

Koolitus „Andmeteadus on Popp“
Prognoosimine ja
mustrite leidmine
andmetest

November 2021

Kristjan Eljand



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Prognoosimine

„Prognooside tegemine on väga raske, eriti tuleviku kohta!”

(Niels Bohr)

„Prognoosimine on nagu kinnisilmi auto juhtimine, järgides juhiseid, mida annab tagaaknast välja vaatav sõber”

(Anonüümne)



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Prognoosimine aitab muuta otsustushorisoni

1. Kirjeldav analüütika – milline oli tulemus minevikus!? (Power BI, Tableau, jne).
2. Ennustav analüütika – milline saab olema tulemus tulevikus!?
3. Soovitav analüütika – mida peaksin tegema, et tulemus tulevikus oleks maksimaalne!?

Prognoosid toovad otsustushorisoni minevikust lähitulevikku.

Mida paremad prognoosid, seda kaugemale saame viia otsusehorisoni. Teisisõnu, **saame teha sama otsuse varem ja tõsta seeläbi äri arengukiirust!**

Mida saame prognoosida: numbriline prognoos

Nädal	Kampaania	Pühad	Allah %	Müük
1	0	0	0%	50
2	1	0	-7%	80
3	0	0	-5%	60
4	0	0	0%	55
5	1	1	-7%	90

Müügi maht on nõ „pidev“
tunnus – see võib omada
kõiki väärtusi.

Mida saame prognoosida: kategooria prognoos

Nädal	Kampaania	Pühad	Allah %	Müük
1	0	0	0%	Madal
2	1	0	-7%	Kõrge
3	0	0	-5%	Keskmine
4	0	0	0%	Madal
5	1	1	-7%	Kõrge

„Kategorilisel“
tunnusel on piiratud
arv väärtusi.

Numbriliselt esitaksime:

Madal = 0

Keskmine = 0.5

Kõrge = 1

Mida saame prognoosida

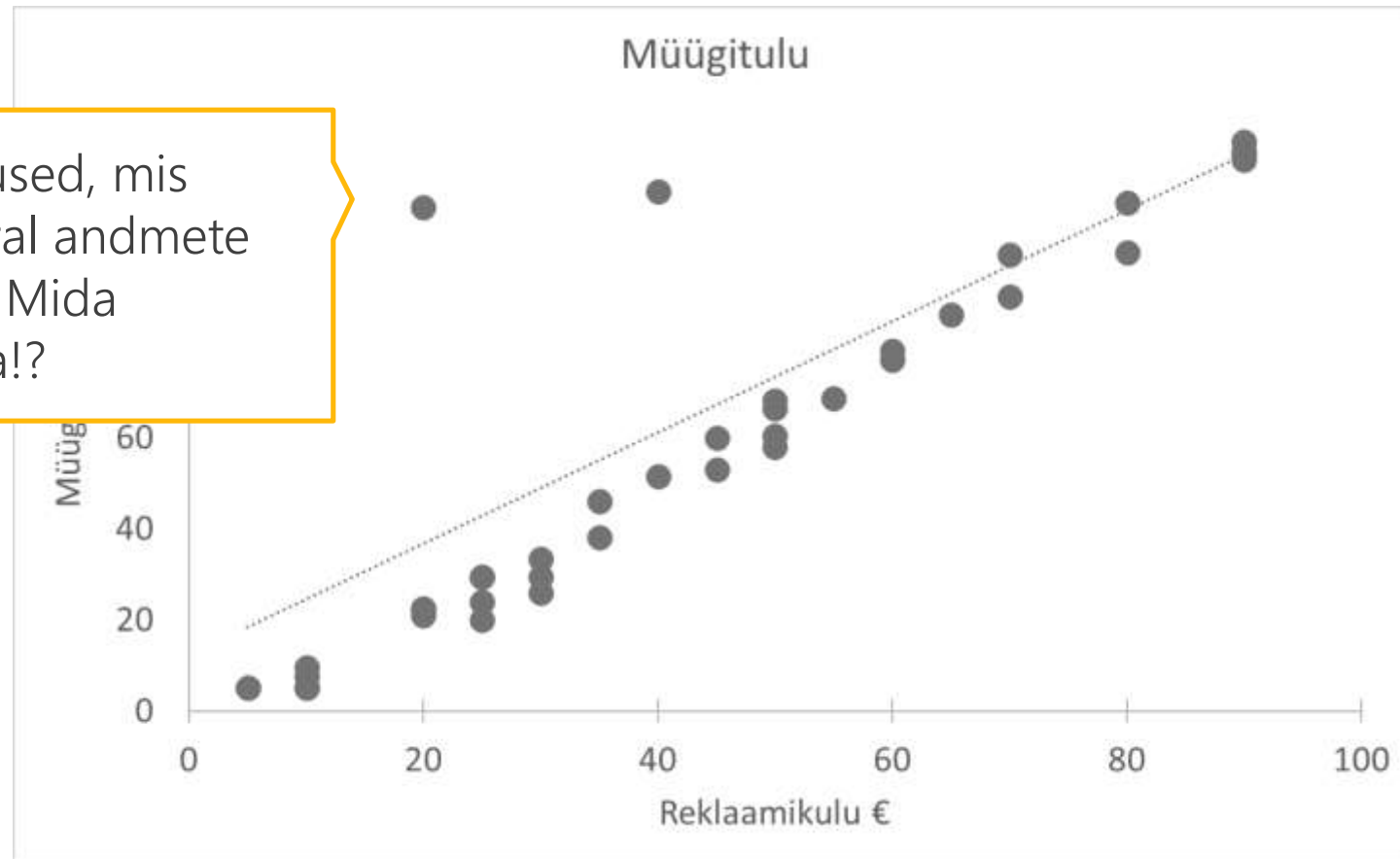
- Numbrilisi väärtusi nagu läbimüük, temperatuur, aku laetuse tase, vee tase kaevanduses, IT-abi pöördumiste arv järgmisel esmaspäeval, jne.
- Kategoorilisi hinnanguid nagu **madal, keskmine, kõrge**.
- Kategoorilisi hinnanguid nagu **läheb/ei lähe** katki järgmise kahe nädala jooksul.

