



HARIDUS- JA NOORTEAMET



Euroopa Liit  
Euroopa Sotsiaalfond



Eesti  
tuleviku heaks

# Suurandmed, avaandmed ja analüütika

Õppematerjalid on loodud Euroopa Liidu Euroopa Sotsiaalfondi meetme „Kaasaegse ja uuendusliku õppevara arendamine ja kasutuselevõtt“ (2014-2020.1.03.15-0001) raames



ProgeTiiger

# Täna saad teada:

- suurandmetest,
- avaandmetest,
- andmeanalüütikast ja
- andmeteadusest.

# Suurandemed

**Suurandmed** (inglise keeles *big data*) on andmed, mis on sedavõrd suure mahuga või keerulised, et nende töötlemiseks ei piisa tavapärastest vahenditest ja ressurssidest.

## KÜSIMUS

**Kust näiteks kogutakse suurandmeid?**

# Avaandmed

**Avaandmed** on kõigile vabalt ja avalikult kasutamiseks antud masinloetaval kujul andmed, millel puuduvad kasutamist ning levitamist takistavad piirangud.

## KÜSIMUS

**Miks kõik andmed pole avaandmed?**

# Näited avaandmetest

- Eesti avaandmete teabevärav <https://avaandmed.eesti.ee/>
- Statistikaamet <https://www.stat.ee/>
- Google Booksi sõnavara Ngram <https://books.google.com/ngrams/>
- Eesti Keeleressursside keskus  
<https://keeleressursid.ee/et/keeleressursid/tekstikorpused>
- Ilmateenistus <https://www.ilmateenistus.ee/kliima/ajaloolised-ilmaandmed/>
- Sotsiaalmeedia avaandmestik <https://data.world/datasets/social-media>

# Andmeanalüütika

**Andmeanalüütika** on andmeteaduse meetodite praktiline rakendamine suurandmete automatiseeritud töötlemisel, mille tulemusena valmivad diagrammid, prognoosid ja soovitused aitavad ka ilma programmeerimis- või statistikahariduseta inimestel langetada tõenduspõhiseid otsuseid.

## KÜSIMUS

**Millised andmeanalüüsi oskused teil on?**

# Kes on andmeteadlane?

## MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

### MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

### DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



### PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

### COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

# Ülesanne

1. Tooge näiteid kus kogutakse suurandmed.
2. Valige välja üks avaandmete keskkond (näiteks eelnevast avaandmete loetelust), koostage 2-3 küsimust, millele nendest andmetest vastuse leiab. Võimalusel püüdke juurde ka vastus kirjutada.