

Miten seurata kirurgisen leikkaushoidon laatua AI-avusteisesti?

Kliininen diagnostiikka ja datan hyödyntäminen osana hoitotyötä:
Varhan tapaus Clavien-Dindo

Maija Lavonius
Tommi Kauko



Auria Health
Data Services

Tehtävän esittely

Terveysthuoltolaki vaatii kirurgisen hoidon laadun valvontaa

Hallitus vaatii vaikuttavaa terveydenhoitoa pienemmällä panostuksella

- Nyt käytössä suoritemittareita, ei vaikuttavuus-/laatumittareita
- Mittareita pitää olla vain muutama
- Hoitoketjua muutettaessa laatua pitää voida seurata
 - Trendejä pitää voida seurata muutoksen aikana
- Mittaaminen ei saa viedä henkilökunnan aikaa
- Mittarin pitää kytkeytyä oikeaan hoitotapahtumaan
- Sairaalahoidon portailla tarvitaan laatu-tietoa eri tarpeisiin
 - Laatu, omavalvonta, taloushallinto, etc.
- Tarvittava tieto saatava johdon eri tasoille helposti
- Laatu-tietoa pitää voida käyttää ennakoiden jo ennen leikkauspäätöstä
- Samaa mittaristoa pitää voida käyttää eri erikoisaloilla
- Samaa mittaristoa pitää voida käyttää eri sairaaloita vertailtaessa



Auria Health
Data Services

Taustaa

Toistuva kysymys *“Olisiko mahdollista ..”*

2018 – Sote-KUVA-mittaristo eli sosiaali- ja terveydenhuollon kustannusvaikuttavuusmittaristo.

“Olisiko mahdollista, että laskisitte Tyksistä tällaisia lukuja?”

2022 – MILA eli mittaaminen ja laatu -hanke.

“Olisiko mahdollista, että laskisitte tällaisia mittareita, joita voisimme vertailla Ruotsin tavoitearvoihin [Vården i siffror; (vardenisiffror.se)]”

2023 – MILA:n jälkeen:

“[Jos alamme kirjata kertomustekstiin Clavien-Dindo-luokan,] Olisiko mahdollista, että tekisitte raportin syöpäkohtaisesti Clavien-Dindo-komplikaatioluokista?”

2025 – Sitra: Ennakoiva kirurgisten komplikaatioiden hallinta -hanke.

“Olisiko mahdollista, että kone ehdottaa klinikolle Clavien-Dindo-luokkaa ja ohjaa tarkistamaan luokituksen määräävät yksityiskohdat?”

Potilastieto ja koneellinen käsittely

Laadunvalvonta toteutettiin ensin käsinkirjaamalla Clavien-Dindo -luokka sairaskertomukseen. Siitä saatiin tämän hankkeen pohjaksi liki 5000 klinikon tekemää oikeaa Clavien-Dindo-kirjausta. Kirjaukset on tehty vuosina 2023-2025 gastrokirurgian yksikössä.

Aurian analyttisessä tietovarastossa on tietoturvalisest konellisest käteltävissä Tyksin erikoissairaanhoidon potilaskertomustiedot, mitkä kattavat muun muassa diagnoosit, tehdyt toimenpiteet, laboratoriomittaukset, lääkemääräykset ja reseptit, sekä kertomustekstit.

Dataa on niin paljon, että sitä ei tarkastella manuaalisesti vaan käsitellään massa-ajona.

