

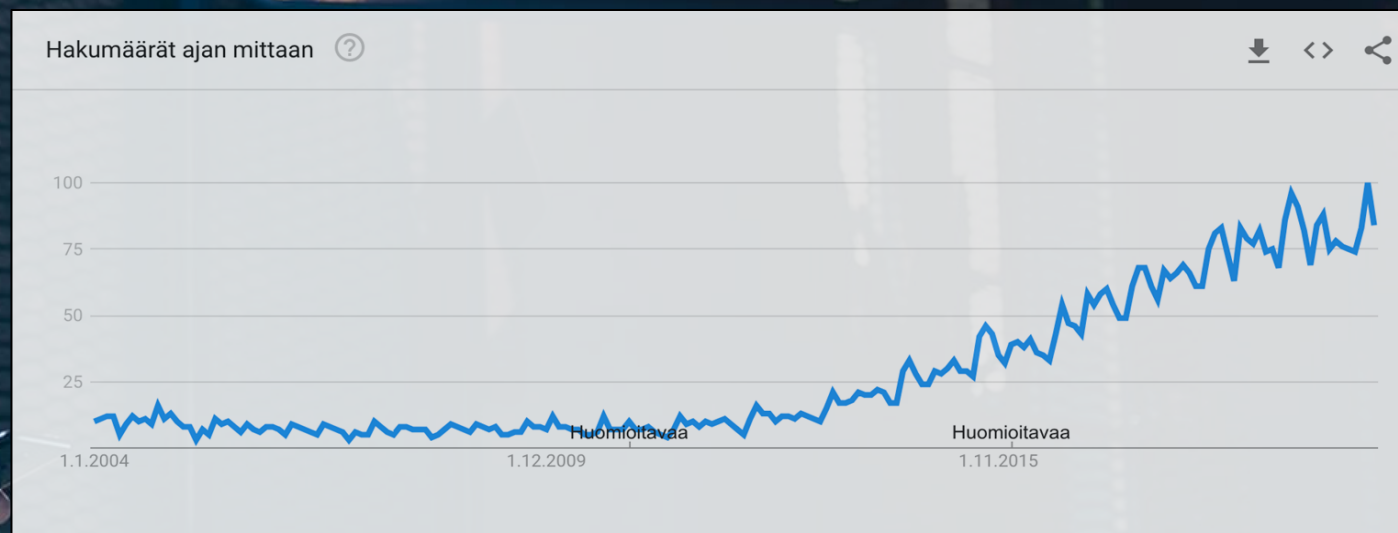
A photograph of three young adults (two men and one woman) standing at night, looking at their smartphones. The man on the left is wearing a blue shirt, the woman in the middle is wearing a denim jacket, and the woman on the right is wearing a black leather jacket. They are all smiling and looking at their phones. The background is blurred with city lights.

Onko Data Science Salatiedettä?

Koneoppimisesta, algoritmeista ja
mallinnuksesta puhutaan paljon.

Mitä niillä oikeasti tehdään?

Google Trends: “Data Science”