Anomaaliate tuvastamine

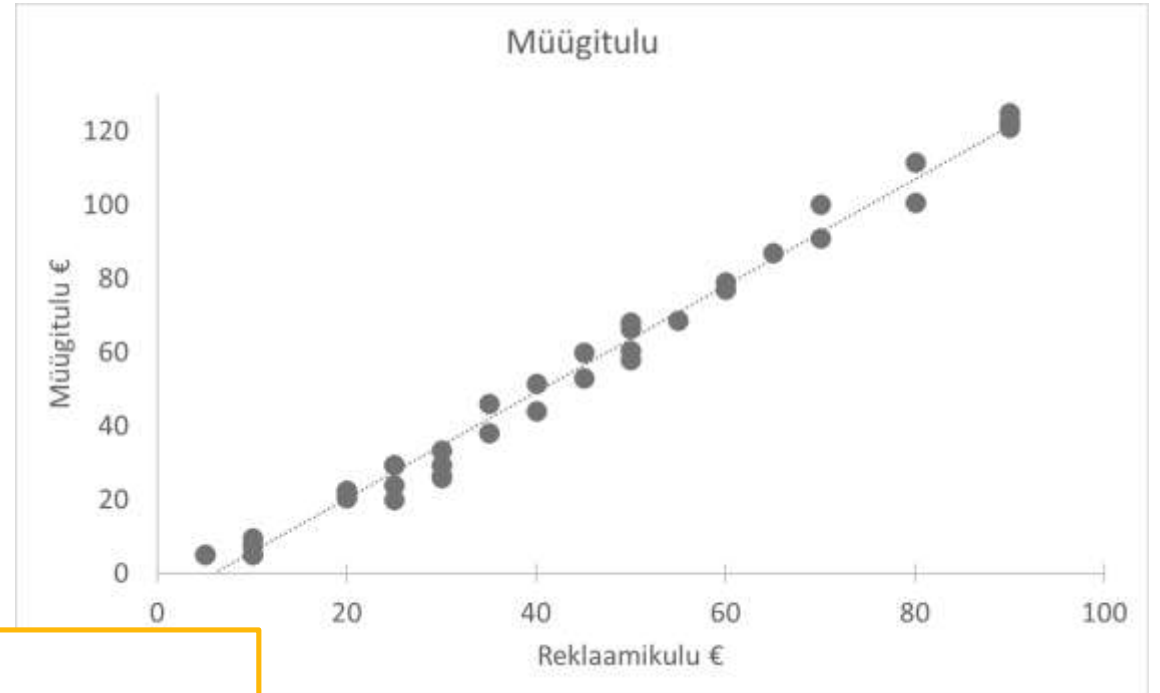
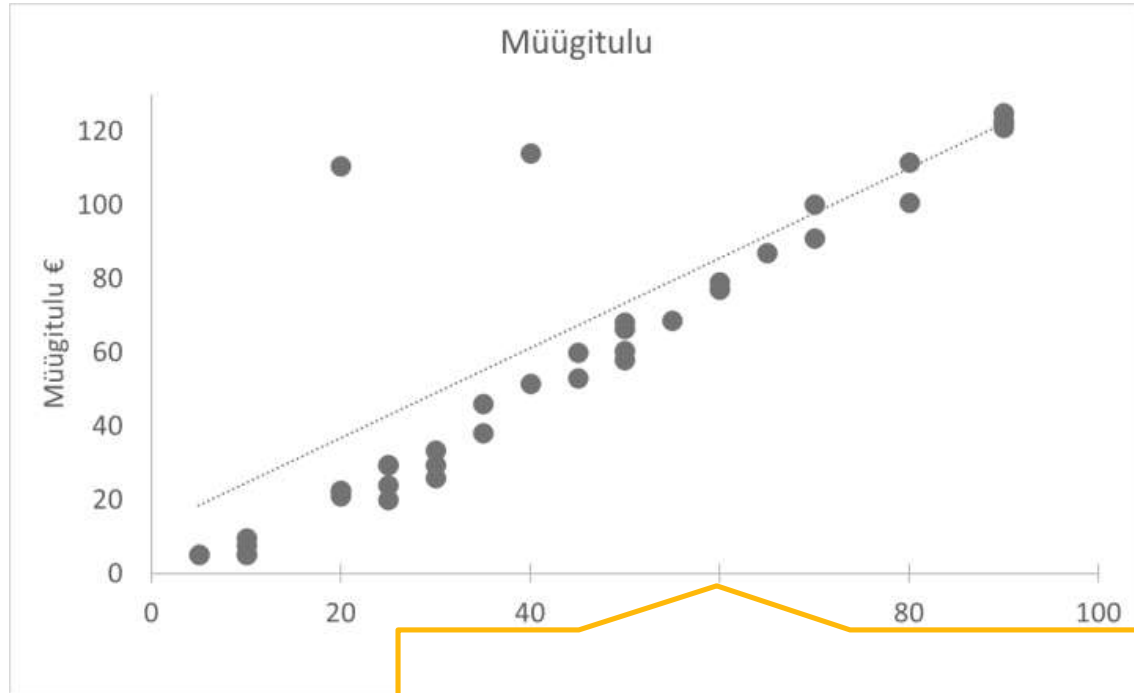
Ehk mida huvitavat võime leida äärealadelt

Anomaaliad

Anomaaliad on väärtused, mis erinevad olulisel määral andmete üldisest dünaamikast. Mida nendega peale hakata!?

