

Tekoälykokeilun loppuraportti

Suosittelualgoritmi Löytö

Helsinki



Tukea digitalisaatiokokeiluihin
kaupungin työntekijöille

Suosittelualgoritmi Löytö

Sari Lehikoinen, Kulttuurin ja vapaa-ajan toimiala,
Helsingin kaupunki

Deloitte Consulting Oy

1. Kokeilun onnistuminen

Tavoitteet

- Kokeilussa haluttiin testata, kuinka tekoälyä voidaan hyödyntää asiakaskohtaisesti kohdennetussa markkinoinnissa
- Ihanteellisena lopputuloksena olisi suosittelualgoritmi, joka olisi niin yksinkertainen, että se voitaisiin monistaa myös muiden toimialojen hyödynnettäväksi

Lopputulokset

- Kokeilun aikana analysoitiin ja mallinnettiin Helsingin kaupungin tapahtumadataa ja rakennettiin käyttöliittymä, jolla algoritmia voidaan testata eri käyttäjien näkökulmasta
- Käyttöliittymä pidettiin yksinkertaisena, mutta sen avulla voidaan havainnollistaa suosittelualgoritmin tarjoamia mahdollisuuksia ja haluttaessa muuttaa algoritmin antamien pisteytyksien painokertoimia
- Algoritmia voidaan monistaa muiden toimialojen käyttöön

2. Kokeilun eteneminen

- Kulttuurin ja vapaa-ajan toimiala markkinoi palvelujaan aktiivisesti, mutta automatiikkaa ja tekoälyä ei vielä hyödynnetä palvelun käytön lisäämiseksi
- Helsingin kaupunki halusi osana kokeilukiihdyttämöä testata suosittelualgoritmin toimintaa ja tarvitsi siinä teknistä tukea
- Kokeilussa testattiin, miten tekoälyä voidaan hyödyntää asiakaskohtaisesti kohdennetussa markkinoinnissa
- Suosittelualgoritmista haluttiin riittävän yksinkertainen, jotta se voitaisiin monistaa myös muiden toimialojen hyödynnettäväksi
- Pidemmällä aikavälillä suosittelutoiminnon taustalla olisivat erilaiset tietokannat, käyttäjien antamat preferenssit (kiinnostuksen kohteet yms.) sekä heidän aikaisempi toimintansa sivustoilla (tykkäykset ja mitä tapahtumia käyttäjä on katsonut sovelluksessa)
- Löytö-palvelun pilottiversiota lähdettiin kokeilemaan MVP periaatteella

2. Kokeilun eteneminen



1. Datat kerääminen

Tavoite: *Analyysissä hyödynnettävä aineisto*

- Haluttiin kerätä data, joka sisältäisi tiedot tapahtumien asiasanoista, päivämäärästä, tapahtumapaikasta ja suosiosta.
- Kokeilu toteutettiin historiallisella datalla, joka ei päivity kokeilun aikana.



2. Datat valmistelu

Tavoite: *Analyysivalmis data*

- Aineistot vietiin analytiikkaympäristöön, data yhdistettiin ja muuttujat käsiteltiin sellaiseen muotoon, että ne olivat seuraavassa vaiheessa hyödynnettävissä. Esimerkiksi määriteltiin, miten usein asiasanat esiintyvät yhdessä.
- Suosiotietoa ei pilotin aikana ollut tapahtumille saatavana. Tämän kokeilun jälkeen, suosiona voidaan hyödyntää tietoja siitä, mitä käyttäjät klikkailevat suosittelualgoritmia käyttäessään.



3. Algoritmin laatiminen

Tavoite: *Yksinkertainen suosittelualgoritmi*

- Määriteltiin algoritmi, joka suosittelee käyttäjille tapahtumia käyttäjän valitsemien häntä kiinnostavien asiasanojen pohjalta.
- Käytettiin yksinkertaista algoritmia, jotta mallin toimivuuden arviointi olisi pilottivaiheessa helpompaa.
- Laadittiin indeksi, joka kuvaa erikseen, kuinka hyvä tapahtuma on käyttäjälle asiasanojen, sijainnin, kohdeyleisön ja tapahtuma-ajan perusteella ja suositellaan näiden summaan/ tuloon perustuvia tapahtumia.