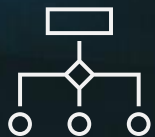




Koneoppiminen



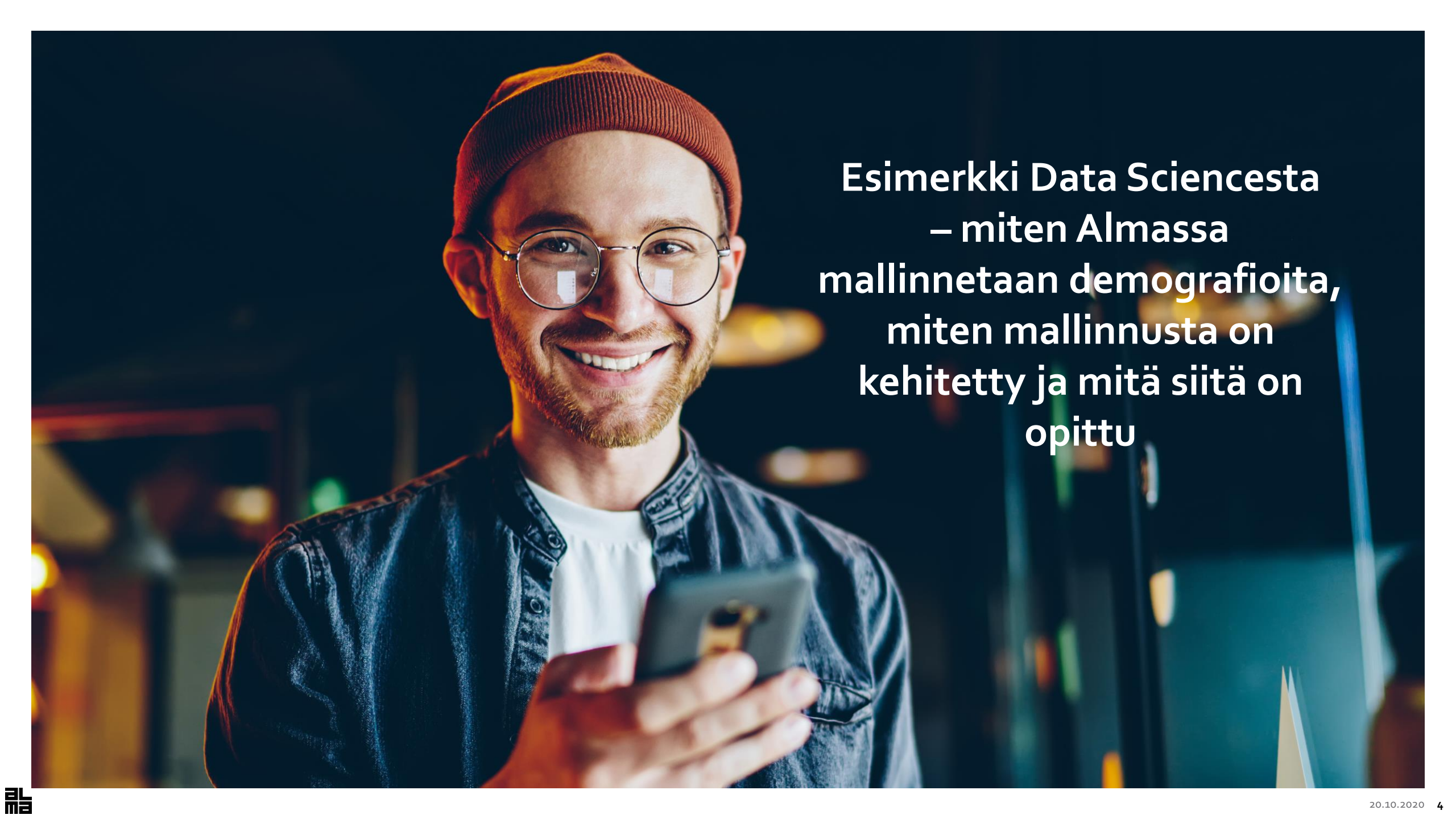
Datan
mallintaminen



Algoritmi



Tietosuojasääntely
ja suostumukset



**Esimerkki Data Sciencesta
– miten Almassa
mallinnetaan demografioita,
miten mallinnusta on
kehitetty ja mitä siitä on
opittu**

Almassa kertyy 1. osapuolen
dataa noin 500 miljoonaa
datasignaalia per kuukausi
kaikista Alman digipalveluista.



