



Aalto-yliopisto  
Perustieteiden  
korkeakoulu

# On Clustering European Energy Data (*aihe-esittely*)

*Iiro Alm*

*30.11.2022*

Ohjaaja: *Pauliina Ilmonen*

Valvoja: *Pauliina Ilmonen*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

# Tausta

- Vuonna 2021 Euroopan kaasusta 40%, kivihiilestä 46% sekä 27% öljystä tuotiin Venäjältä
- Venäjän hyökkäyksen Ukrainaan 2022 aiheuttamien pakotteiden johdosta energian tuonti Venäjältä Eurooppaan on pitkälti pysähtynyt
- Tilanne on johtanut energiakriisiin Euroopassa
  - Esimerkiksi: 1) Saksassa kotitalouksien sähkölaskut ovat nousseet jopa 185 prosenttia viime vuodesta, 2) euroalueen inflaatio nousi 10.7 prosenttiin lokakuussa 2022, 3) suurimmilta Suomen valtion omistama Fortum kärsi lähes kuuden miljardin euron häviön sijoituksessaan saksalaiseen energiayhtiöön, Uniperiin

# Tavoitteet

- Työn tarkoituksena on selvittää Euroopan maiden lähtötilannetta energiakriisiin alussa sekä analysoida maiden samanlaisuutta energiamuuttujien perusteella
- Maiden samanlaisuutta analysoidaan *K:n keskiarvon klusterointimenetelmä* -koneoppimisalgoritmilla, joka järjestää otoksen maat samankaltaisiin ja erillisiin ryhmiin energiamuuttujien perusteella
- Analyysin tuloksia voidaan peilata eri Euroopan maiden tämän hetkiseen energiatilanteeseen
  - Miten samankaltaiset maat ovat pärjänneet energiakriisin aikana?

# Rajaukset

- Datan saatavuus
  - Kaikkia energiamuuttujia ei ole saatavilla kaikille Euroopan maille
  - Uusimmat datapisteet ovat vuoden 2020 lopun lukuja: työssä pystytään selvittämään eri maiden lähtötilanne energiakriisiin mentäessä, mutta vuoden 2022 aikaisiin maakohtaisiin eroihin / ongelmiin tutkimuksella ei suoraan pystytä ottamaan kantaa datan puutteellisuuden takia

# Tietolähteet ja aineistot

- Kaikki data yhdeksälle energiamuuttujalle sekä 29 Euroopan maalle kerätään Eurostatista, Euroopan komission ylläpitämästä tietokannasta
- Energiakriisin uutisia sekä maakohtaisia energialähteitä käytetään tulosten analysoimiseen sekä johtopäätösten tekemiseen
- Metodologiaosuuteen (K:n keskiarvon klusterointimenetelmä) käytetään internet- sekä oppikirjalähteitä

# Työkalut

- Python -ohjelmointikieltä sekä sen Open Source – kirjastoja käytetään datan analysointiin ja visualisointiin
- Osa kuvaajista ja taulukoista visualisoidaan Excelillä sekä PowerPointilla
- Työ kirjoitetaan LaTeX:lla

# Aikataulu

- Datan kerääminen ja analysointi 09/2022-11/2022
- Tekstin kirjoittaminen 10/2022-12/2022
- Aiheen esittely 30.11.2022
- Valmiin työn esittely 21.1.2023