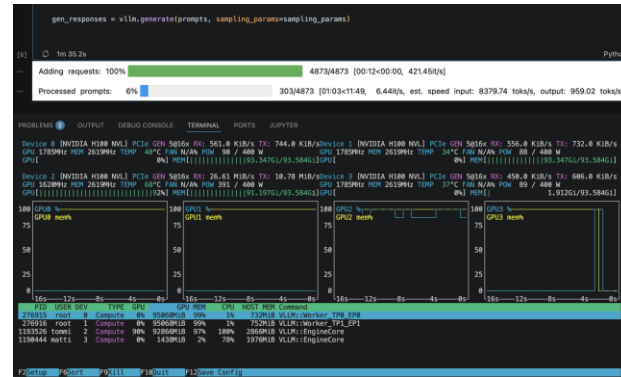
```
218
219 kohortti_data %>%
220   left_join(kohortti_pseudot, by = "henkilotunnus",
221     select(henkilotunnus, tapahtuma_alkaleima, mv_
222       nakyma_selite, hoitotapahtuma_tunnus, h
223     clavier_dindo_teksti, clavier_dindo_teksti_luoki
224     maaryksen_toimipiste, maaryksen_pvm,
225     dindo_luokitus) %>%
226   write.table(file = paste0(tallennuspolku, "/"),
227     rm(kohortti_data)
228
229 kohde_dgn_hoikojen_kaikki_palvelut %>%
230   select(henkilotunnus, palvelu_numero, varaus.n
231     yksikko_nimi, tulosyksikko, tulosyksikko_selite,
232     hoiko_alkupvm, hoiko_loppupvm, hoiko_alkamissy-
233     jatkoh_laitos_nimi, jatkoh_toimipiste_nimi, jh_s
234     dg_syy_selite, dg_oire_koodi, dg_oire_selite, pa
235     collect()) %>%
236   left_join(kohortti_pseudot, by = "henkilotunnus",
237     write.table(file = paste0(tallennuspolku, "/"),
238     rm(kohde_dgn_hoikojen_kaikki_palvelut)
239
240 kohde_dgn_radut %>%
241   left_join(kohortti_pseudot, by = "henkilotunnus",
242     select(henkilotunnus, radulahete_id, radututki
243     lahetehetki, ajanvaraushetki, tilaava_yksikko, h
244     %>%
245     write.table(file = paste0(tallennuspolku, "/"),
246     rm(kohde_dgn_radut)
247
248 kohde_dgn_laakkeet %>%
249   left_join(kohortti_pseudot, by = "henkilotunnus",
250     write.table(file = paste0(tallennuspolku, "/"),
251     rm(kohde_dgn_laakkeet)
252
253 tahystykset %>%
254   left_join(kohortti_pseudot, by = "henkilotunnus",
255     select(henkilotunnus, palvelu_numero, toimenpi
256     kayntityyppi_selite, toimenpide_koodi, toimenpid
257     vo_toimipiste_koodi, vo_toimipiste_nimi, toimenp
258     write.table(file = paste0(tallennuspolku, "/"),
259     rm(tahystykset)
260
261 anestesioidu %>%
262   left_join(kohortti_pseudot, by = "henkilotunnus",
```

Poimintaskriptin alku

Kielimalli

Hankkeessa askelmerkit olivat:

- 1) Toisinnettavissa olevan ympäristön pystytys sopivalle palvelimelle, jossa on vLLM:llä GPU-kortteja käytettävissä – mahdollista siirtää ympäristöstä toiseen.
- 2) Kielimallin lataaminen kortin muistiin ja promptin testaus.
- 3) Kielimallin promptaaminen isompina batcheina (nopeus!)
- 4) Promptin hienosäätö ja eri kielimallien benchmarkkaus – parhaiden kandidaattien valinta muun muassa F1 scoreen perustuen [jolla mitataan yhtä aikaa mallin herkkyyttä (sensitivity) ja positiivista ennustearvoa (precision)]
- 5) Mallin valintaan vaikutti myös ns. konteksti-ikkunan koko (potilaskertomus-proosa on pitkää!) ja paljonko dataa kielimallille voi yhdessä promptissa syöttää.