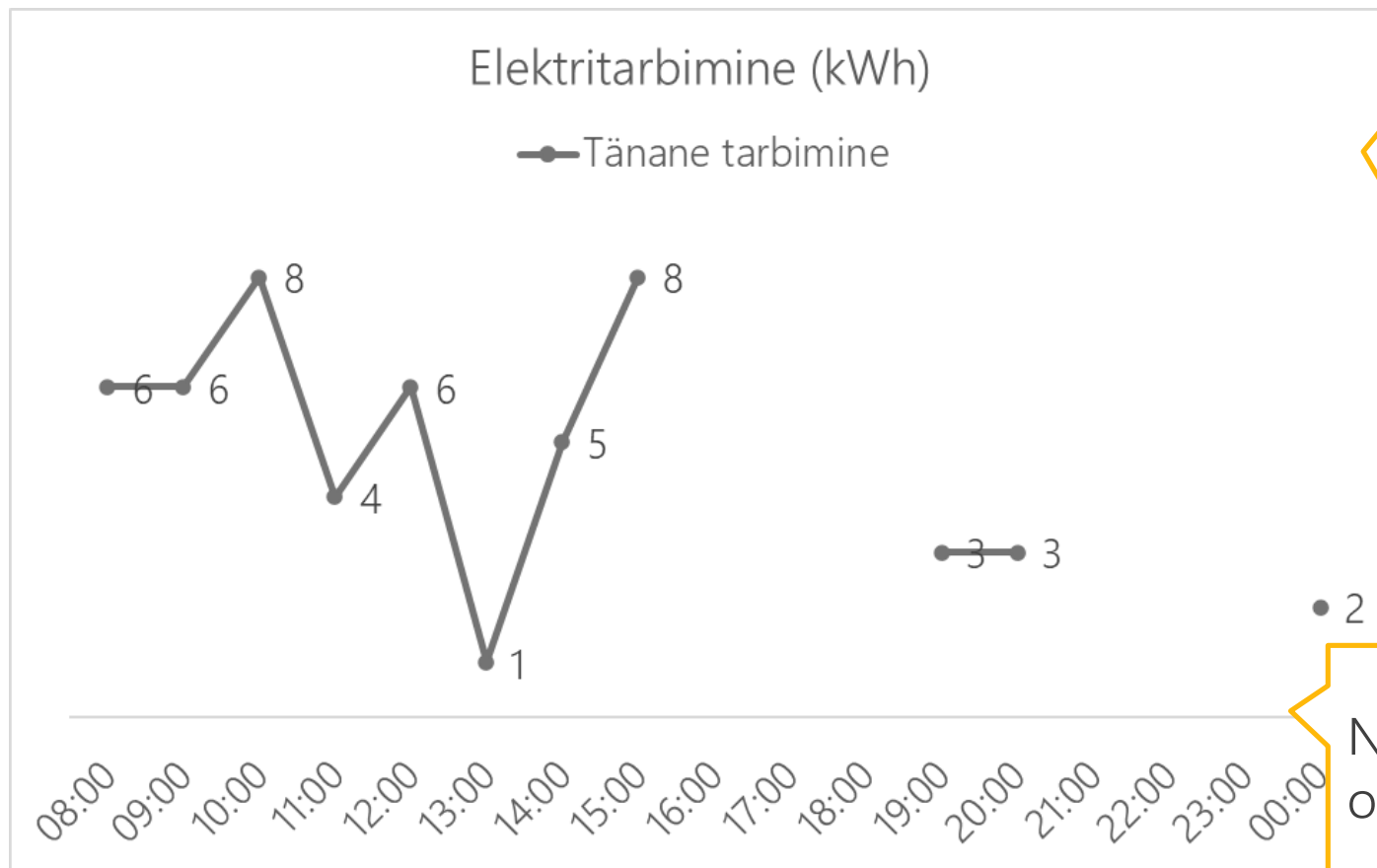


Anomaaliad prognooside tegemisel



Prognoosimudelitest eraldatakse mõnikord anomaaliad, kuna nad vähendavad prognoosi täpsust (tõmbavad prognoosijoone enda suunas)