ILTALEHTI

Kauppalehti

ARVOPAPERI

**Tekniikka
& Talous**

 ETUOVI.com

tivi

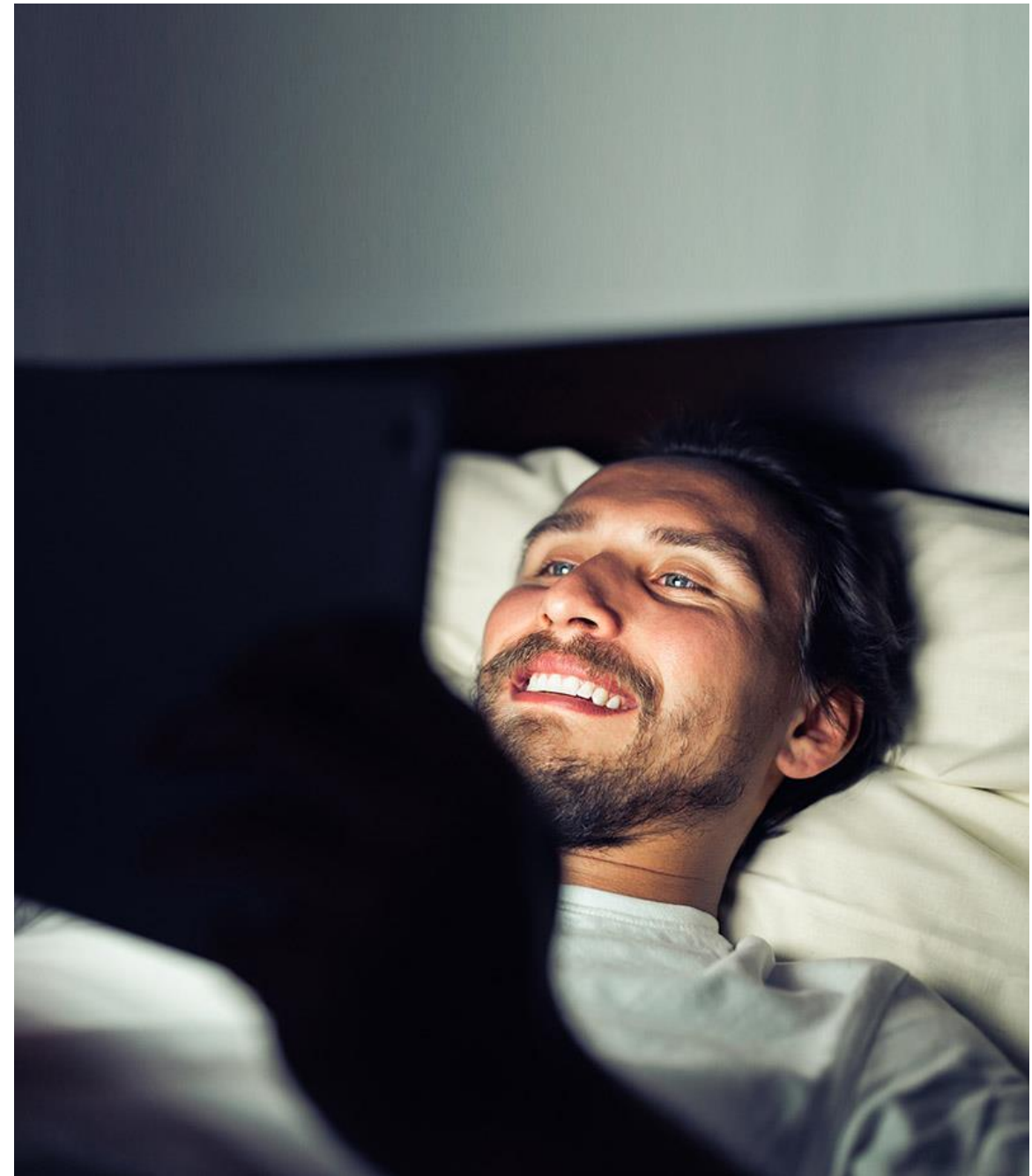
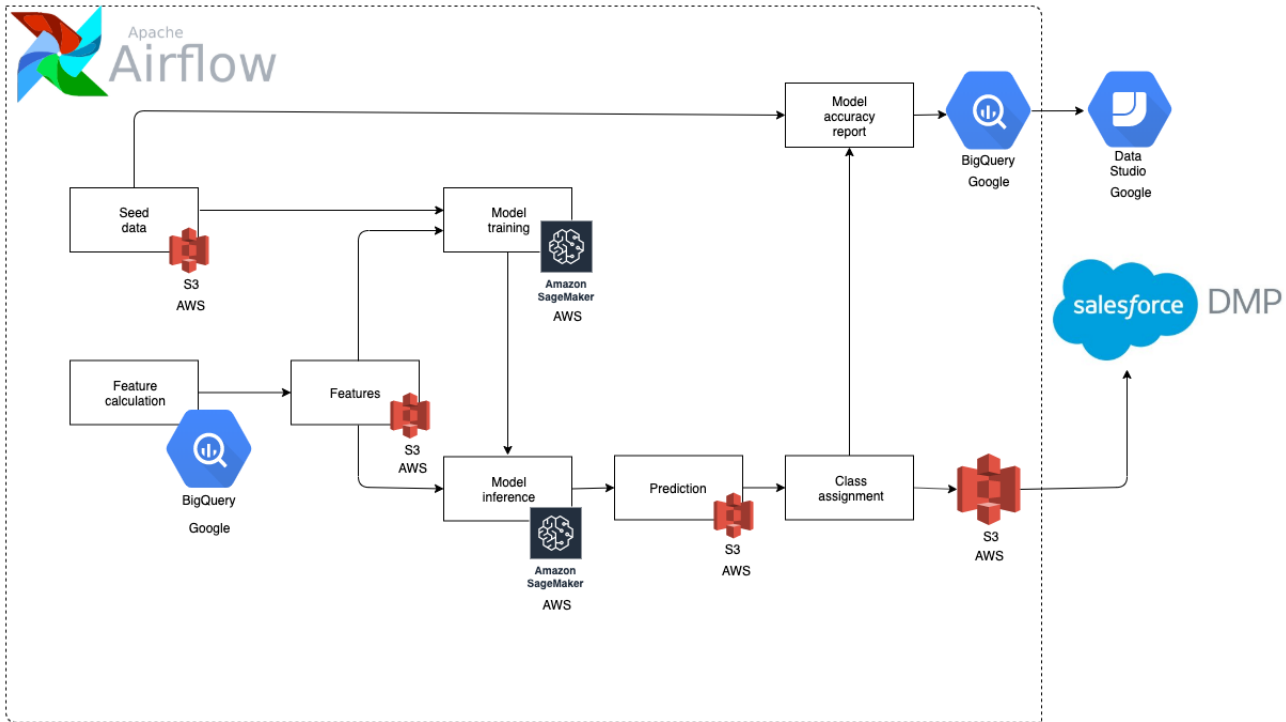
AUTOTALLI.com