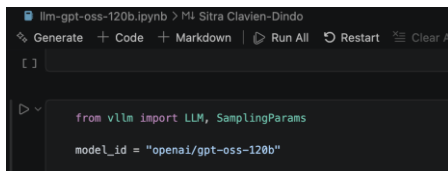
```
tommi — tommy@mlops-dev: ~/.cache/huggingface/hub

models--Qwen--Qwen2.5-32B-Instruct
models--Qwen--Qwen3-Next-80B-A3B-Instruct
models--Qwen--Qwen3.6-35B-A3B
models--google--medgemma-27b-text-it
models--meta-llama--Llama-4-Scout-17B-16E-Instruct
models--openai--gpt-oss-120b
models--openai--gpt-oss-20b
models--unsloth--MiniMax-M2.7
tommy@mlops-dev: ~/.cache/huggingface/hub$ _
```

Kielimalli, jatkoa

Dataa rajattiin ajallisesti kolmeen eri hetkeen:

- 30 pv leikkauksesta,
- 90 pv leikkaudesta ja
- siihen hetkeen asti mitä klinikolla oli dataa käytettävissään tehdessään alkuperäistä luokitusta.



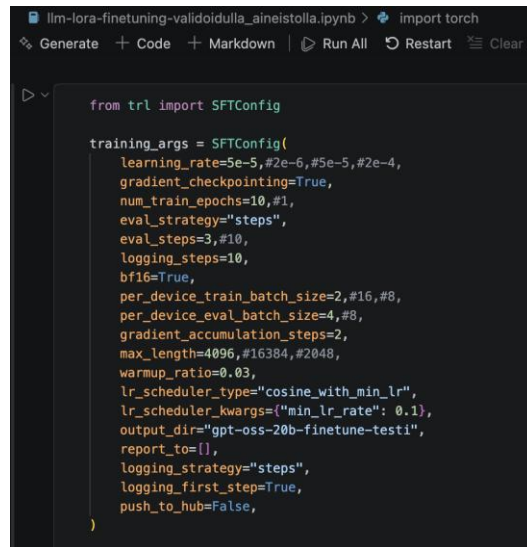
```
from vllm import LLM, SamplingParams

model_id = "openai/gpt-oss-120b"
```

Hankkeen alussa soveltuvimmaksi malliksi valikoitui OpenAI:n

gpt-oss-120b.

Kliinikon kanssa samaa mieltä **~87%** C-D-luokista.



```
from trl import SFTConfig

training_args = SFTConfig(
    learning_rate=5e-5, #2e-6, #5e-5, #2e-4,
    gradient_checkpointing=True,
    num_train_epochs=10, #1,
    eval_strategy="steps",
    eval_steps=3, #10,
    logging_steps=10,
    bf16=True,
    per_device_train_batch_size=2, #16, #8,
    per_device_eval_batch_size=4, #8,
    gradient_accumulation_steps=2,
    max_length=4096, #16384, #2048,
    warmup_ratio=0.03,
    lr_scheduler_type="cosine_with_min_lr",
    lr_scheduler_kwargs={"min_lr_rate": 0.1},
    output_dir="gpt-oss-20b-finetune-test1",
    report_to=[],
    logging_strategy="steps",
    logging_first_step=True,
    push_to_hub=False,
)
```

Mallia pyydettiin antamaan Clavier-Dindo-luokka ja lyhyt perustelu sille miksi luokka olisi juuri tämä.

Kliinikot ovat käyneet läpi kielimallin perusteluita ja kirjanneet ylös omat huomionsa missä kielimalli on mennyt metsään. Tätä perusteluiden validointia on laajennettu gastrokirurgialta neurokirurgiaan, kaulavaltimokirurgiaan ja gynekologiaan, sekä viimeisimpänä polven ortopediaan. **Validoituja tapauksia on nyt noin 500.**

Validoinnin pohjalta on tarkoitus jatkaa kielimallin hienosäädön ja kouluttamisen kanssa.

Rajoitteet

- Lopullisen kielimallin valinta tulee pitkälti riippumaan siitä miten kielimallia on tarkoitus promptata tuotantojärjestelmässä. Millainen on alustapalvelin missä varsinaista kielimallia ajetaan?
- Kielimallien kehitys menee erittäin vauhdikkaasti koko ajan eteenpäin - miten taataan myös tuotannossa kielimallin päivittäminen tai vaihtaminen sovelluksen taustalla?
- Tarvitseeko ratkaisu lääkinällisen laitteen leiman?
- Millainen tulee olemaan loppukäyttäjien vastaanotto?
- Miten ratkaisu skaalautuu kaikkiin erikoisaloihin?