4. Tulosten visualisointi

Tavoite: *Tulosten esittäminen muodossa, joka mahdollistaa käyttäjille algoritmin toiminnan arvioinnin*

- Luotiin visualisointi, joka antaa käyttäjille mahdollisuuden valita itseään kiinnostavia asiasanoja ja nähdä, mitä tapahtumia algoritmi tällöin suosittelisi hänelle.
- Kokeilussa hyödynnettiin valmista visualisointiohjelmaa, jotta voitiin keskittyä olennaisimpaan, eli algoritmin laatimiseen.

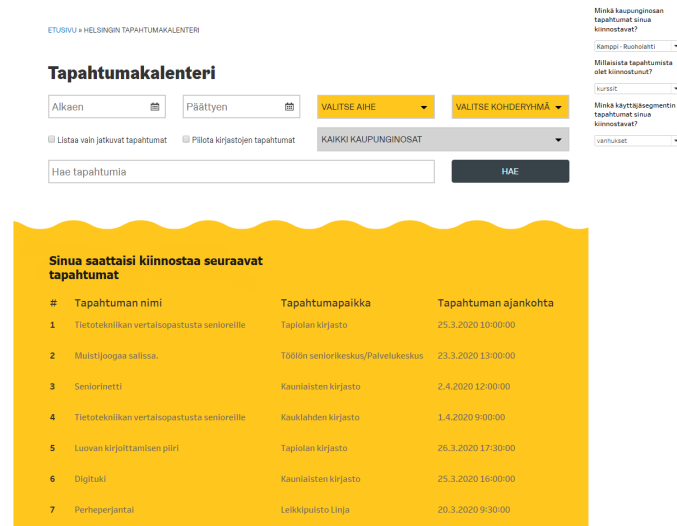
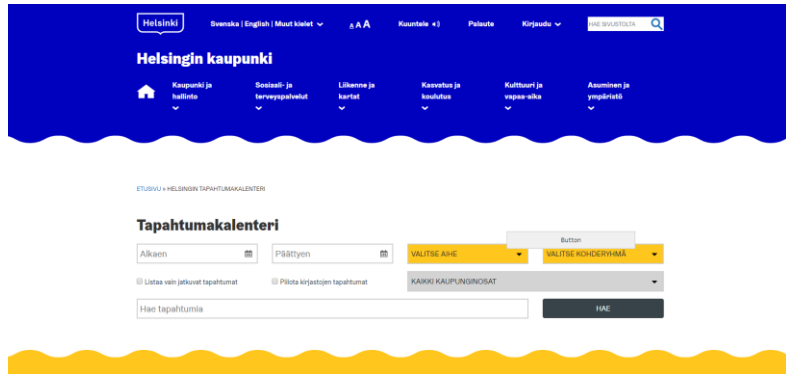


5. Loppuraportti

Tavoite: *Kevyt loppuraportti*

3. Kokeilun tuotokset

- MVP suosittelualgoritmi, sekä
- MVP Tableau -työkirja, joka havainnollistaa Löydön TOP 10 suosituksukset kuvitteelliselle käyttäjälle Helsingin kaupungin tapahtuma-kalenterin yhteydessä



3. Kokeilun tuotokset

ETUSIVU » HELSINGIN TAPAHTUMAKALENTERI

Tapahtumakalenteri

Alkaen Päättyen

VALITSE AIHE VALITSE KOHDERYHMÄ

Listaa vain jatkuvat tapahtumat Piilota kirjastojen tapahtumat

KAIKKI KAUPUNGINOSAT

Hae tapahtumia HAE

Minkä kaupunginosan tapahtumat sinua kiinnostavat?

