

Loppuraportti

Vietin vuoden Columbian yliopistossa New Yorkissa ja valmistuin datajournalismin maisteriksi toukokuussa 2019. Kahdessatoista kuukaudessa opin koodaamaan Python-ohjelmointikielellä, keräämään ja analysoimaan isoja tietomääriä sekä visualisoimaan dataa ja kirjoittamaan pitkiä uutisjuttuja dataan perustuen.

Tässä lyhyt yleiskatsaus opintojen rakenteeseen:

Kesä 2018:

- 1) Data Foundations: Python-kielen perusteet, Pandas-kirjaston käyttö, skreippaus, command linen käyttö
- 2) Databases: Opettelimme SQL tietokantoja, skreippausta, teimme oman projektin.
- 3) Data Studio: Joka viikko uusi datan analysointi ja visualisointiprojekti.
- 4) Reporting: Toimittajan työn perusteiden harjoittelua. Aiheena oli New Yorkin oikeusistuimet, joista kirjoitimme päivittäisuutisia ja pitempiä juttuja. Kurssi oli kiinnostava, mutta ei ammatillisesti kovin hyödyllinen jos on jo tehnyt toimittajan töitä. Siksi päätinkin keskittyä enemmän koodausta käsitteleviin kursseihin.

Syksy 2018:

- 1) Data & Computation and Innovation. Jatko-osa kesän Foundations-kurssiin. Opimme lisää datan visualisoinnista JavaScript- ja D3-koodikielten avulla. Teimme ns. scrollytelling -tarinoita ja interaktiivisia grafiikoita.
- 2) Writing with Data. Kirjoitusharjoituksia tilastoista ja numeroista. Tehtävänä esimerkiksi kirjoittaa juttu vapaavalintaisesta tilastosta tai tutkimuksesta mutta niin, että jutussa ei käytetä isompaa numeroa kuin kymmenen.
- 3) Invesigative Techniques with Data. Tutustui erilaisiin tietokantoihin ja tietolähteisiin. Eniten yhdysvaltalaisia lähteitä, mutta myös joitakin kansainvälisiä.
- 4) Law. Kävimme läpi yhdysvaltalaisia mediaa koskevia lakeja. Hyödyllistä tietoa lopputyötä tehdessä, joskin datajournalismia koskevia lakeja olisi voinut käsitellä tarkemmin.
- 5) Ethics. Vierailevat toimittajat keskustelivat kohtaamistaan eettisistä dilemmoista.
- 6) Master's project. Lopputyö ei ollut akateeminen tutkielma vaan pitkä, 5000-sanainen juttu dataelementteineen. Minä päätin tutkia muutaman eri ruokayrityksen ruokien ravintoarvoja. Lisäksi haastattelin ravitsemusasiantuntijoita ja ruokayritysten asiakkaita.

Kevät 2019:

- 1) Algorithms. Opettelimme perusteet tilastotieteestä ja algoritmeista. Keskityimme siihen, kuinka tehdä uutisia algoritmeista: mihin kiinnittää huomiota ja kuinka tutkia yritysten ja julkisen sektorin käyttämiä algoritmeja?
- 2) Computational Journalism. Valinnainen, hieman kertaava kurssi jo opittuun: pythonia, skreippausta, ryhmätöitä ja hackaton. Opettelin myös käyttämään optimisaatio-ohjelmaa.
- 3) Food Writing. Valinnainen kurssi. Keväällä piti valita kaksi kurssia, joista toinen ei saanut liittyä dataan. Valitsin ruokatoimittamisen. Kurssimme tuotoksia julkaisimme www.nytable.com.
- 4) History. Journalismin historiaa Yhdysvalloissa ja maailmalla. Oli superkiinnostavaa lukea vanhoja klassikkolehdistöjuttuja.
- 5) Master's Project. Lopputyötä tehtiin myös kevätlukukaudella. Ohjaajaa tapasin säännöllisesti noin parin viikon välein.

Vuosi oli työntäyteinen. Opin enemmän kuin uskalsin uneksiakaan mutta myös nukuin vähemmän kuin koskaan aiemmin. Vietin arkipäivät yliopistolla yhdeksästä viiteen joko luennoilla tai läksyjä tekemässä, ja iltaisin kun lapset olivat nukahtaneet, tein vielä 3-4 tuntia läksyjä joka ilta. Kaiken kaikkiaan tein töitä arviolta vähintään 60-70 tuntia viikossa.

Seuraavaksi kooste havainnoistani datajournalismin trendeistä Yhdysvalloista tällä hetkellä:

Datajournalismi Yhdysvalloissa juuri nyt

1. Scrollytelling

Scrollytelling on huippusuosittu kerrontamuoto Yhdysvalloissa juuri nyt. Scrollytelling, vapaasti suomennettuna selauskerronta, tarkoittaa nettiuutisjuttuja, joissa grafiikka tai kuvitus muuttuu kun käyttäjää selaa ruutua alaspäin. Klikkauksia tai muita toimenpiteitä ei tarvita. Perustelut tähän kuuluvat: "toimittajan kuuluu tehdä datan analysointityö ja esittää lukijalle vain kaikkein tärkein tieto. Lukijaa ei tule vaivata liiallisilla vaihtoehdoilla ja klikkauspäästösten teolla." Ja toden totta, scrollytelling-jutut ovatkin usein todella informatiivisia, viihdyttäviä ja kaiken kaikkiaan yksinkertaisesti näyttäviä. Silti mielestäni joskus voi olla vielä informatiivisempää ja viihdyttävämpää, joskus jopa välttämätöntä, että lukija itse klikkaa tai valitsee, mistä haluaa tarkemmin tietoa. Toimittaja ei millään voi aina tietää, mikä tieto on kullekin lukijalle se tärkein. Niinpä yksi datajournalistin tehtävistä mielestäni on, paitsi selvittää yleiskuva, myös tehdä monimutkaisesta tietomassasta helposti lähestyttävää ja tarjota lukijalle keinoja löytää tietomassasta juuri häntä kiinnostavat tiedot.

Esimerkki scrollytelling -jutusta:

Could Trump Really Deport Millions of Unauthorized Immigrants?

<https://www.nytimes.com/interactive/2016/11/29/us/trump-unauthorized-immigrants.html>

Esimerkki jutusta, jossa lukija pääsee määrittelemään, mistä tarkemmin ottaen haluaa tietoa:

Vain 11 prosenttia synnyttää ilman kipulääkkeitä - Lue sinulle räätälöity juttu siitä, miten lapsesi tulisi maailmaan

<https://dynamic.hs.fi/2017/synnytyssairaalat>

2. Tekstin analysoiminen

Massiivisten tekstimassojen analysointi tietokoneavusteisesti on hyvin suosittua Yhdysvalloissa datajournalistien keskuudessa. Esimerkiksi Los Angeles Times loi algoritmin, jonka avulla se analysoi tuhansia poliisiraportteja ja huomasi, että vakavat rikokset olivat usein merkitty pienemmiksi rikkeiksi.

Tekstianalyysi on toki paljon helpompaa englanninkielisestä tekstistä kuin suomen kielestä, koska erilaisia analyysikirjastoja on tarjolla englanniksi pilvin pimein, suomeksi ei niinkään. Lisäksi suomen kielen taivutusmuodot lisäävät oman hankaluutensa soppaan. Mutta toisaalta, tekstidatan määrä on Suomessa usein pienempi kuin Yhdysvalloissa, jolloin tekstin luku manuaalisesti ilman koneita voi olla hyvinkin mahdollista.

(Esimerkki jättimäisestä tekstianalyysistä tietokoneavusteisesti:)

LAPD underreported serious assaults, skewing crime stats for 8 years

<https://www.latimes.com/local/cityhall/la-me-crime-stats-20151015-story.html>

(Esimerkki Suomesta, jossa toimittajat lukivat itse kaikki Suomessa annetut kunnianloukkaustuomiot. Koska analysoituja tapauksia oli vain 867, ne oli suhteellisen helppo lukea itse)

Solvaukset syyttäjän pöydällä

<https://www.hs.fi/sunnuntai/art-2000006028892.html>

3. Algoritmien toiminnan vahtiminen

Toimittajat kirjoittavat algoritmeista ja niiden toiminnasta paljon Yhdysvalloissa. He analysoivat miten yritysten tai julkisen sektorin käyttämät algoritmit toimivat, toimivatko ne kuten niiden pitäisi ja mitä se tarkoittaa ihmisten ja oikean elämän kannalta, että algoritmit toimivat kuten toimivat. Sanalle sanoen: toimittajat ovat aina olleet vallan vahtikoiria, nyt he ovat myös algoritmien vahtikoiria.

Esimerkiksi ProPublica tutki ohjelmistoa, jota oikeusistuimet käyttävät rikollisten riskiarviointien tekemiseen. Ohjelmisto ei kysy tuomittavan rotua, mutta ProPublica sai selville, että algoritmi syrjii mustia. Vai syrjiikö? Washington Post kirjoitti ProPublican skrupin jälkeen artikkelin, jonka mukaan syrjinnän olemassaolo riippuu siitä, millaisella

mittaristolla algoritmiä tarkastelee. Ja tärkein oppi tässä kaikessa oli se, että kumpikin tiedotusvälineistä oli oikeassa: nykyisellä demografialla on mahdotonta luoda algoritmiä, joka olisi kaikista näkökulmista neutraali. Opetus siis oli, että ohjelmistot, koodaus ja algoritmit eivät koskaan ole täysin neutraaleja ja objektiivisia, vaan että ihmiset tekevät päätöksiä siitä, minkä suhteen algoritmit optimoidaan eli mitä niillä viime kädessä haetaan. Siten algoritmeissa voi aina olla tarkoituksellista tai vahingossa syntynyttä puolueellisuutta.

Esimerkkejä:

“Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks.”

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

“A computer program used for bail and sentencing decisions was labeled biased against blacks. It’s actually not that clear.”

https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/?utm_term=.1e6ed6e00c30

4. Skreippaus

Skreippaus, eli tietojen tallentaminen nettisivulta - tai sivustolta koodin tai tietokoneohjelman avulla on suosittu tapa hankkia dataa datajournalistien keskuudessa Yhdysvalloissa. Esimerkiksi New York Times skreippasi juoksijoita ja heidän välineitään koskevia tietoja juoksusivustolta ja analysoi, pitikö Niken mainoslause todella paikkansa. Nike väitti, että eräs heidän kenkämallinsa tekisi juoksijoista nopeampia, ja yllättävää kyllä, New York Timesin analyysin mukaan näin saattaa hyvinkin olla.

Ongelma skreippauksessa Yhdysvalloissa tosin on se, että sitä koskeva lainsäädäntö on sekavaa ja epäselvää, joten isot skreippausprojektit ovat usein mahdottomia. Mutta on myös tapoja hillitä näitä ongelmia. Esimerkiksi, ProPublica teki ison, jatkuvasti päivittyvän jutun Facebookin poliittisista mainoksista. Datan ProPublica hankki lukijoidensa avustuksella. Lukijat asensivat nettiselaimensa pienen lisäosan, joka tallensi mainostiedot ja lähetti tiedot sitten ProPublicalle. Näin ollen ProPublica ei skreipannut tietoja itse, kun taas lukijat saivat tehdä omille tiedoilleen mitä tahansa, kuten lahjoittaa ne ProPublicalle.

Esimerkkejä:

Facebook Political Ad Collector. How Political Advertisers Target You

<https://projects.propublica.org/facebook-ads/?lang=en-US>

Nike Says Its \$250 Running Shoes Will Make You Run Much Faster. What if That’s Actually True?

<https://www.nytimes.com/interactive/2018/07/18/upshot/nike-vaporfly-shoe-strava.html>

5. Tietyt teemat: Yksityisyys, Twitter ja populaarikulttuuri

Tietyt teemat näyttävät kiinnostavan datajournalisteja erityisen paljon Yhdysvalloissa. Paljon juttuja tehdään esimerkiksi ihmisten yksityisyydestä internetin aikakaudella, Twitteristä ja populaarikulttuurista kuten musiikista ja amerikkalaisesta jalkapallosta.

Esimerkkejä:

(Hyvä esimerkki tekstianalyysistä, scrollytelling-kerronnasta ja yksityisyydestä:)

We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.

<https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html>

(Hyvä esimerkki Twitteriä käsittelevästä uutisoinnista. Tätä Pulitzer-palkittua juttua oli tekemässä myös Columbiassa opettava Mark Hansen, joka veti Computational Journalism- kurssiani.)

The Follower Factory:

<https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html>

(Hyvä esimerkki scrollytelling-tekniikasta ja populaarikulttuuria käsittelevästä data journalismista:)

The unlikely odds of making it big

<https://pudding.cool/2017/01/making-it-big/>

Loppusanat

Columbian yliopistossa viettämäni vuosi oli yksi elämäni parhaista vuosista. Olen ikuisesti kiitollinen Helsingin Sanomien Säätiölle sen mahdollistamisesta. Vähintään yhtä kiitollinen olen puolisololleni ja lapsilleni: olitte parasta mahdollista seuraa New Yorkin seikkailuissa.