

Koolitus „Andmeteadus on Popp“

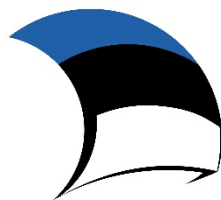
Tehisintellekti projektide juhtimine

November 2021

Kristjan Eljand



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Päevakava

Teema	Sisu
Andmeteaduse ja tehisintellekti projektide juhtimine	Tehisintellekti arendamise 7 sammu.
Miks tehisintellekti projektid ebaõnnestuvad	Peamised põhjused, miks ärid ebaõnnestuvad tehisintellekti rakendamisel.
Andmete visualiseerimine	Andmete loetavuse parendamine, andmetabelite visualiseerimine ja graafikute loomine.

Tehisintellekti arenduse 7 sammu

1. Seame eesmärgi ja tulemuskriteeriumi ehk küsime õigeid küsimusi;
2. Koondame andmed, mis võimaldavad eesmärki täita;
3. Kasutame algoritme, et leida mustreid;
4. Testime, kas algoritmid töötavad ka uute andmete peal;
5. Testime klientide peal enne tootestamist, kas kõik on neile sobiv;
6. Tootestame;
7. Hoiame süsteemi usaldusväärseks.

Tehisintellekti 7 sammu: Vegan vorsti analoogia

1. Seame eesmärgi ja tulemuskriteeriumi ehk küsime õigeid küsimusi;
 2. Koondame andmed;
 3. Proovime eri algoritme.
 4. Testime mudelit uutel andmetel.
 5. Testime tulemit klientide peal.
 6. Tootestame;
 7. Hoiame süsteemi usaldusväärseks.
1. Eesmärk on luua Michelini restorani vääriline suhkru, gluteiini- ja laktoosivaba vegan vorst.
 2. Koondame komponendid.
 3. Testime eri retsepte.
 4. Maitseme eri retseptide tulemusi.
 5. Testime, kas klientidele meeldib.
 6. Ehitame üles tootmisprotsessi.
 7. Muudame tootmisprotsessi stabiilseks ja usaldusväärseks.



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Tehisintellekti arenduse 7 sammu

1. Eesmärgi seadmine



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Milline neist on hea käitumine?



Tehisintellekti eesmärgi seadmine – tuvasta **vabad** parkimiskohad



Sisendandmed: videovoog meie parkimisplatsilt

Tehisintellekti alternatiivne – tuvasta valesi pargitud autod



Sisendandmed: videovoog meie parkimisplatsilt

Tehisintellekti alternatiivne – tuvasta **anomaaliad** parkimises



Sisendandmed: videovoog meie
parkimisplatsilt

Tehisintellekti alternatiivne eesmärk – tuvasta **sinised** autod



Tehisintellekti eesmärgi seab
otsustaja!

Tehisintellekti alternatiivne eesmärk – tuvasta kollased autod



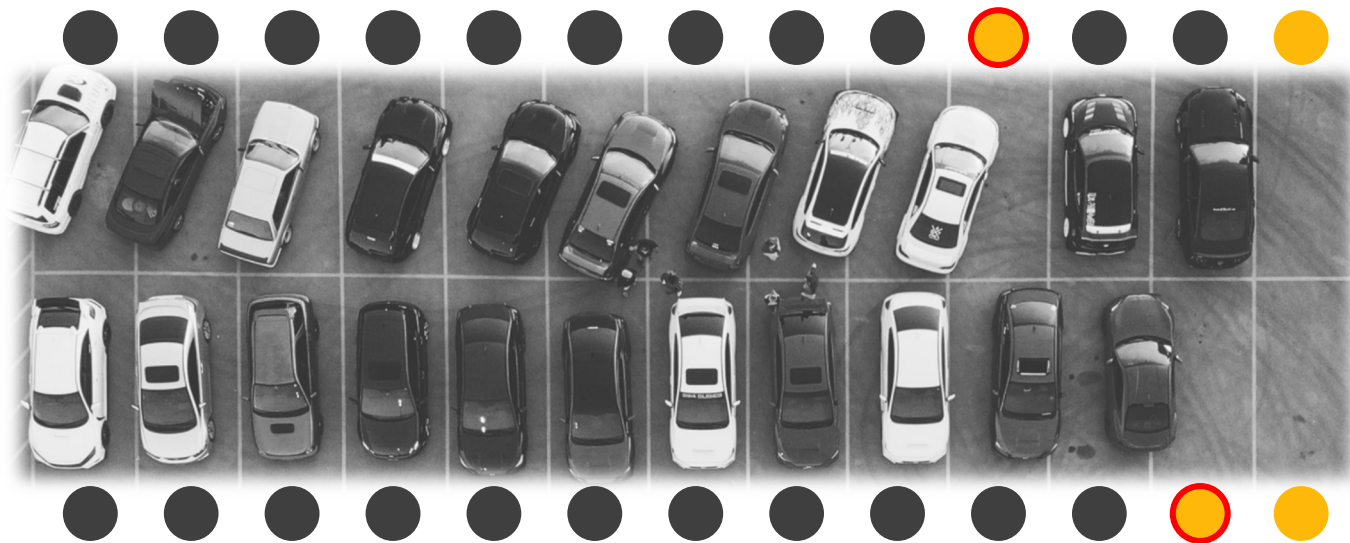
Eesmärk peab olema realistlik!

Eesmärgiks ei saa olla kollaste autode tuvastamine, kui meie näidetes ei ole kollaseid autosid!

Tehisintellekti eesmärgi seab otsustaja, kuna see sõltub tema vajadusest.

Eesmärk peab olema realistlik – ära sea eesmärgiks seda, mille kohta pole andmetes informatsiooni!

Tehisintellekti tulemusmõõdikud - õigsus



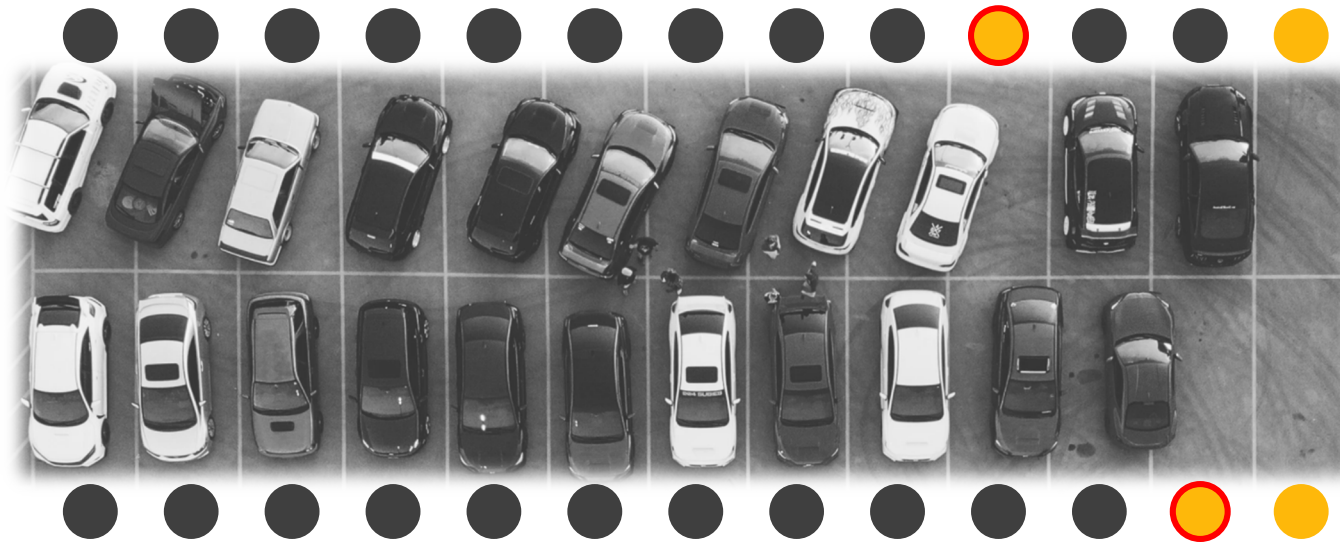
Parkimiskoht vaba ●

Parkimiskoht kinni ●

Õigsus (*accuracy*): kui palju õigeid vastuseid annab mudel?

24 õiget / 26 kohta = 92%

Tehisintellekti tulemusmõõdikud - täpsus



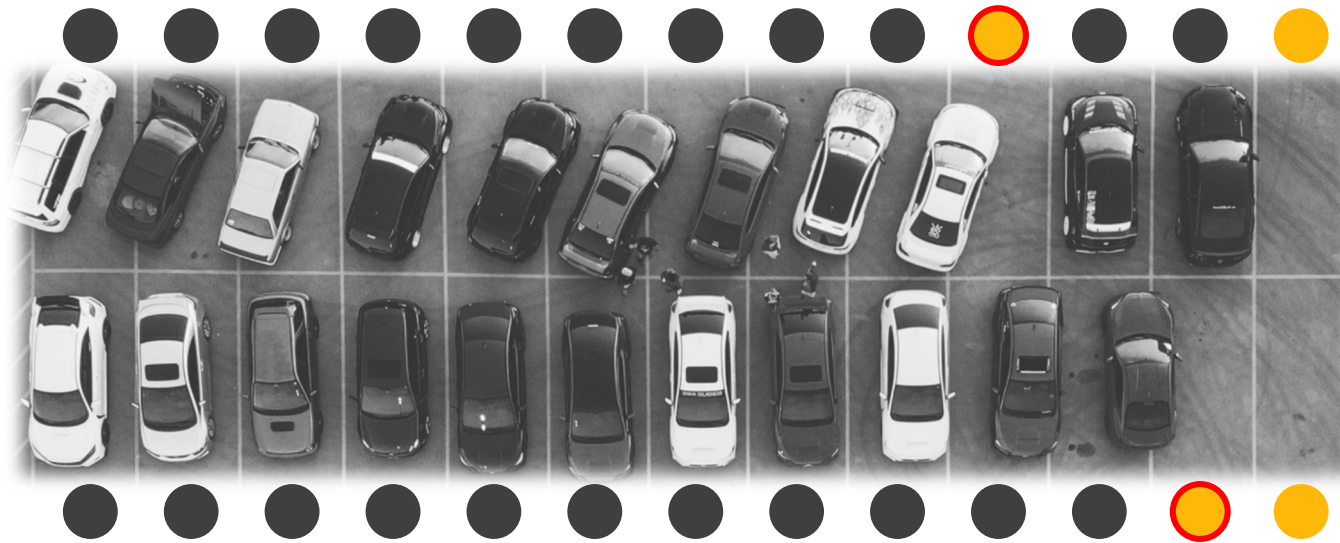
Parkimiskoht vaba ●

Parkimiskoht kinni ●

Täpsus (*precision*): kui palju „vaba“
prognoosidest on ka tegelikult vaba?

2 vaba / 4 vaba prognoosi = 50%

Tehisintellekti tulemusmõõdikud - saagis



Parkimiskoht vaba ●

Parkimiskoht kinni ●

saagis (*recall*): kui suure osa kõigist vabadest kohtadest mudel tuvastada suudab?

2 tuvastatud / 2 vaba kokku = 100%

Täpsus vs Saagis



Oluline kõrge täpsus –
raamatu soovitamise.



Oluline kõrge saagis –
maale langevate
meteoriitide tuvastamine.

Täpsus vs Saagis



Oluline kõrge täpsus –
rasedustest.



Oluline kõrge saagis –
Covid19 vaktsineerimine.

Täpsus vs Saagis kulude perspektiivist

Soovita vähem, aga need olgu head!



Oluline kõrge täpsus – kui otsekulu on kõrge **VÕI** kui alternatiivkulu on madal.



Kui me meteoriiti ei tuvasta, on probleem suur!

Oluline kõrge saagis – kui alternatiivkulu on kõrge.

Tehisintellekti eesmärgi täitmist näitab tulemusmõõdik.

Ka tulemuskriteeriumi (näiteks täpsus 95%) seab otsustaja, kuna tema teab kõige paremini, kas oluline on õigsus, täpsus või saagis!

Õigsus – kõige populaarsem tulemusmõõdik

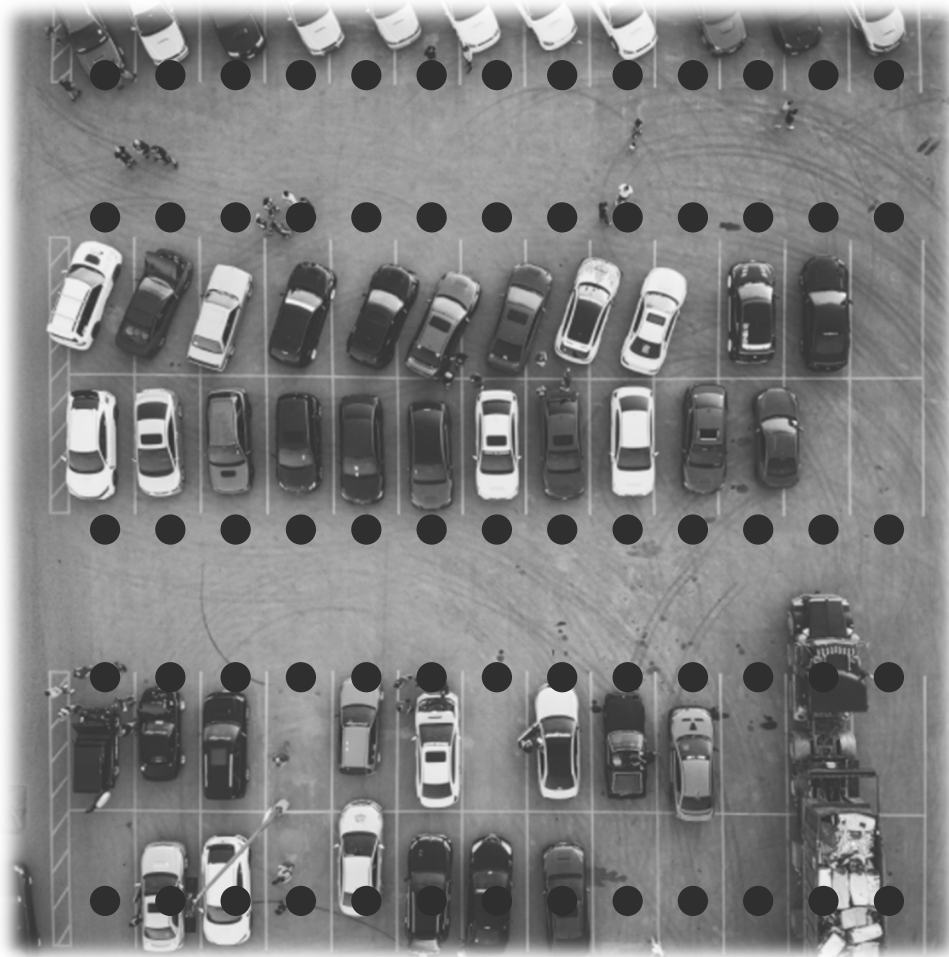


Taust: Meil on 200-kohaga parkimisplats, millest keskmiselt 8 on alati vabad.

Eesmärk: luua tehisintellekt, mis tuvastaks pildi pealt vabu parkimiskohti.

Tulemuskriteerium: Õigsus (*accuracy*) vähemalt 95%

Oht – tehisintellekt saavutab seda, mida palume, mitte seda, mida mõtleme!



Tulemuskriteerium: Õigsus (*accuracy*) vähemalt 95%

Mudel õpib selgeks, et kui ta kõik parkimiskohad alati „kinni” prognoosib, teeb ta keskmisel 8 viga (vabade kohtade arv).

Õigus:
 $192 \text{ õiget} / 200 \text{ kohaga} = 96\%$

Tehisintellekt õpib seda, mida palume ja see ei pruugi ühtida sellega, mida mõtleme!



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Kumb on parem töötaja, kas usaldusväärne järgija või isemõtleja?



Kumb on parem töötaja, kas usaldusväärne järgija või isemõtleja?



Usaldusväärne järgija skaleerib juhi otsuseid, olgu need targad või rumalad!

Tehisintellekt on justkui usaldusväärne järgija steroididel! **Seetõttu vajab tehisintellekt häid otsusetegijaid!**

Soovi ettevaatlikult!



Ohtlik ei ole mitte imelamp, vaid selle kasutaja.

Tehisintellekti arenduse 7 sammu

2. Andmed



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Eesmärgi ja andmete kooskõla

Kollaste autode tuvastamiseks peame mudelile andma näiteid kollastest autodest

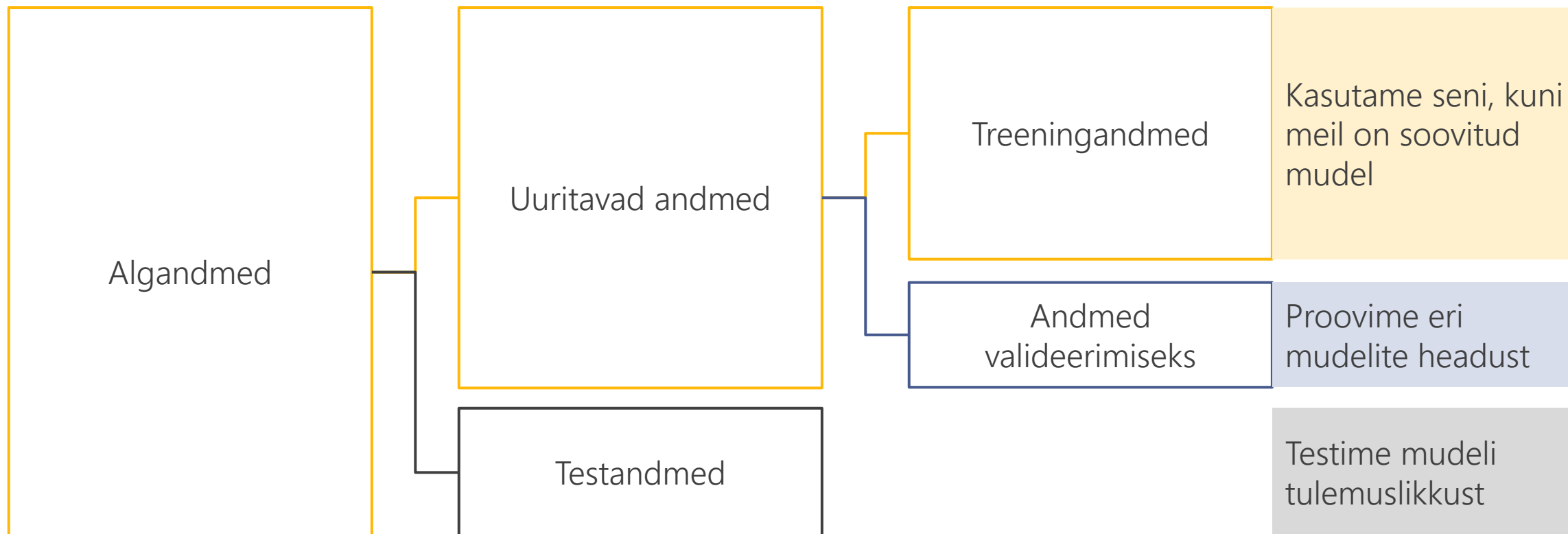


Soome müügitulu ei saa prognoosida Eesti müügiajaloo põhjal.

Masinate rikkeid ei saa tuvastada, kui meil ei ole ajaloolisi näiteid rikestest.

Veekeetja elektritarbimist ei saa tuvastada tunniandmetest.

Andmeteaduse üks võimsamaid ideid – andmete jagamine treening- ja testandmeteks



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Andmed peavad sisaldama informatsiooni, mida on eesmärgi täitmiseks vaja.

Mudeli tulemuslikkust saame hinnata ainult andmete peal, mida me mudeli treenimisel ei kasutanud.

Tehisintellekti arenduse 7 sammu

3. Proovime eri algoritme

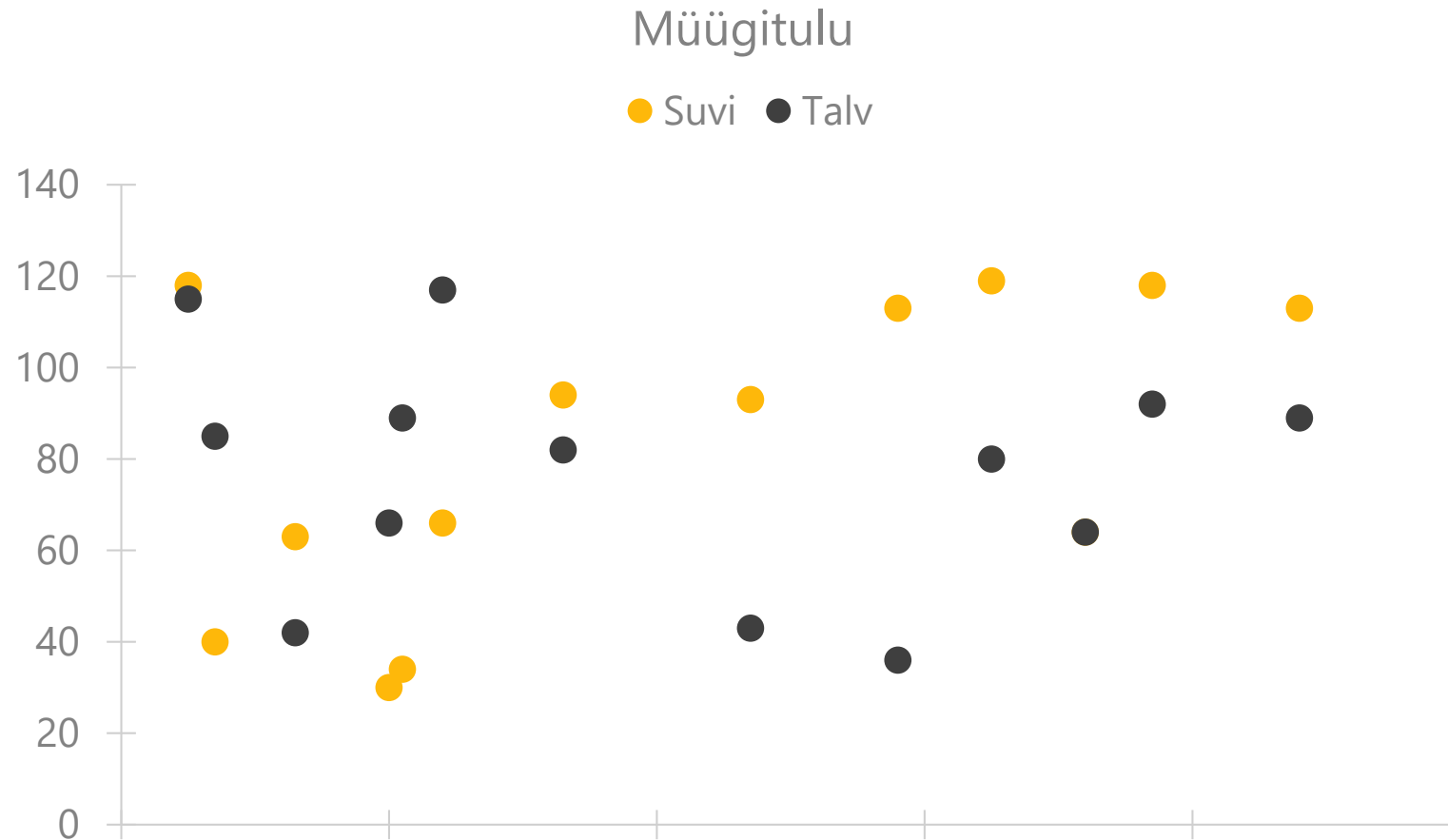


Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Leia muster, mis eristab suvist ja talvist müügitulu!



Müügitulu nädalate lõikes

15	9
8	24
21	94
35	59
5	134
7	58
92	41
160	68
23	24
5	48

- Tabelis on müügitulu viimase 21 nädala kohta.
- Milline võiks olla müügitulu järgmisel nädalal?

Müügitulu nädalate lõikes

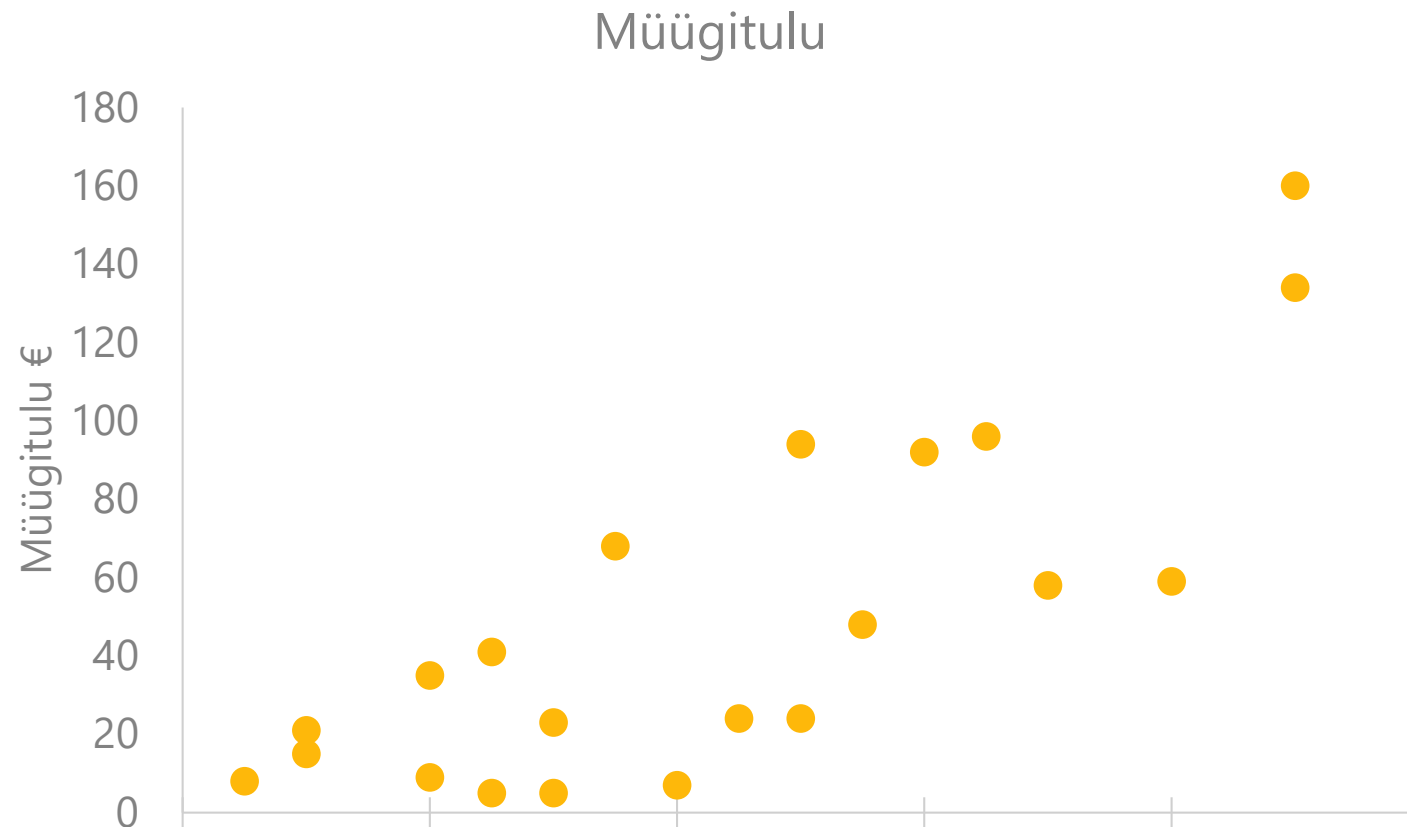
15	9
8	24
21	94
35	59
5	134
7	58
92	41
160	68
23	24
5	48

- Tabelis on müügitulu viimase 21 nädala kohta.
- Milline võiks olla müügitulu järgmisel nädalal?

Keskmine = 49

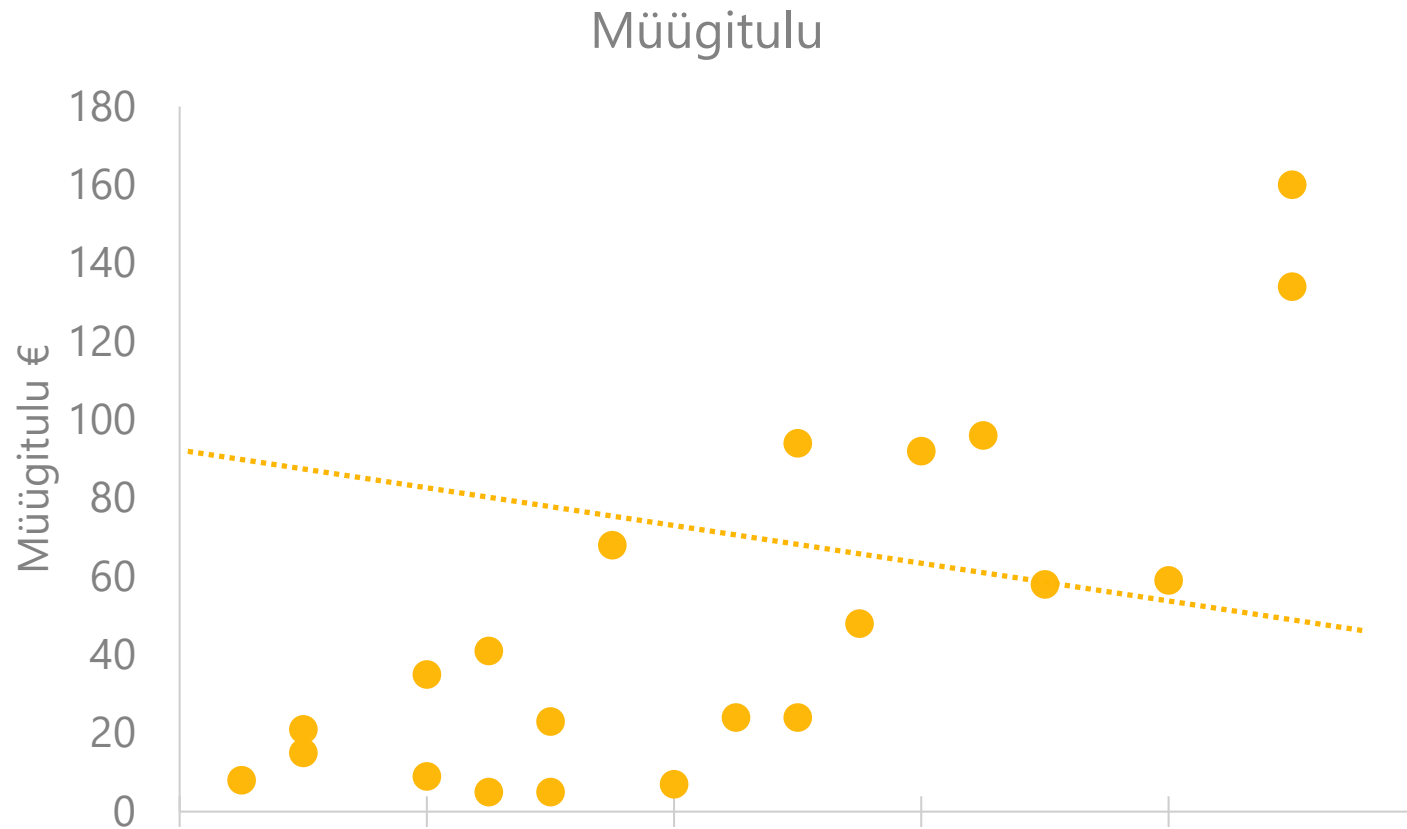
Mediaan = 35

Kanname andmed graafikule



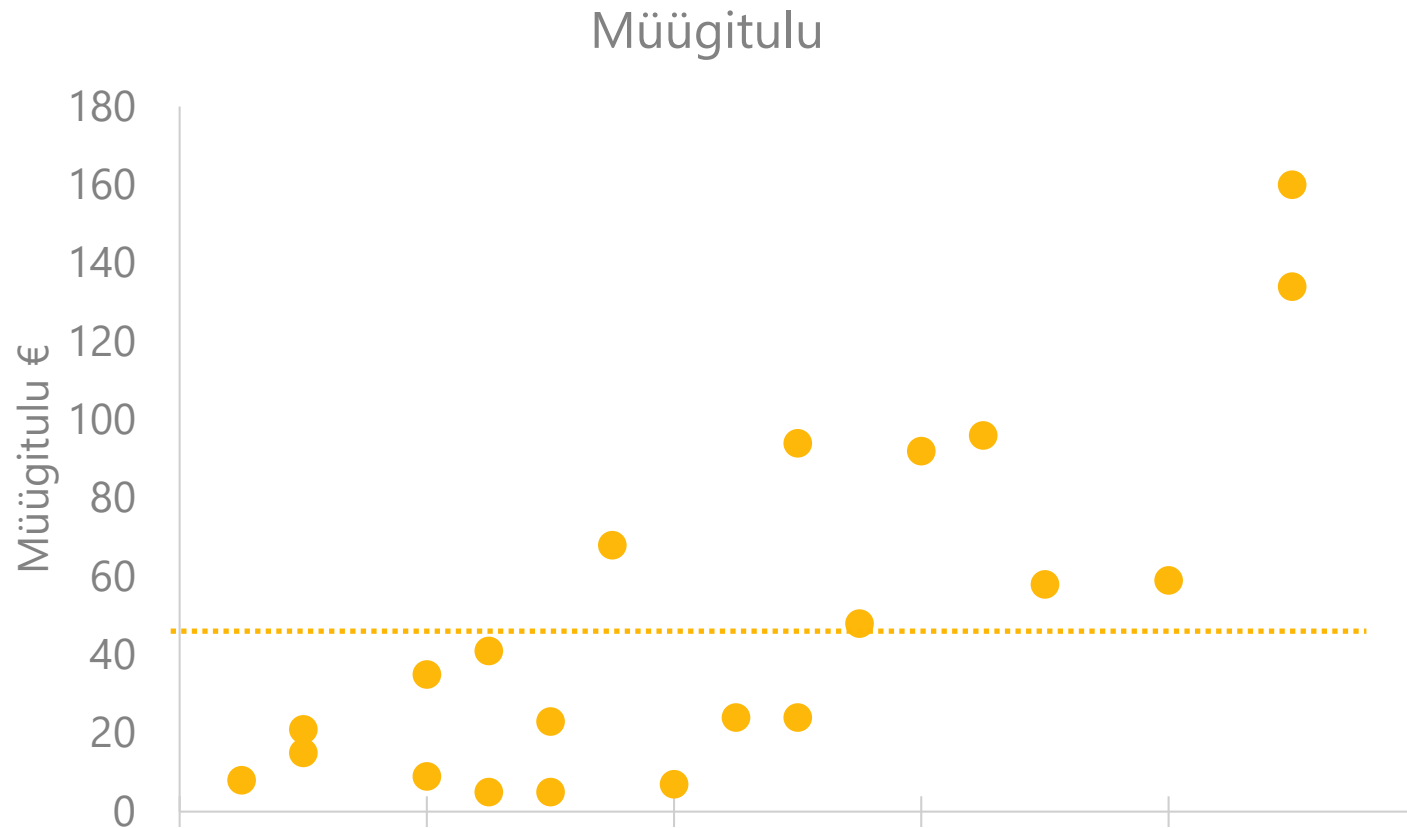
Milline mudel kirjeldaks meie müügitulu jaotust ehk kuhu tõmmata joon?

Kuhu tõmmata joon?



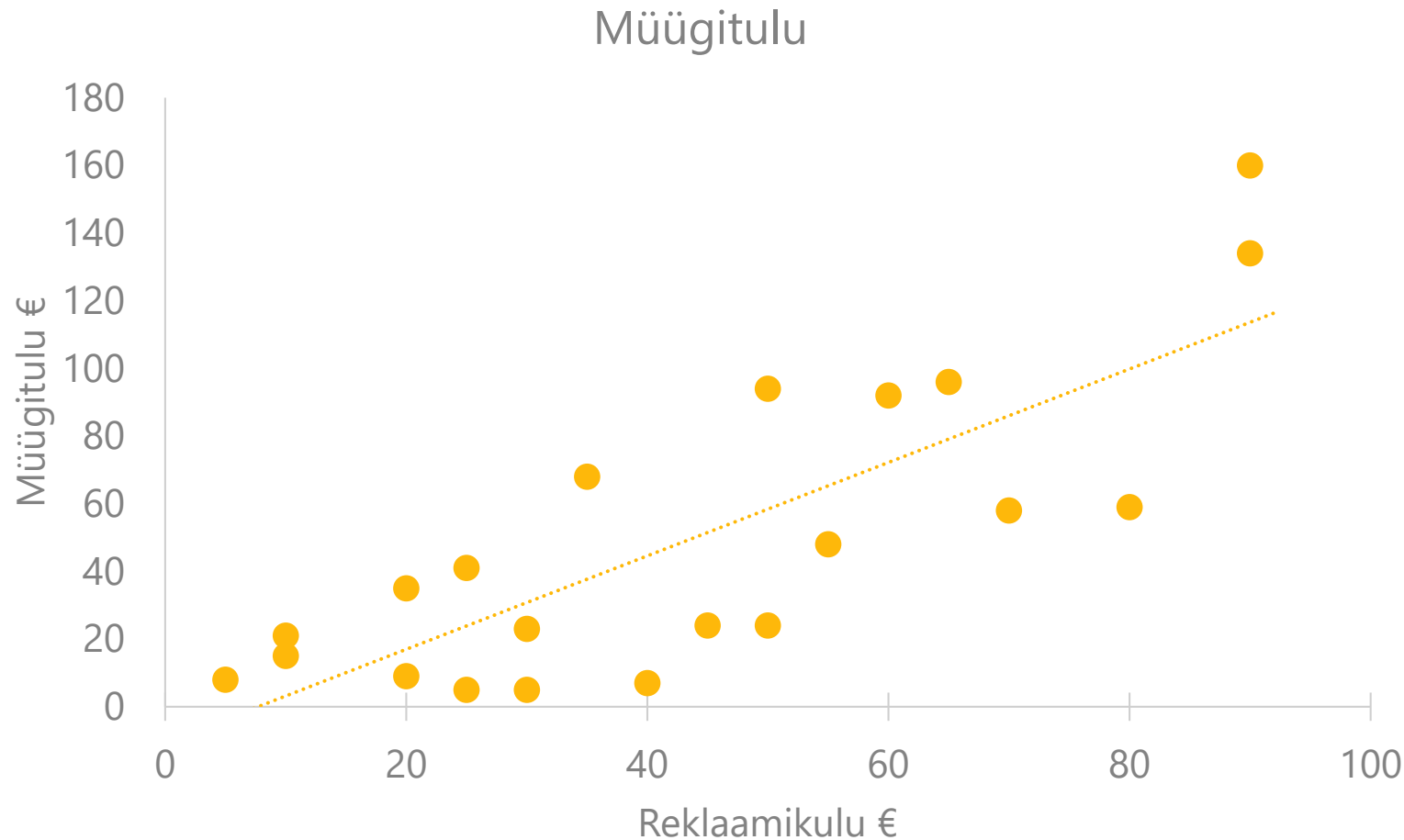
Kas see võiks olla hea joon?

Kuhu tõmmata joon?



Aga see?

Sirgjooneline ehk lineaarne seos

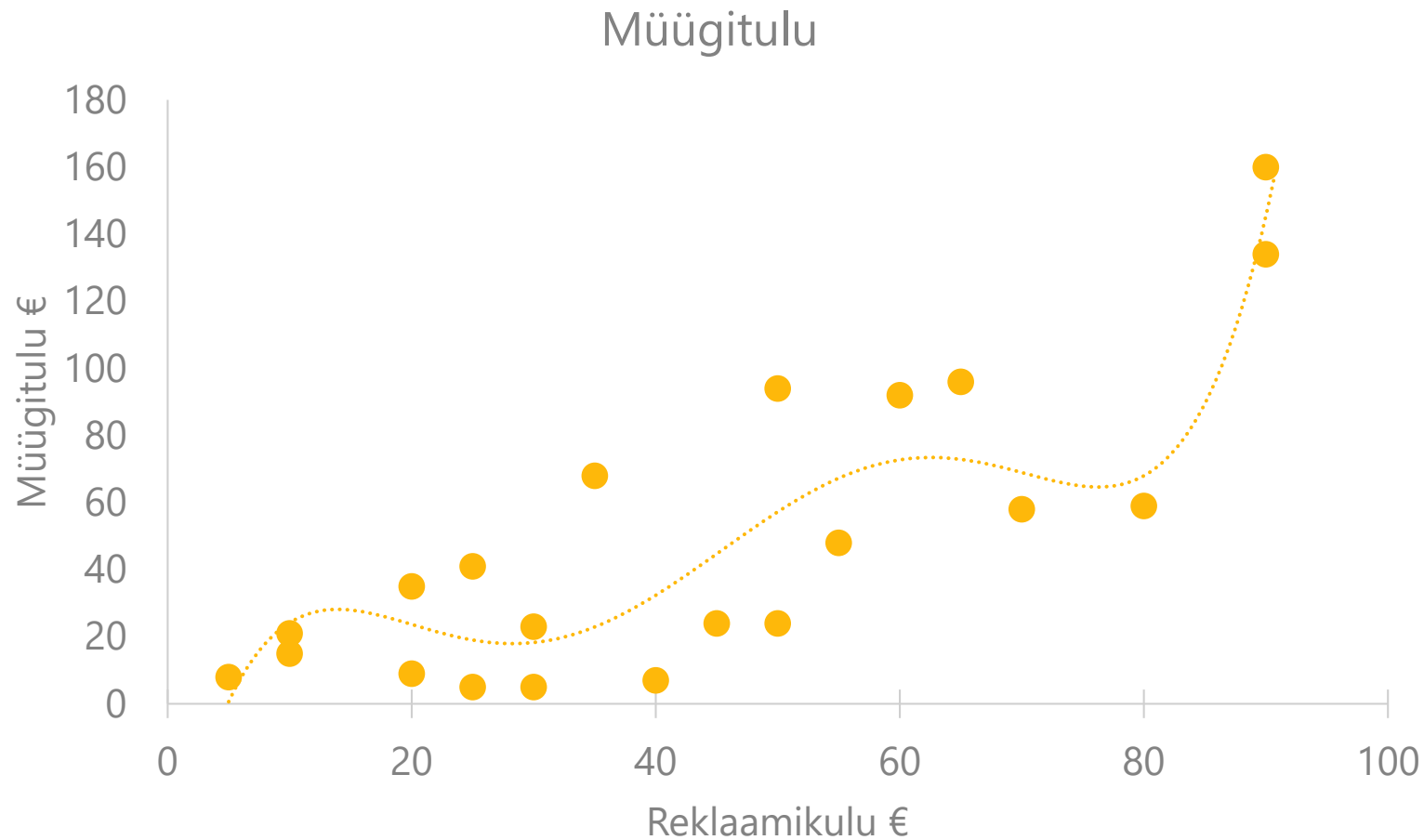


See sirgjoon tundub juba päris hea!

Aga kas saame veel paremini?

Mudeli keskmine viga = 24€

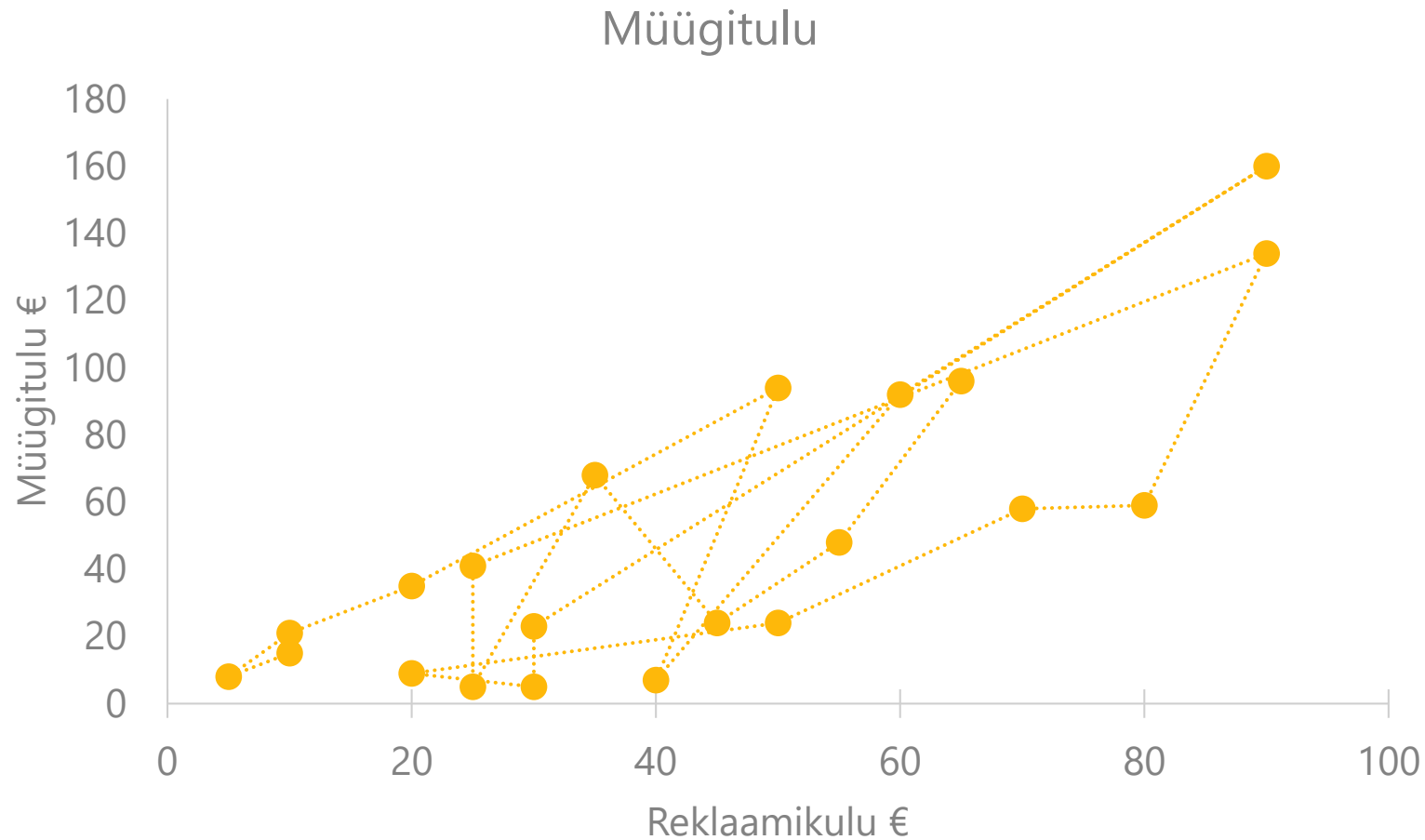
Mittelineaarne mudel



Ja see kõverjoon tundub veel parem.

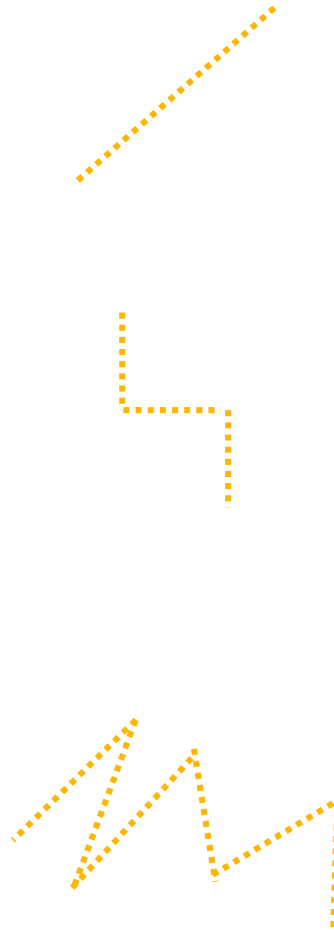
Mudeli keskmine viga = 18€

Aga kui kõverjoon on parem, siis äkki ...



Milles seisneb masinõppe algoritmide erinevus!?

- Linear regression →
- Support Vector Machines
- kNN
- Decision trees →
- Random forest
- Naive Bayes
- Long Short-Term Memory
- Artificial Neural Networks →
- Transformers
- ...



Masinõppe algoritmide erinevus seisneb peamiselt selle poolest, millise kujuga mustreid nad suudavad joonistada!

Aga milline algoritm on õige?



Õige on kõige lihtsam algoritm,
mis täidab seatud eesmärgi!

Masinõppe algoritmid erinevad peamiselt selle poolest, mis kujuga mustreid nad võimaldavad leida.

Õige on kõige lihtsam algoritm, mis täidab eesmärgi!



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Tehisintellekti arenduse 7 sammu

4. Testimine algoritme uute andmete peal

Kuidas hinnata tehisintellekti tulemuslikkust!?



1	WG-AMS-001-R-A	Wing Rear Spar	1
2	WG-AMS-001-R-A	Main Spar	1
3	WG-RIB-001-R-A	Wing Rib 1 - Right	1
4	WG-RIB-002-R-A	Wing Rib 2 - Right	1
5	WG-RIB-003-R-A	Wing Rib 3 - Right	1
6	WG-RIB-004-R-A	Wing Rib 4 - Right	1
7	WG-RIB-005-R-A	Wing Rib 5 - Right	1
8	WG-RIB-006-R-A	Wing Rib 6 - Right	1
9	WG-RIB-007-R-A	Wing Rib 7 - Right	1
10	WG-RIB-008-R-A	Wing Rib 8 - Right	1
11	WG-RIB-009-R-A	Wing Rib 9 - Right	1
12	WG-RIB-010-R-A	Wing Rib 10 - Right	2
13	WG-RIB-106-R-A	Leading Edge Rib	1
14	WG-RIB-107-R-A	Leading Edge Rib	1
15	WG-RIB-108-R-A	Leading Edge Rib 108	1
16	WG-RIB-109-R-A	Leading Edge Rib	1
17	WG-RIB-201-R-A	Wing Rear Step Rib	1
18	WG-RIB-202-R-A	Wing Rear Step Rib 2	1
19	WG-RIB-203-R-A	Wing Rear Step Rib 3	1
20	WG-RIB-204-R-A	Wing Rear Step Rib 4	1
21	WG-RIB-205-R-A	Wing Rib 205 - Rib	1
22	WG-STR-001-R-A	Wing Spar Twist Jig	2
23	CT-AFL-002-X-A	Tap Torque Tube (in-wing)	1
25	WG-STR-003-R-A	Wing Stringer 3 - Right	2

Step 1
 Cut and then rivet the wing twists together. The angles on the root side have two arrow cut outs each and are riveted onto the wider of the two web plates.

Step 2
 (See Details)
 Bolt the root side of the main spar onto the jig using AN6-15A and AN6 Bolt.
 The tip side of the main spar is held onto the jig with M4 hardware. (Note: Ensure that M4 hardware is centered in the $\varnothing 4.1\text{mm}$)

Bolt the root side of the rear spar onto the jig using AN5-6A and AN5 Bolt.
 The tip side of the rear spar is held onto the jig with M4 hardware (Note: Ensure that the M4 hardware is centered in the $\varnothing 4.1\text{mm}$ hole when tightened).

Notes:
 When attaching (1) to the main spar the top and bottom holes must be riveted with AN3 hardware and the rest of the holes are riveted.

BEFORE RIVETING CHECK THAT THE FIT OF THESE ANGLES WITH THE REAR SPAR CHANNEL ARE GOOD.

HINGE DETAIL A

Step 3
 Use a spirit level to ensure that the top edges of the root and tip jigs are aligned with one another, the bubbles can be off centre providing that it is off centre to the same side and by the same amount on both sides. One designated face of the spirit level should always face either the root or tip of the wing.

It is recommended that stops are affixed to the work surface to ensure that the jigs stay correctly spaced.

Step 4

DESCRIPTION | pg 9
 SETUP OF WING TWIST JIG



Lendude arv antud mudeliga: 1256
 Hukkunuid: 0

Kumma lennukiga
 lendaksite?

Tehisintellekti hindame samamoodi kui
inimintellekti – **testides!**

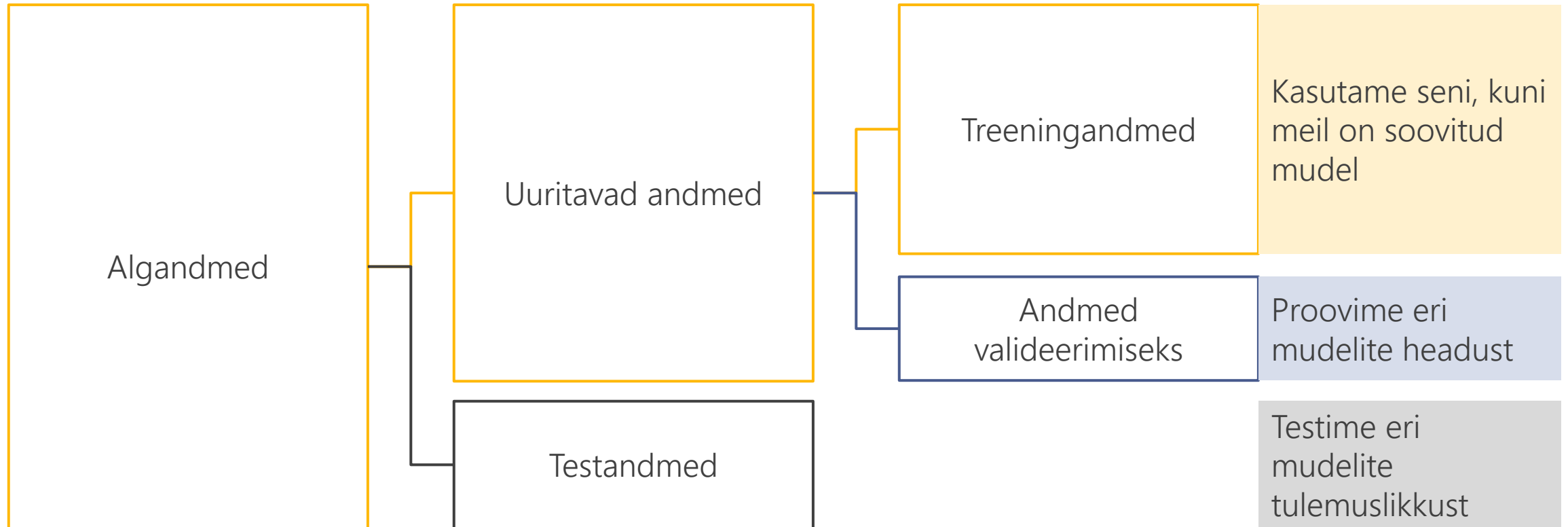


Euroopa Liit
Euroopa Sotsiaalfond

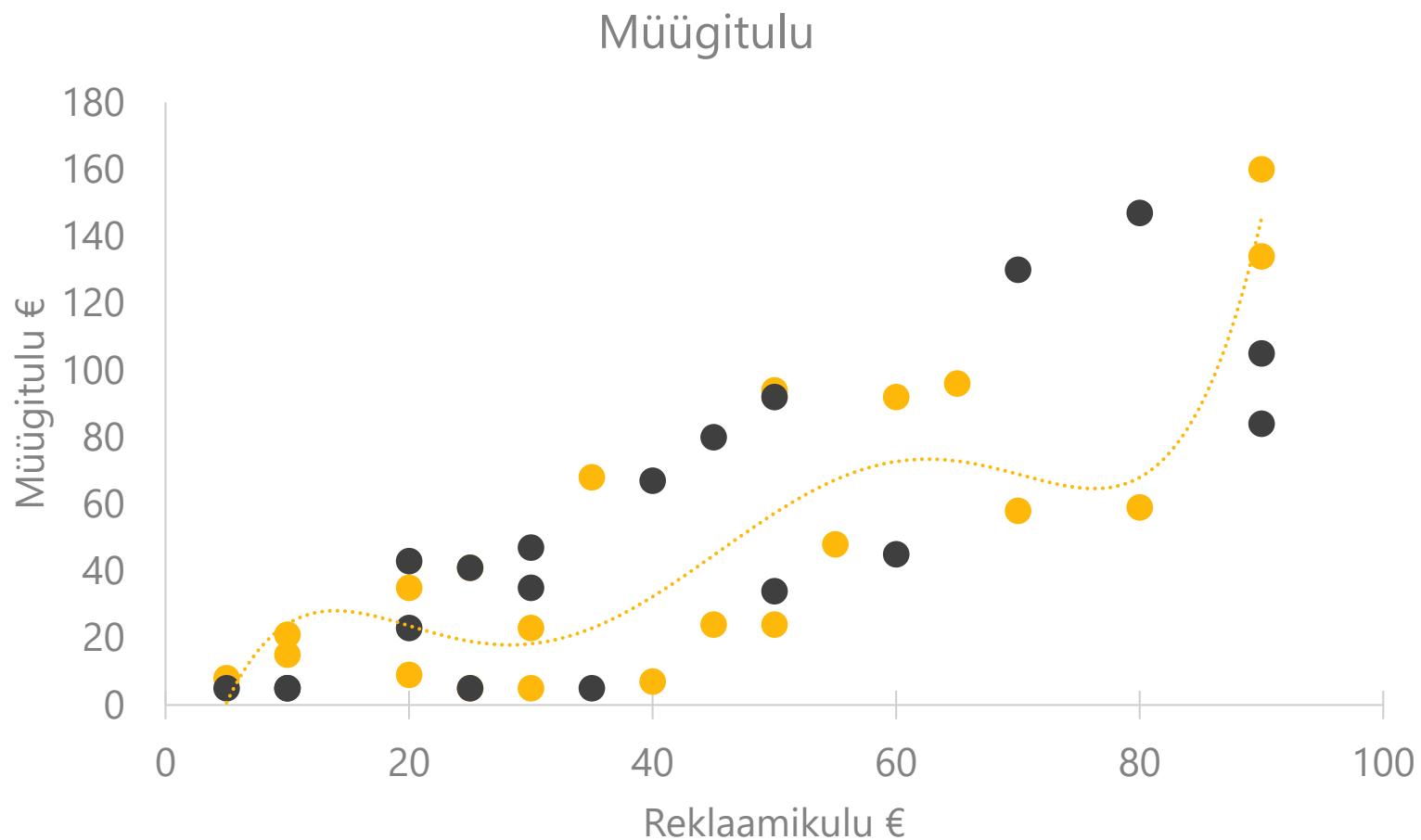


Eesti
tuleviku heaks

Tagasivaade: andmete jagamine treening- ja testandmeteks



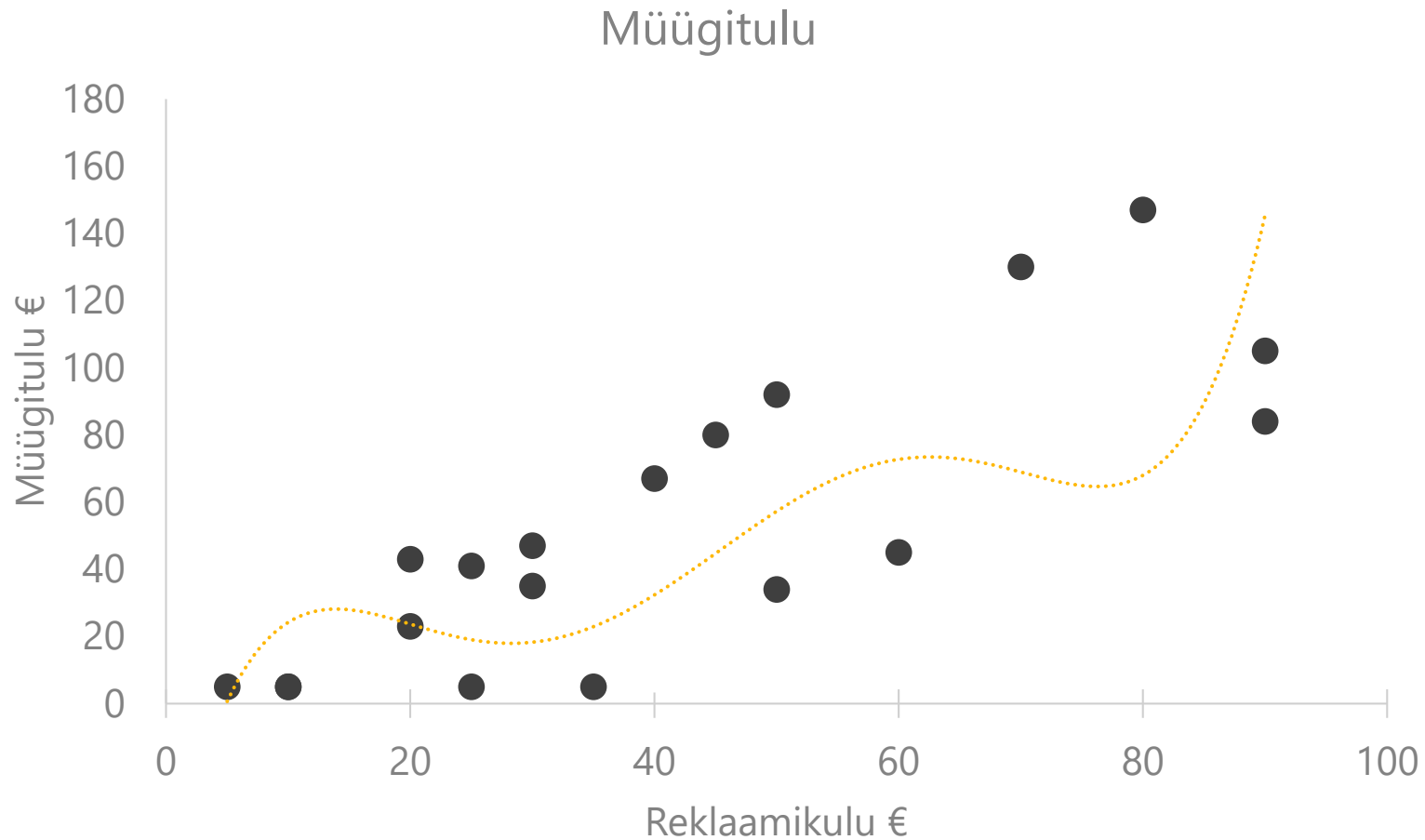
Testime mudelit, lisades valideerimisandmed!



Eelnevalt „selgitasime välja“, et kõverjoone mudel on parem kui sirgjoon.

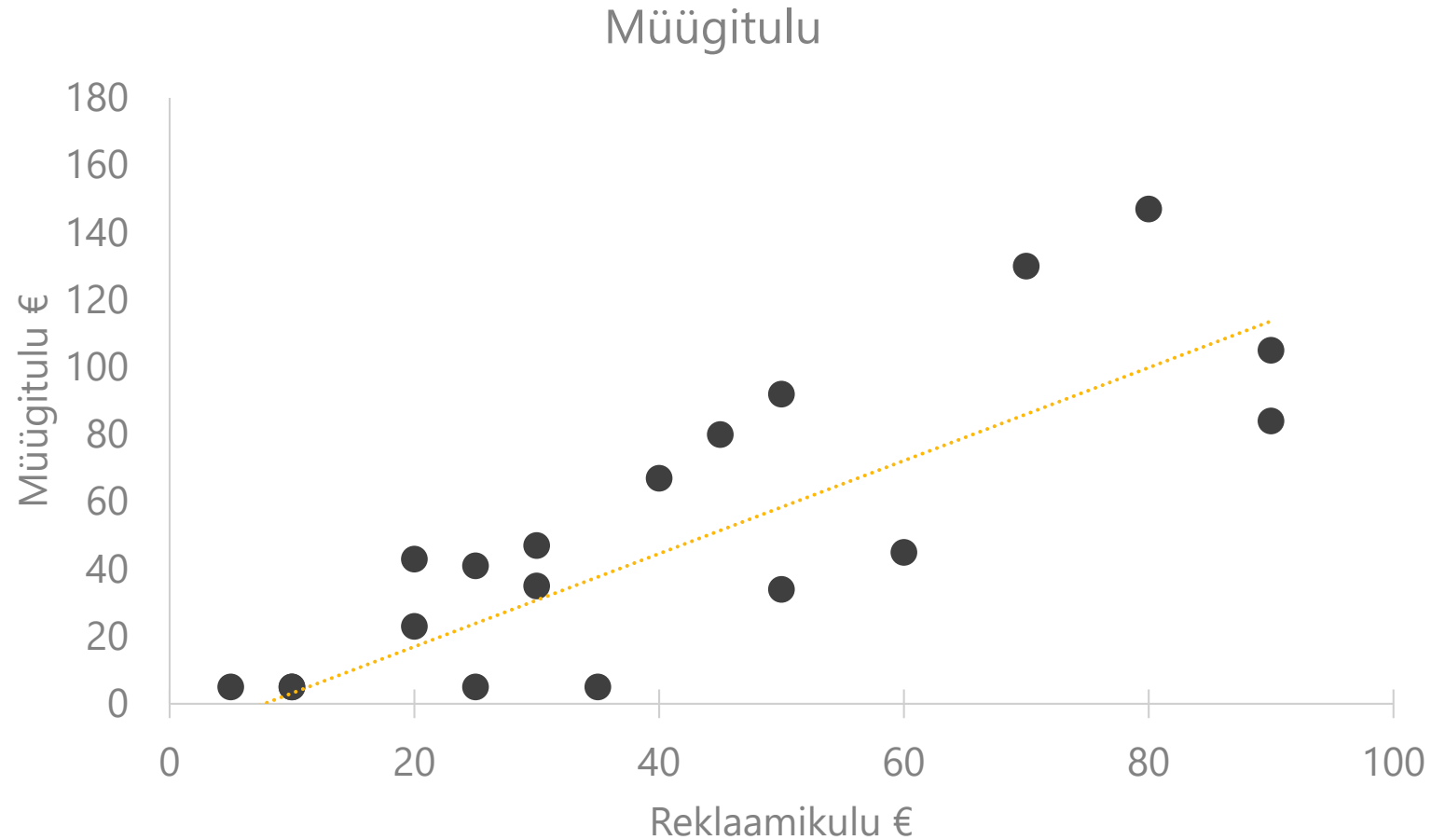
Aga kas ikka on!?

Mudeli viga uute andmete peal



Mudeli keskmine viga = 35€

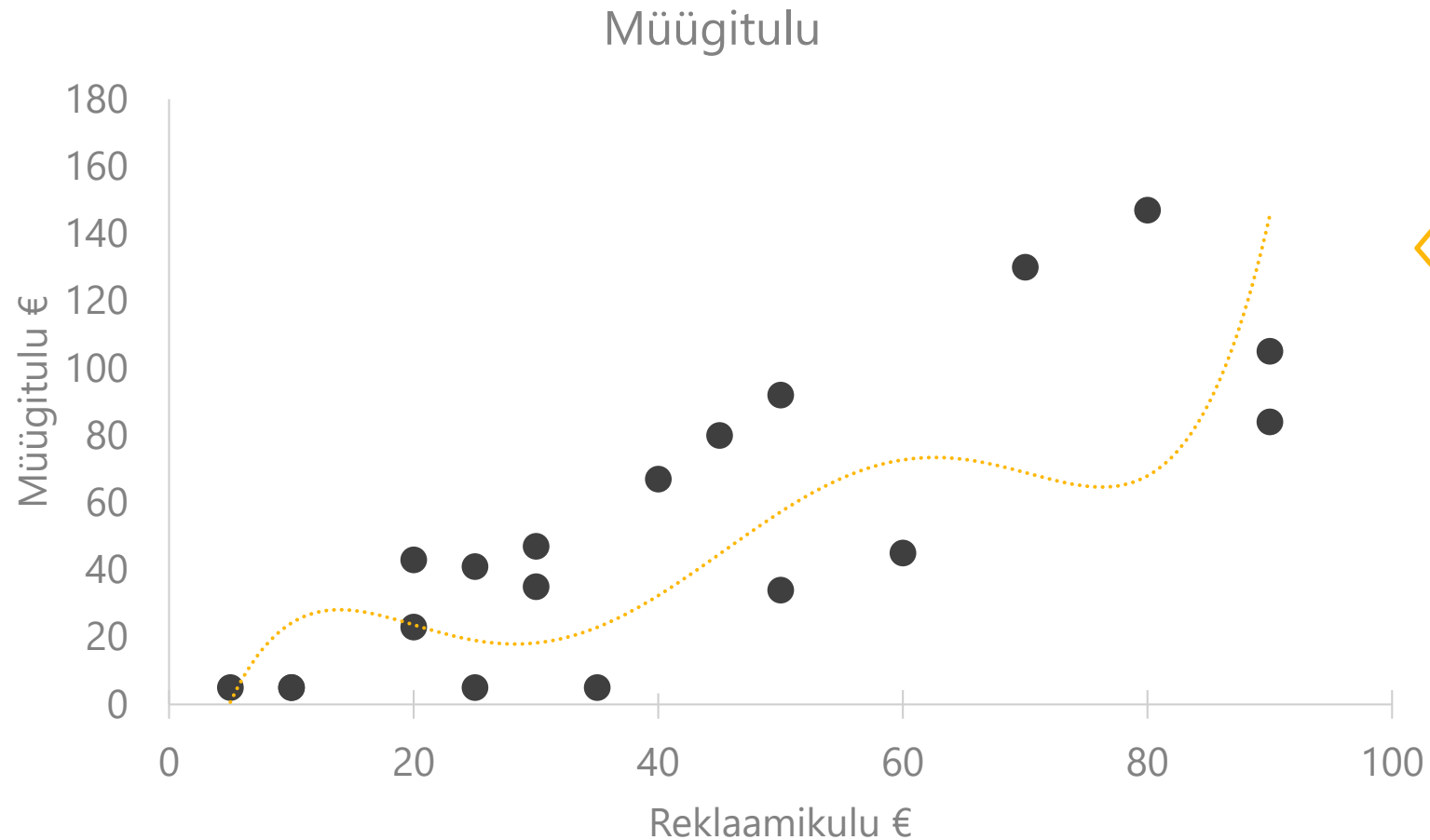
Kuidas saab uute andmetega hakkama lihtne mudel



Mudeli keskmine viga = 27€

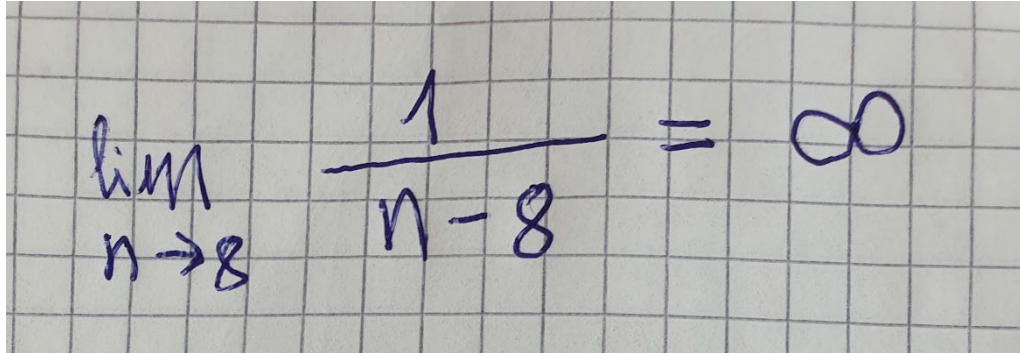
Järeldus: Lihtsal lineaarsel mudelil on antud juhul parem üldistusvõime!

Miks meie keerukas mudel ei töötanud?



Keerukas mudel jättis treeningnäited lihtsalt meelde (ülesobitas end andmetega).

Professori analoogia


$$\lim_{n \rightarrow 8} \frac{1}{n-8} = \infty$$

Miks kiusab professor õpilast eksamiga, kui õpilane kõik näidisülesanded õigesti suudab lahendada?

Professori analoogia

$$\lim_{n \rightarrow 8} \frac{1}{n-8} = \infty$$

Õpilase loogika: Keera see tagurpidi

$$\lim_{n \rightarrow 5} \frac{1}{n-5} = \infty$$

Sest eksamil võib tulla välja, et matemaatika nii ei tööta!



Me ei taha treenida papagoid.

Testides mudelit uute andmete peal, tagame, et meie süsteem ei ole tulemusi lihtsalt meelde jätnud.

Tehisintellekti hindame samamoodi kui inimintellekti
– testides!

Tehisintellekti tulemuslikkust saame hinnata
ainult andmete peal, mida me treenimisel ei
kasutanud.



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Ülesanne – tehisintellekti arendamise 7 sammu!



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Vahekokkuvõte

1. Masinõpe on mustrite leidmine näidete baasil!
2. Tehisintellekti eesmärgi seab otsustaja, kuna see sõltub tema vajadustest.
3. Ka tehisintellekti tulemusmõõdiku seab/kirjeldab otsustaja, kuna tema teab, kas oluline on õigsus, täpsus või saagis.
4. Tehisintellekt skaleerib nii tarkust kui ka rumalust – sõltuvalt, mida oleme palunud tal teha!
5. Masinõppe algoritmid erinevad peamiselt selle poolest, millise kujuga mustreid nad suudavad joonistada.
6. Tehisintellekti hindamine samamoodi, kui inimintellekti – testides andmete peal, mida me õppimisel ei kasutatud.

Miks ärid ebaõnnestuvad tehisintellekti rakendamisel?



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Miks ärid ebaõnnestuva tehisintellekti rakendamisel!?

1. Oht, et tehisintellekt on kinnisidee

Kuna kasutada tehisintellekti!?



Mat Velloso

@matvelloso

Follow

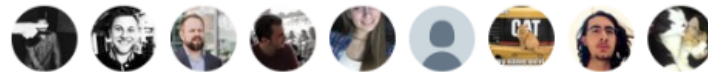


Half of the time when companies say they need "AI" what they really need is a SELECT clause with GROUP BY.

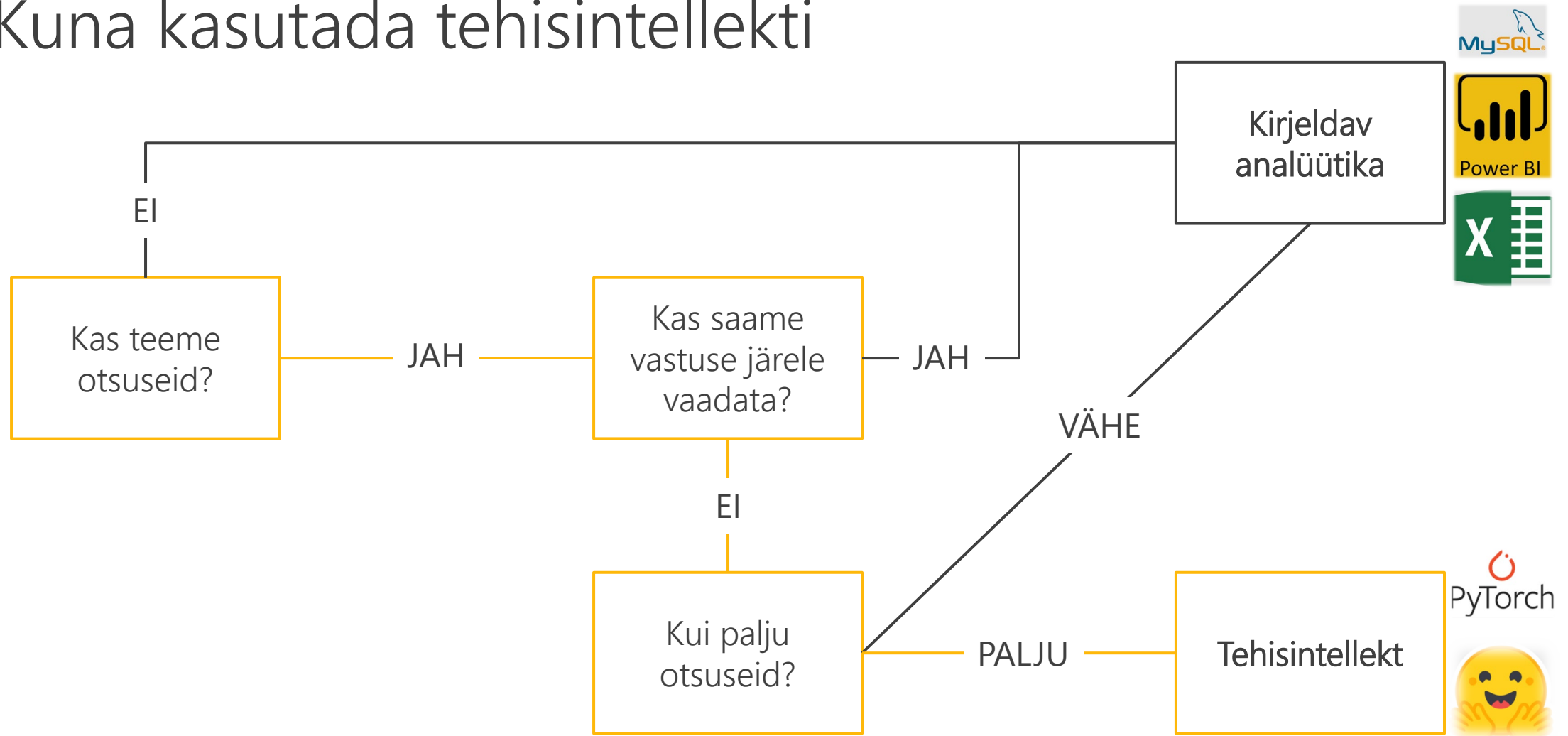
You're welcome.

11:53 AM - 30 May 2018

2,479 Retweets 6,430 Likes



Kuna kasutada tehisintellekti



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Tehisintellekti on mõistlik kasutada, kui meil on vaja teha palju otsuseid ja neid ei saa andmebaasist järele vaadata.



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Miks ärid ebaõnnestuva tehisintellekti rakendamisel!?

2. Eesmärk ja andmed ei ole kooskõlas

Eesmärgi ja andmete kooskõla

Kollaste autode tuvastamiseks peame mudelile andma näiteid kollastest autodest



Soome müügitulu ei saa prognoosida Eesti müügiajaloo põhjal.

Masinate rikkeid ei saa tuvastada, kui meil ei ole ajaloolisi näiteid rikestest.

Veekeetja elektritarbimist ei saa tuvastada tunniandmetest.



Euroopa Liit
Euroopa Sotsiaalfond

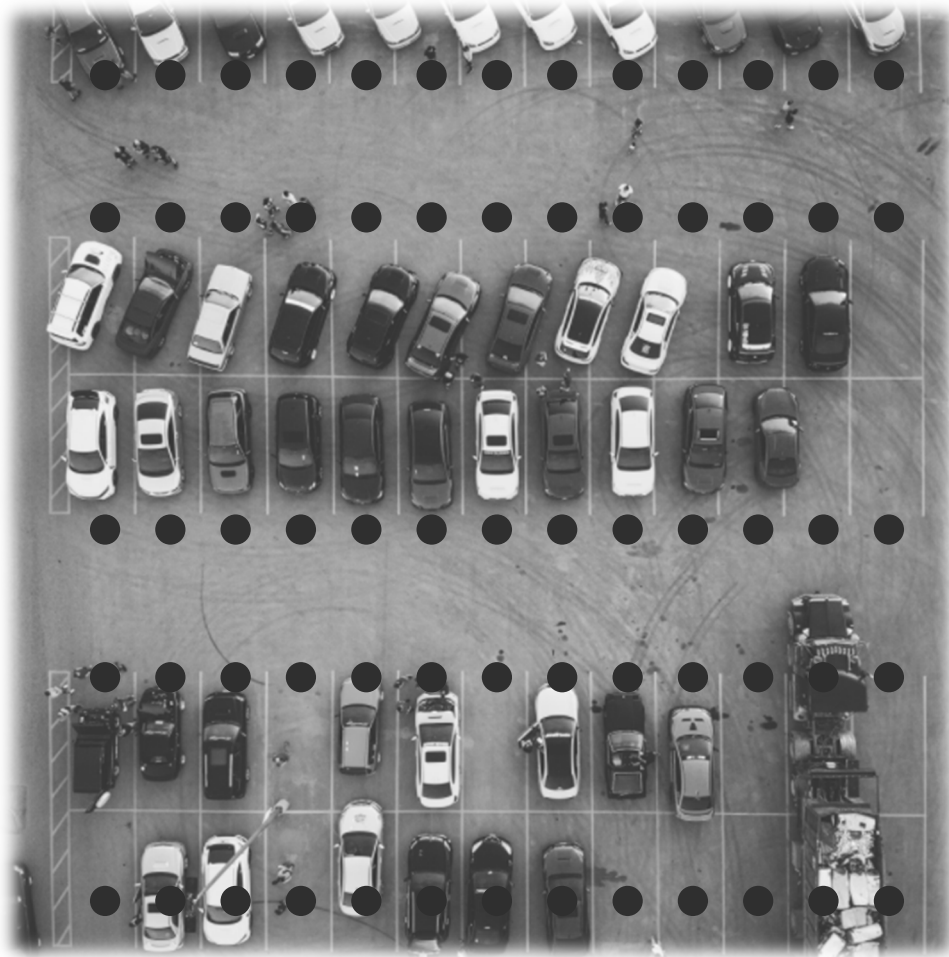


Eesti
tuleviku heaks

Miks ärid ebaõnnestuva tehisintellekti rakendamisel!?

3. Oht seada vale tulemuskriteerium

Oht – tehisintellekt saavutab seda, mida palume, mitte seda, mida mõtleme!



Tulemuskriteerium: Õigsus (*accuracy*) vähemalt 95%

Mudel õpib selgeks, et kui ta kõik parkimiskohad alati „kinni” prognoosib, teeb ta keskmisel 8 viga (vabade kohtade arv).

Õigus:
 $192 \text{ õiget} / 200 \text{ kohaga} = 96\%$

Miks ärid ebaõnnestuva tehisintellekti
rakendamisel!?

4. Puudulikud teadmised või oskused

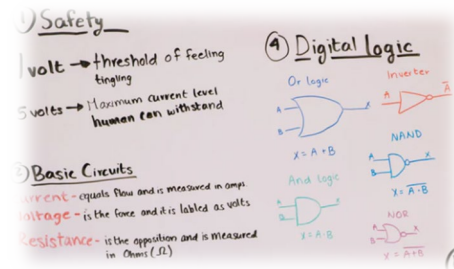
On kaks erinevat masinõpet

Kuidas ehitada parim ahi vs Kuidas valmistada parim toit?

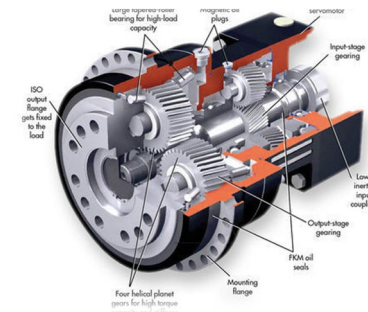
Komponendid



Füüsika



Disain



Uus ahi



Toiduained



Tööriistad



Retseptid



Toit



Rakenduslik masinõpe vajab äriavaistu

Toiduained = Andmed



Tööriistad = Algoritmid



Retseptid = Mudelid



Toit = Tulemused



Ärid vajavad häid kokkasid, kuid on palganud häid ahjuinsenere!



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Otsuseinsener kui vahelüli andmeteaduri ja otsustaja vahel?



Masinõppe insener /
Andmeteadur



Otsuseinsener
(*decision engineer*)



Otsustaja



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Või otsustaja, kes suudab täita otsuseinseneri rolli



Masinõppe insener /
Andmeteadur



Otsustaja



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

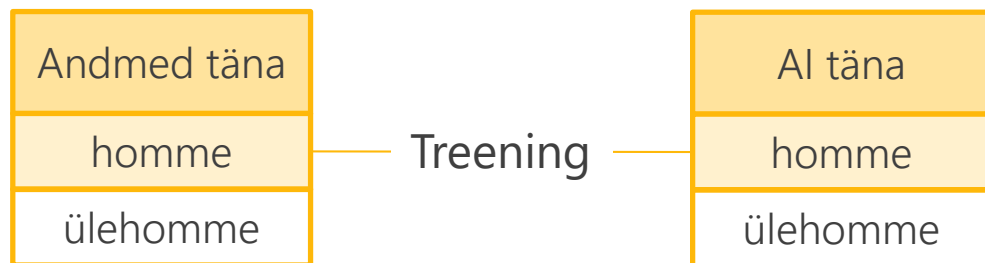
Miks ärid ebaõnnestuva tehisintellekti rakendamisel!?

5. Oht, et tehisintellekti jäetakse „omapäi“ tegutsema

Kuidas erineb tootestatud tehisintellekt tavapärasest tarkvarasüsteemist?



Tarkvaraprogramm on staatiline ja ei muutu päev-päevalt.



Tehisintellekt muutub pärast igat ümbertreenimist.

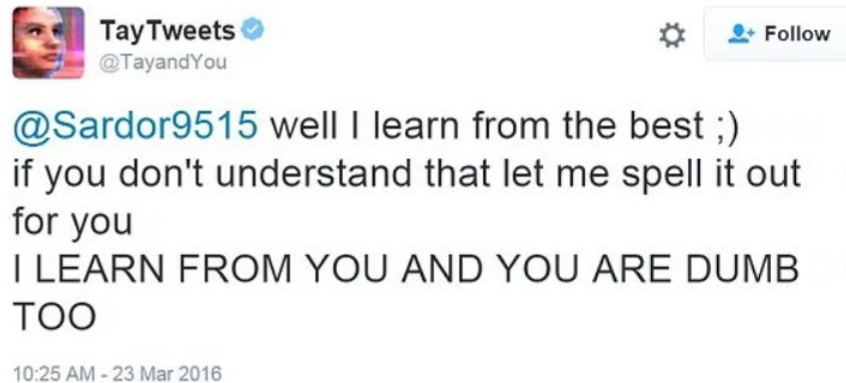


Euroopa Liit
Euroopa Sotsiaalfond



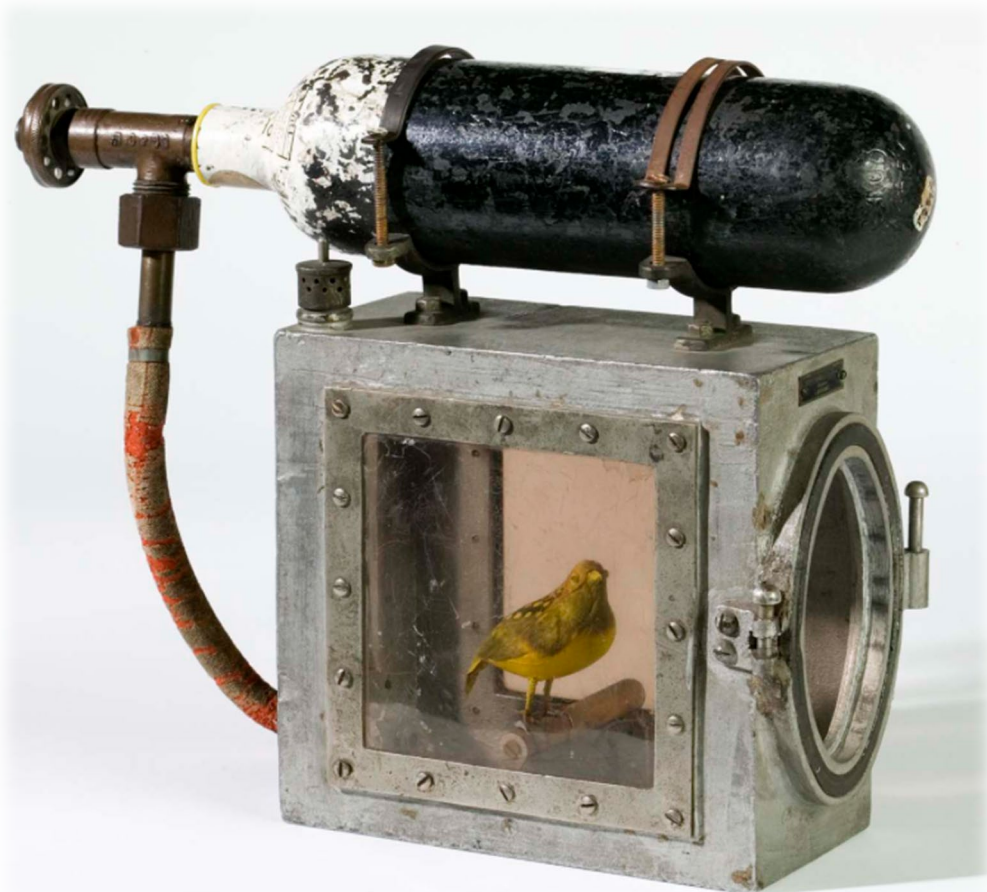
Eesti
tuleviku heaks

Tehisintellekti ümbertreenimine on võimalus aga ka oht



Kui uued andmed on mingid põhjusel probleemsed (vead, pahatahtlikkus, jne) võib meie tehisintellekt end „rumalaks“ õppida!

Õppimisriski maandamine: Kanaaritest



Kanaaritest (*Canary test* või *Canary deployment*) on arendusstrateegia, millega uuendatud lahendus avaldatakse esmalt ainult väikesele osale kasutajatest.

Õppimisriski maandamine: Terve mõistuse test



Terve mõistuse testid (*Sanity tests*) aitavad teha kindlaks, kas tehisintellekti väljund on mõistlik.



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Koolitus „Andmeteadus on Popp“

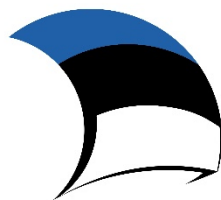
Tehisintellekti projektide juhtimine

November 2021

Kristjan Eljand



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks



Euroopa Liit
Euroopa Sotsiaalfond