymys, johon kaipaam vastauksia?
- ✓ Mitkä toimintamallit tuntuvat tällä hetkellä tehottomilta?
- ✓ Ideointi ja uuden innovointi: toimisiko malli A paremmin, jos tietäisin enemmän asiasta B?

Kannattaa siis kuluttaa hetki kartoittaen ulkoisen datan lähteitä ja seuraten pienempiäkin ideantynkiä. Näin saatujen ajatusten ja nykyisten datalähteiden yhdistelystä datan tarve saattaa kirkastua nopeastikin. Täten oikeiden lähteiden rajaaminen omien valmiuksien ja käyttötarkoitusten mukaan alkaakin olla maustamista vaille valmista!

Selvitä tarpeet datan suhteen

- ✓ Määritä, millaiset tarpeet sinulla on datan suhteen: mitä sisäistä dataa löytyy, millaista ulkoista dataa tarvitsen lisäksi?
- ✓ Minkä tyyppistä dataa tarvitsen?

Valitse oikeat datasetit

- ✓ Paras tulos saavutetaan yleensä käyttämällä useita lähteitä, ei pelkästään yhtä lähdeä
- ✓ Mitä työkaluja minulla on käytössäni, mitä pystyn tekemään itse?
- ✓ Mitä datalla on tarkoitus tehdä?

Aina tämä ei kuitenkaan aivan helppoa. Kaikkea ei tarvitse eikä kannatakaan tehdä itse. Tukeutuminen asiantuntijoiden apuun voi helpottaa ja nopeuttaa omaa työtaakkaa huomattavasti. **Tutustu tästä [Location Intelligence Innovation -työpajaamme](#)**. Työpajan tuloksena saat konkreettisen ja perustellun näkemyksen hedelmällisimmistä sijaintianalytiikkaan liittyvistä kehityskohteista.

Ulkoisen datan käyttötapauksia



ESIMERKKI 1

Lähes reaaliaikainen metsätuhojen arviointi myrskyn jälkeen

Datalähteet:

- ✓ Kaupallinen satelliittidata (tulkinta koneoppimisen keinoin)
- ✓ Avoin metsädata (ks. [Luke](#))
- ✓ Avoin säädata
- ✓ Avoin data asuinalueista
- ✓ Avoin data kiinteistöjen omistuksista



ESIMERKKI 2

Lähes reaaliaikainen asiakasvirtojen analyysi kaupan sisällä ja sen ympäristössä

Datalähteet:

- ✓ Videoanalytiikan avulla liikkeen sisäänkäynneiltä kerätty data, jota analysoidaan koneoppimisen menetelmin: datasta voidaan tunnistaa asiakkaiden sukupuoli, ikä ja tunnetila. Yksilöitä ei tällaisesta datasta voida tunnistaa, joten ratkaisu on täysin GDPR-säännösten mukainen
- ✓ Kaupallinen mobiililaitteiden sijaintidata paljastaa liikkeet kaupan ulkopuolella
- ✓ Vertailemalla ylläolevia dataa voidaan mitata myös myyntikonversioita



ESIMERKKI 3

Toimipaikkasuunnittelu

Datalähteet:

- ✓ Sisäinen data samankaltaisten toimipaikkojen kävijöistä/asiakkaista
- ✓ Kaupallinen data menneistä ja nykyisistä liikennevirroista kohdesijaintien ympäristössä
- ✓ Avoin demografiatieto ja väestödata kohdesijaintien läheisyydessä

Mitä haasteita ulkoiseen dataan liittyy?

Ulkoisen data tarjoaa huikkeitä mahdollisuuksia omien datalähteiden rikastamiseen sekä analytiikan tehostamiseen. Miksi ulkoista dataa hyödynnetään silti edelleen vähän?

Usein ulkoisen datan hyödyntämiseen tarvitaan sijaintianalytiikkaa – ja sen osaamista. Vaikka organisaatiossa tunnistettaisiinkin ulkoisen datan arvo, on huippuosajia vaikea löytää. Sijaintianalytiikka on kokonaan oma tieteenalansa ja kokonaisuutensa, joka vaatii muun muassa kaukokartoituksen, 3D-mallintamisen, geoinformatiikan ja maantieteen sekä menetelmällistä että teknistä osaamista.

Datan saatavuus ja laatu vaihtelee

Suomessa ulkoisen datan saatavuus ja laatu sekä avoimen että kaupallisen datan osalta on tällä hetkellä suhteellisen hyvä. Kansainvälisellä tasolla tarkasteltuna data on kuitenkin hajanaista ja puutteellista tai siihen on vaikea päästä käsiksi. Poikkeuksena voidaan pitää kaukokartoitusaineistoja sekä perinteistä tilastollista dataa (YK, Maailmanpankki), jota on saatavilla hyvin myös kansainvälisessä laajuudessa.

Ota meihin yhteyttä



CHRIS WINQUIST

Chief Commercial Officer
christoffer.winqvist@advian.fi

Kiinnostaako ulkoinen data? Anna meidän auttaa! Kysy lisää ja varaa aika Chrisin kalenterista [täältä](#):

Konsultointi, innovointi ja suunnittelu
Sopivien datalähteiden kartoitus ja etsiminen
Datalähteiden yhdistäminen & rikastaminen
Ulkoisen datan prosessoinnin automaatiot
Ulkoiseen dataan perustuvat edistyneet analytiikkamallit
Analyysit & automaatiot

www.advian.fi