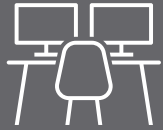
BT

 **telkku.com**



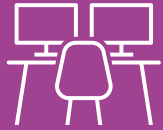
Demografiamallinnuksen tietojärjestelmä-arkkitehtuuri





Miten luodaan mallinnettuja yleisöjä mainonnan kohdentamiseen?





Almassa mallinnetaan mm. :



Demografiasegmenttejä

Mainostaja-asiakkaiden dataa

Molemmissa tapauksissa mallinnuksen peruslogiikka on sama,
vaikka seed-datan lähde on eri.



Miten demografiat mallinnetaan?

Kaikille Alman sivustojen kävijöille
ennustetaan demografiat



Malli koulutetaan opetusdataalla



ID	NAINEN	MIES	OMAKOTI-TALO	RIVITALO	KERROS-TALO	Space	Politics	Educatio	
A	1	0	1	0	0	0,2	0,5	0,1	
B	0	1	0	1	0	0,5	0,1	0	
C	1	0	1	0	0	0	0,6	0,2	
D	1	0	0	0	1	0,1	0	0	
E	0	1	0	0	1	0	0	0,7	
F	1	0	0	1	0	0	0,2	0	
G	1	0	0	0	1	0	0,3	0	



Opetus-data



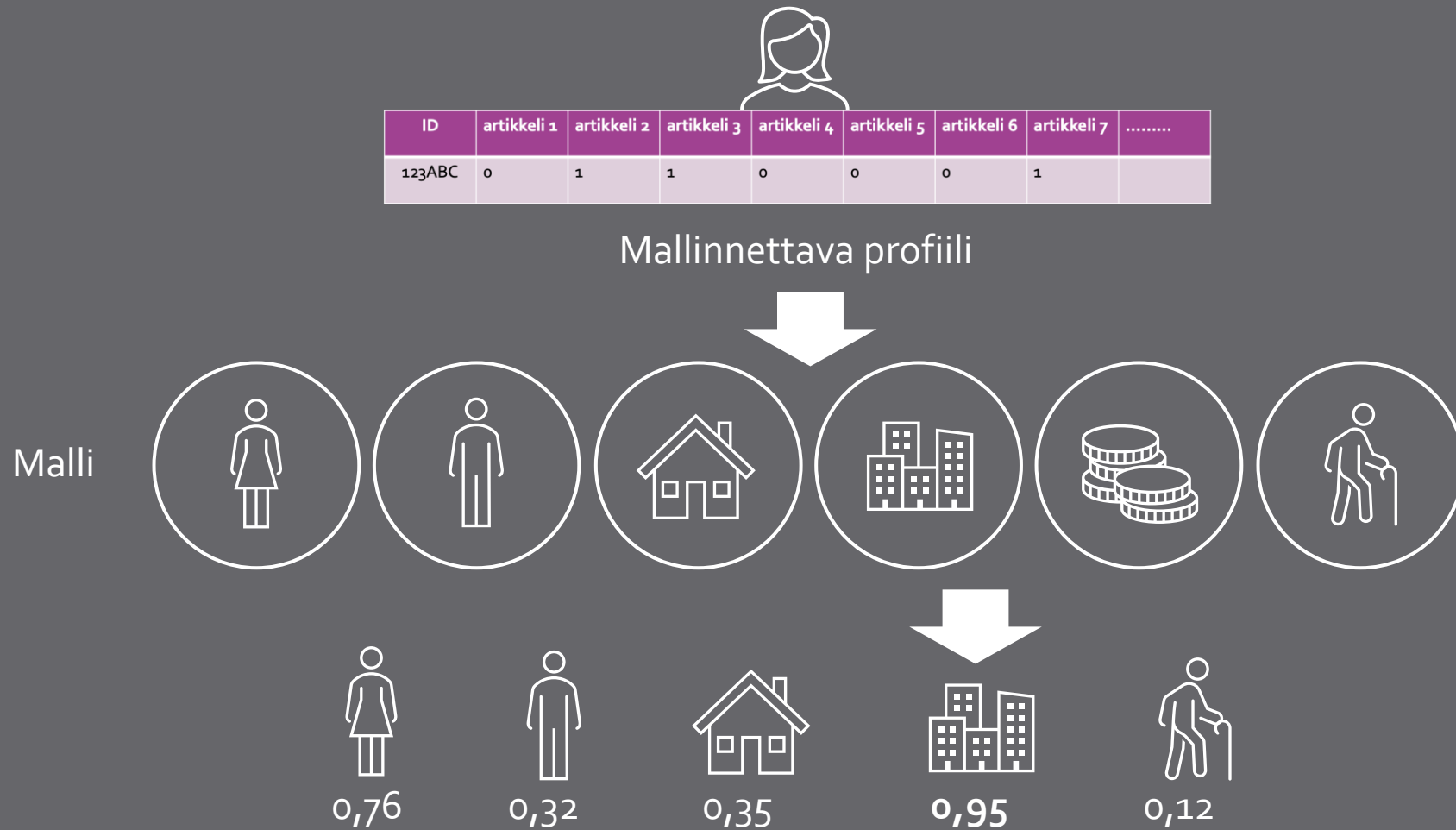
Malli: Hybrid matrix factorization



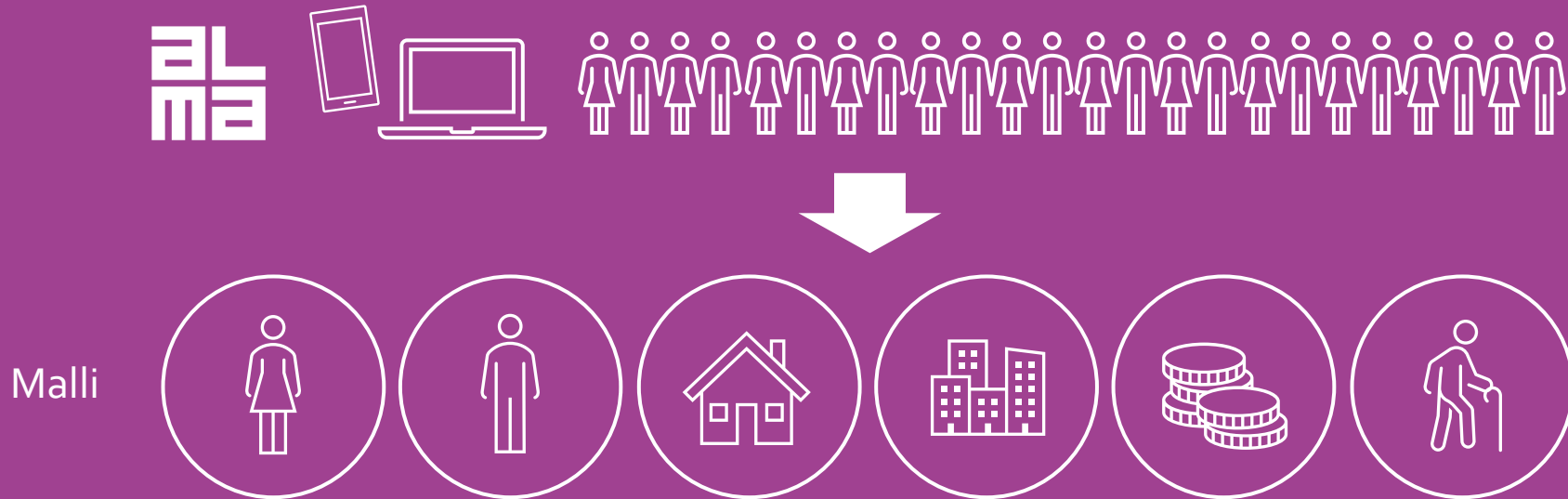
Malli



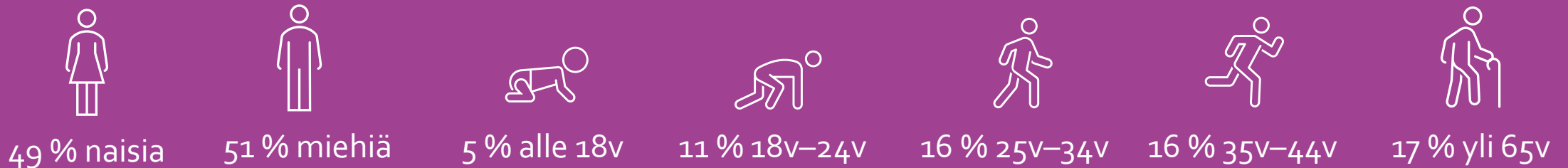
Tuntemattomalle käyttäjälle luodaan ennusteet käyttäjäprofiilin perusteella



Sivustojen käyttäjät ajetaan päivittäin mallinnuksen läpi



Käyttäjät jaotellaan demografioihin digiyleisön jakauman mukaisesti



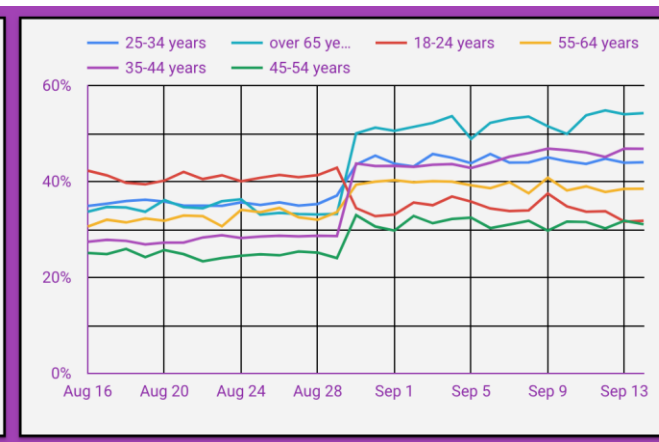
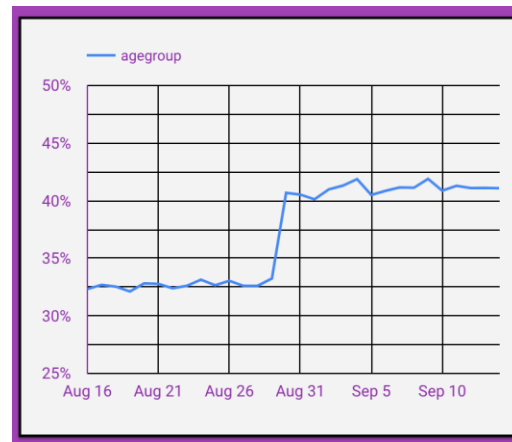
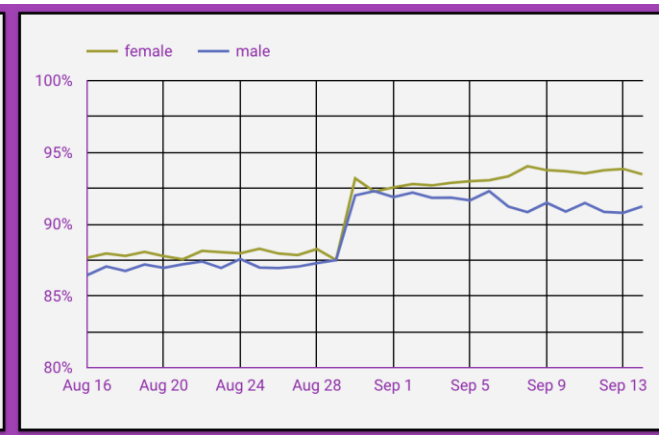
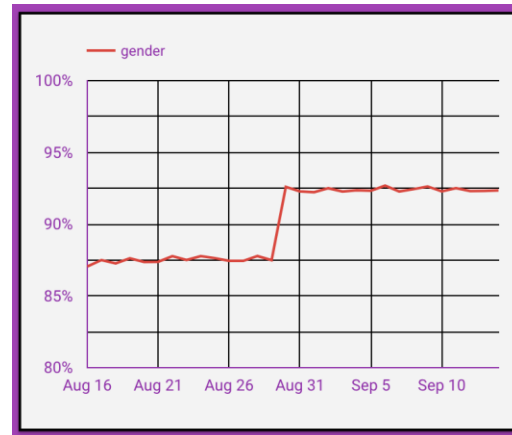


Alman vahvuutena on
omat data scientistit

Mitä tapahtui, kun muutimme mallinnustapaa?

Segmenttien mallinnustarkkuus
kehittyi huimasti:

- Per demografiaryhmä
tarkkuuden nousu ka +18 %
- Per demografiasegmentti
tarkkuuden nousu ka +40 %



Mitä muutimme?

Vaihdoimme mallinnuksessa käytetyn datan

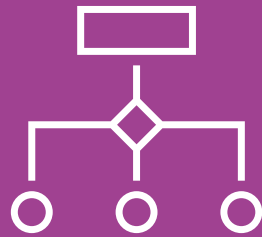
- Aikaisemmin semanttinen asiasanoitus, nyt suora artikkelitieto
- Huomattava datamäärän lisäys

Vaihdoimme mallinnusalgoritmin

- Pystymme mallintamaan isomman datamassan



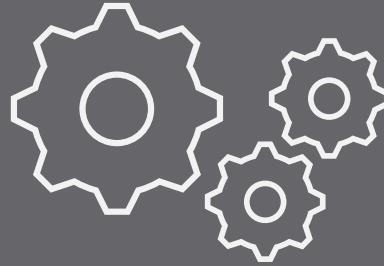
Jatkuva kokeileminen ja testaaminen tuo uusia näkemyksiä



Laaja-alaisen artikkelisisällön kautta tehty mallinnus ei tarkkuudeltaan juurikaan parane lisäämällä url-tietoa tai muuta sivusto- tai osiotietoa.



Käyttäjien metadatan lisääminen kasvattaa merkittävästi datamäärää ja siten mallinnuksen laskenta-aikaa.



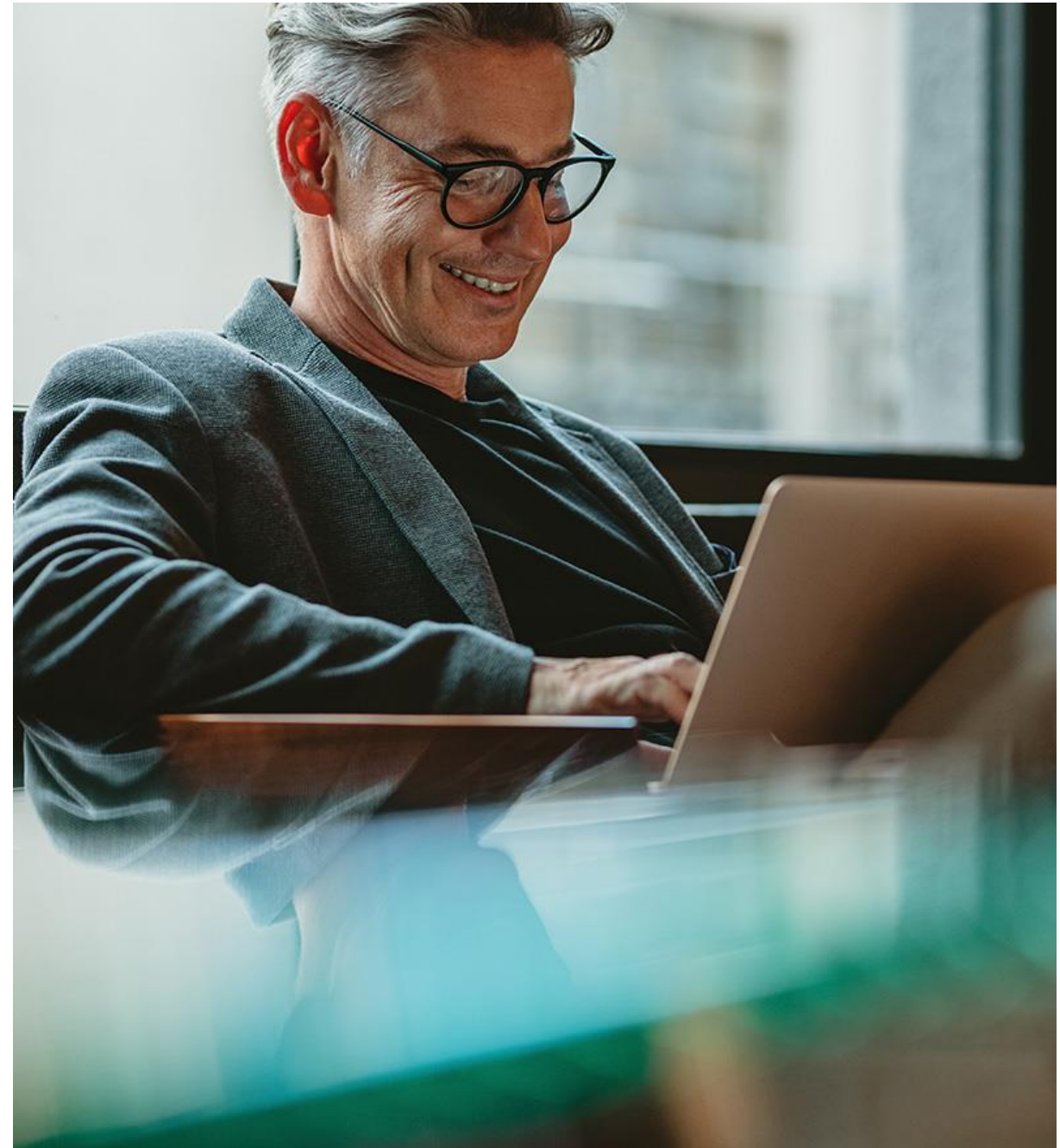
Kokeilemme erilaisia datalähteitä ja
algoritmeja

Pohdimme, kannattaako painottaa
segmenttipopulaatiota vai
mallinnustarkkuutta

Ennakoimme tehtävien muutosten
vaikutuksen

Onko Data Science siis salatiedettä?

Almassa ei ainakaan ole, vaan se on jokapäiväistä tekemistä niin mediamyynnin, sisältömyynnin kuin sisällön personoinninkin näkökulmasta katsoen.



**Data science on
jatkuvaa
kehitystyötä
– aina voi tulla
paremmaksi**





Kiitos!

Merja Pyhälä

puh. 0400 679 724
merja.pyhala@almamedia.fi