Anomaaliate tuvastamine kui alternatiiv juhendatud õppele

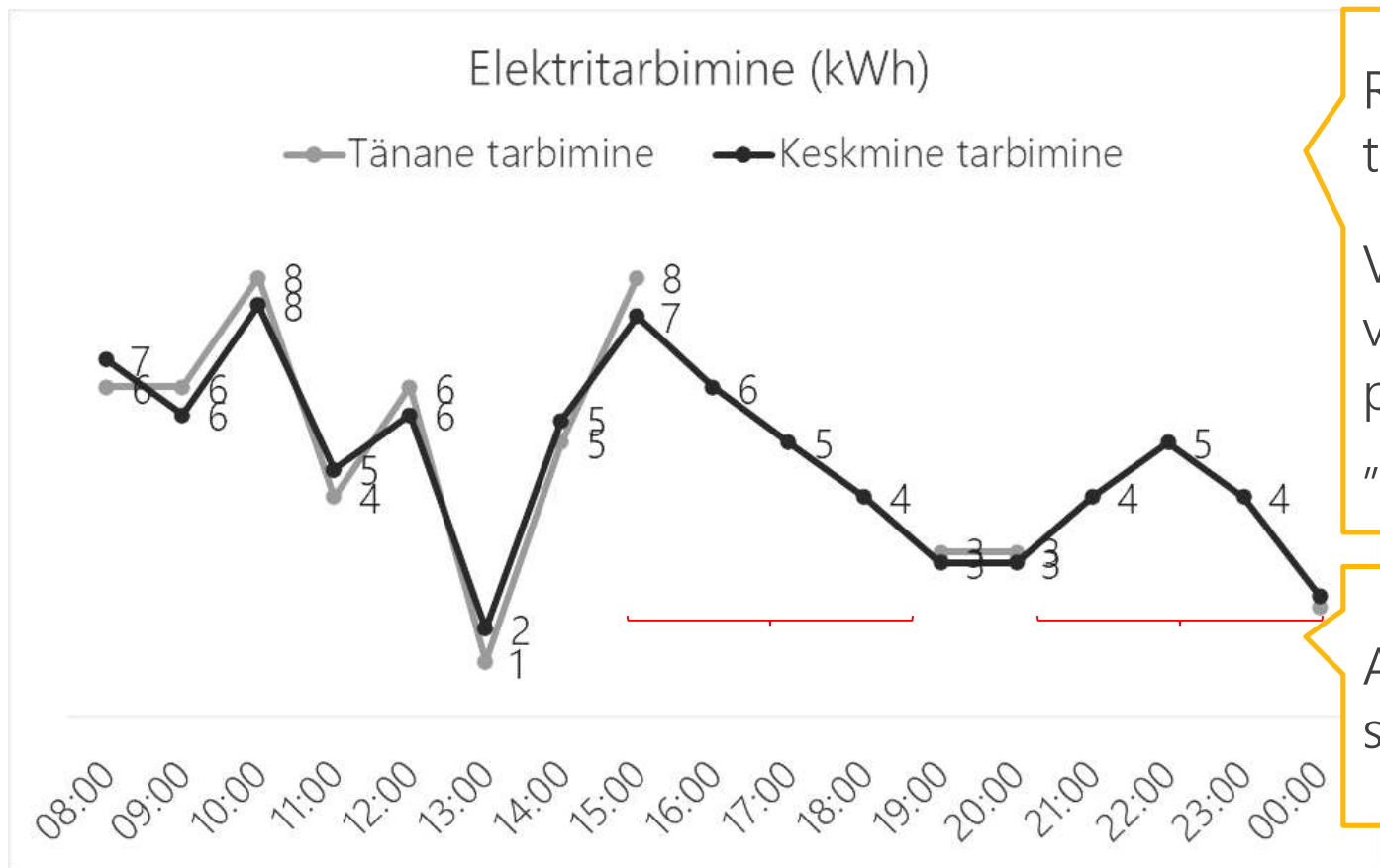


Rikete ennetamise juhendatud õppega:

Kogume näited varasemate rikete kohta
-> märgendame andmed -> treenime
masinõppemudeli leidmaks sarnaseid
mustreid.

Näiteks ütleme, et tühimikud tarbimises
on varasemalt olnud rikked.