Kamppi - Ruoholahti

Millaisista tapahtumista olet kiinnostunut?

kurssit

Minkä käyttäjäsegmentin tapahtumat sinua kiinnostavat?

vanhukset

Sinua saattaisi kiinnostaa seuraavat tapahtumat

#	Tapahtuman nimi	Tapahtumapaikka	Tapahtuman ajankohta
1	Tietotekniikan vertaisopastusta senioreille	Tapiolan kirjasto	25.3.2020 10:00:00
2	Muistijoogaa salissa.	Töölön seniorikeskus/Palvelukeskus	23.3.2020 13:00:00
3	Seniorinetti	Kauniaisten kirjasto	2.4.2020 12:00:00
4	Tietotekniikan vertaisopastusta senioreille	Kauklahden kirjasto	1.4.2020 9:00:00
5	Luovan kirjoittamisen piiri	Tapiolan kirjasto	26.3.2020 17:30:00
6	Digituki	Kauniaisten kirjasto	25.3.2020 16:00:00
7	Perheperjantai	Leikkipuisto Linja	20.3.2020 9:30:00

Pilotissa käyttäjä valitsee häntä kiinnostavat asiasanat valikoista

3. Kokeilun tuotokset

Kuinka suosittelukone toimii:



4. Opit tekoälyn mahdollisuuksista

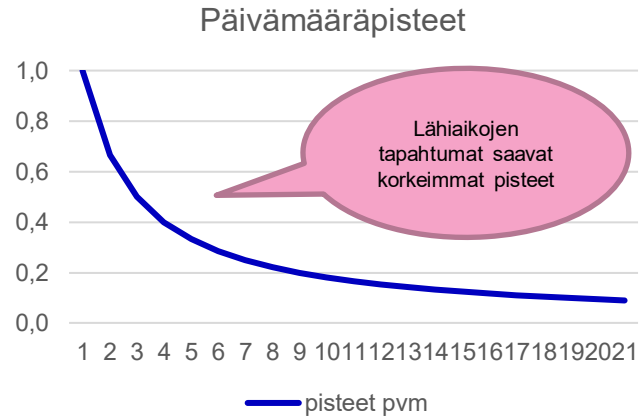
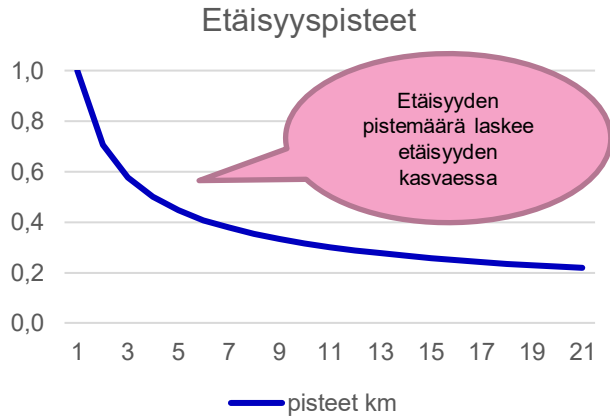
- Kaupungilla yleensä ja kulttuurin ja vapaa-ajan toimialalla on laaja ja moninainen käyttäjäryhmä, minkä johdosta käyttäjien ohjaaminen palveluiden luokse esimerkiksi digikanavissa on hankalaa
- Palveluiden löytäminen voi vaatia kohtuuttomasti aikaa ja vaivaa käyttäjältä
- Tekoälyn avulla voidaan lyhentää palveluiden asiakaspolkua ja tuoda palvelut ja tapahtumat lähemmäksi asiakasta
- Erityisesti tästä voisi olla hyötyä sellaisten käyttäjäryhmien kanssa, jotka eivät aktiivisesti etsi palveluja tai joiden tiedonhakuun käyttämä aika ja vaiva on vähäinen
- Tekoälyn avulla voitaisiin lisäksi suositella sellaisia tapahtumia tms. kulttuurin ja vapaa-ajan toimialan palveluja, joita asiakas ei tunne entuudestaan
- Ihmisen käyttäytymisen ennustaminen on vaikeaa, tekoälyä hyödyntämällä voi osua oikeaan suositukseen tai sitten ei > asiakastyytyväisyyden jatkuva seuranta tärkeää

5. Opit tekoälyn kehittämisestä

- Kokeiluvaiheessa ei ollut dataa siitä, millaisista tapahtumista eri käyttäjäryhmät ovat olleet kiinnostuneita
- Tätä kutsutaan termeillä 'Cold-start-problem' ja on tyypillinen suosittelusysteemien kehittämiselle ja vaikuttaa vahvasti siihen, millaisia menetelmiä kokeilussa on mahdollista käyttää tekoälyn kehittämiseen
- Kokeilussa ratkaisimme ongelman antamalla käyttäjälle mahdollisuuden syöttää preferenssejään
- Mikäli kokeilun jälkeen siirrytään tuotantovaiheeseen, voidaan kerätä tietoa siitä, mistä tapahtumista käyttäjät ovat kiinnostuneita ja käyttää tätä tietoa algoritmin kehittämiseen
- Emme pystyneet tässä projektissa matemaattisesti vertaamaan eri vaihtoehtoisia tekoälymalleja siinä, kuinka hyviä suosituksia ne antavat
- Tästä johtuen projektissa ei rakennettu black box ratkaisua vaan malli, jonka syötteet ovat ymmärrettäviä ja ihmisarvioiden perusteella voidaan muuttaa mallin kertoimia
- Mallia voitaisiin jatkossa kehittää siten, että se tarjoaisi käyttäjälle mahdollisuuden vaikuttaa eri muuttujien painoarvoon (esim. kun käyttäjä määrittelee, että sijainti ei ole niin tärkeä, suosittelualgoritmin pisteytystä voitaisiin muuttaa siten, että sijainnille annettaisiin pienempi painoarvo).

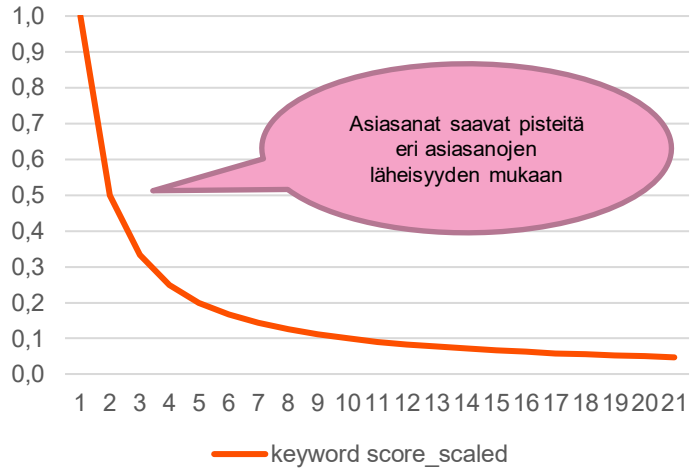
6. Opit kokeilemisesta

- Pisteytyksen painojen muuttaminen yksinkertaisella tavalla antaa mahdollisuuden lähteä kokeilemaan mikä toimii parhaiten
- Pisteytyksen toimivuutta ja osumatarkkuutta kannattaa testata jatkuvasti

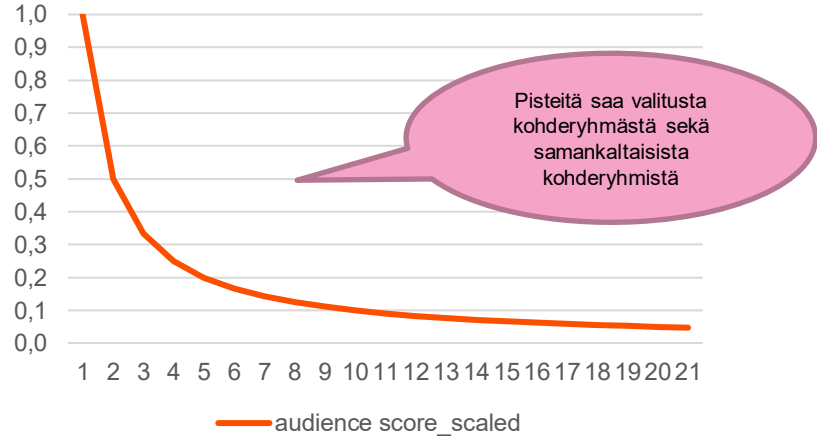


6. Opit kokeilemisesta

Asiasanojen pisteet



Kohderyhmän pisteet



6. Opit kokeilemisesta

pisteet	km	pisteet p	pvm ero	audience määrät	audience score_scaled	keyword määrät	keyword score_scaled
1,0	1	1,0	0	1	1,0	1	1,0
0,7	2	0,7	1	2	0,5	2	0,5
0,6	3	0,5	2	3	0,3	3	0,3
0,5	4	0,4	3	4	0,3	4	0,3
0,4	5	0,3	4	5	0,2	5	0,2
0,4	6	0,3	5	6	0,2	6	0,2
0,4	7	0,3	6	7	0,1	7	0,1
0,4	8	0,3	7	8	0,1	8	0,1
0,3	9	0,2	8	9	0,1	9	0,1
0,3	10	0,2	9	10	0,1	10	0,1
0,3	11	0,2	10	11	0,1	11	0,1
0,3	12	0,2	11	12	0,1	12	0,1
0,3	13	0,1	12	13	0,1	13	0,1
0,3	14	0,1	13	14	0,1	14	0,1
0,3	15	0,1	14	15	0,1	15	0,1
0,3	16	0,1	15	16	0,1	16	0,1
0,2	17	0,1	16	17	0,1	17	0,1
0,2	18	0,1	17	18	0,1	18	0,1
0,2	19	0,1	18	19	0,1	19	0,1
0,2	20	0,1	19	20	0,1	20	0,1
0,2	21	0,1	20	21	0,0	21	0,0

=1/keyword määrät

7. Opit resursoinnista

- Kokeilun budjetti käytettiin kokonaan ja se pakotti fokusoimaan tekemisen tarkasti.
- Suunnitteluun käytettiin minimiaika ja pyrittiin toteuttamaan yksi versio MVP-hengessä.
- Tapaamiset: 10 h
- Datankäytön hankinta: 20 h
- Algoritmi: 40 h
- Tableau-työkirjat: 5 h
- Loppuraportti: 5 h

8. Kokeilun tekninen ympäristö

- <http://api.hel.fi/> REST-rajapinta
 - Linkedevents-tapahtumat
 - Päälähde datalle
- **Python**
 - Datan haku REST-rajapinnasta
 - Asiasanoista muodostetun matriisin luominen, jonka avulla pystytään tunnistamaan jokainen mahdollinen asiasana-kombinaatio
 - Asiasanojen suhteellisen esiintyvyyden laskeminen
 - Tapahtumien lokaalitiedon yhdistäminen ulkopuoliseen tietolähteeseen, josta tunnistetaan tapahtumien kaupunginosat
 - Tapahtumien lokaatiotiedon avulla laskettu eri kaupunginosien välinen etäisyys
 - Eri lokaatiokombinaatioista luotu matriisi, johon on laskettu kaupunginosien välinen etäisyys
 - Tapahtumista luotu taulu, joka sisältää kaikki tapahtumat, asiasanat ja lokaatiot
- **Tableau**
 - MVP
 - Visualisointi perustuen tapahtumatauluun

9. Kokeilun data

- Kokeilussa käytettiin Linkedevents-tapahtumat –dataa ajanjaksolta 01/2014 – 09/2020
- Data sisälsi tapahtumien asiasanat ja lokaatiot
- Data haettiin pythonilla <http://api.hel.fi/> REST-rajapinnasta

10. Jatkopäätökset ja -ideat

- Jatkosta, rahoituksesta tms. ei ole ollut mahdollista keskustella vallitsevan poikkeuksellisen tilanteen vuoksi
- Asiaan palataan, kun on palattu normaalioloihin