Anomaaliate tuvastamine kui alternatiiv juhendatud õppele

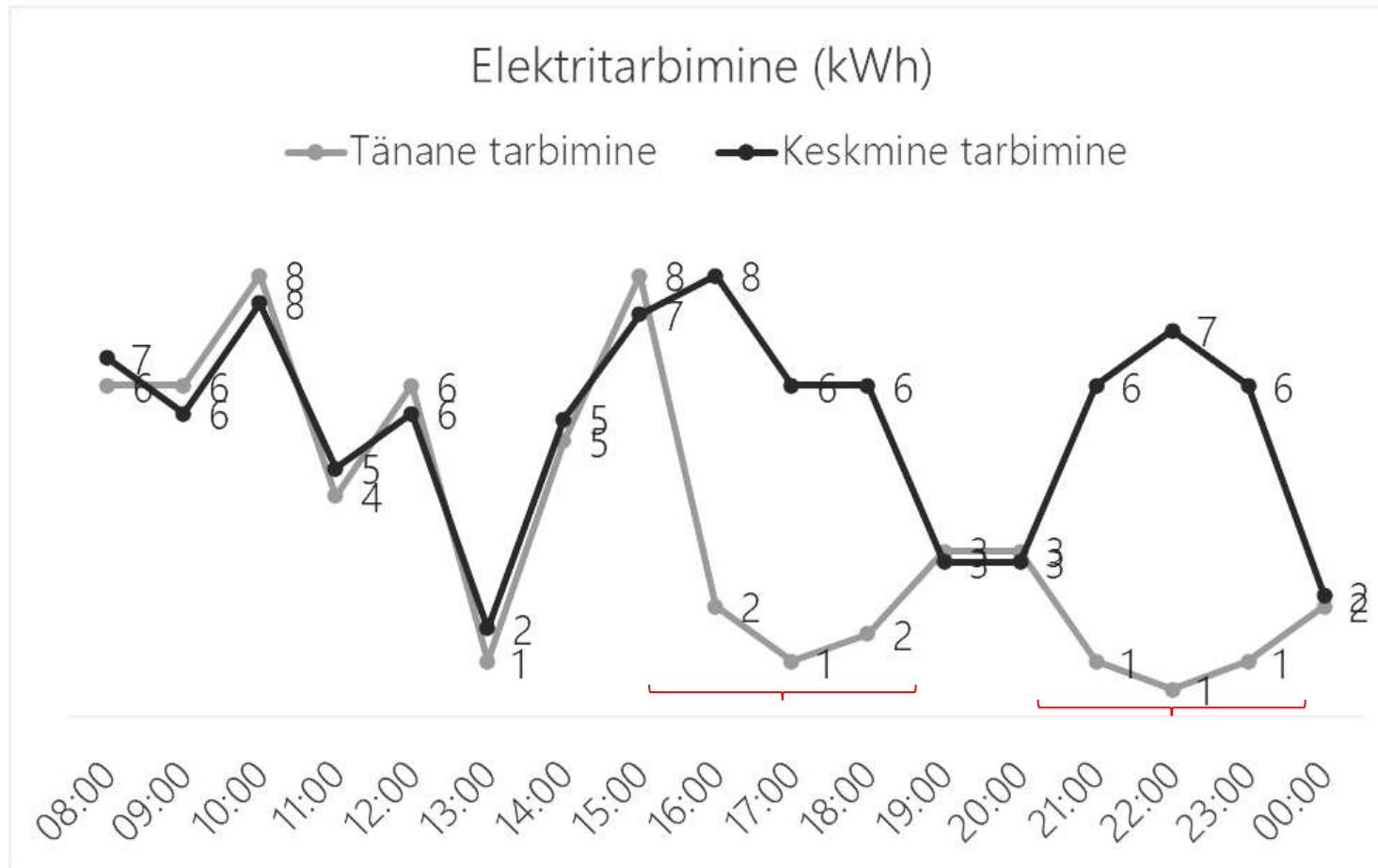


Rikete leidmine anomaaliate tuvastamisega:

Võrdleme oleviku tarbimisandmeid varasema tarbimisdünaamikaga ja proovime anda hinnangu tarbimise „normaalsusele“

Antud juhul annaks tühimikud andmetes signaali anomaalsest mustrist.

Oht: anomaalia on üldtermin



Aga ka keskmisest oluliselt erinev tarbimine on anomaalia, kuid ei viita rikkele.

Näited anomaaliate tuvastamise kasutuslugudest

1. Tootmisliinil on ummistus (normaalsus on, et ummistust pole).
2. Pinge ja voolutugevus ei vasta füüsikaseadustele.
3. Tarbimine on ebaühtlane.
4. Tarbimine erineb oluliselt tavapärasest.
5. Kliendi käitumine ei sarnane ühegi kliendigrupi käitumisega.

Koolitus „Andmeteadus on Popp“
Prognoosimine ja
mustrite leidmine
andmetest

November 2021

Kristjan Eljand



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks