

Samu Syrjänen | Opiskelija Helsingin yliopistossa

samu.syrjanen@gmail.com
+358 404161217
[LinkedIn](#)
Helsinki

Ala:
Datatiede
Koneoppiminen
Tietojenkäsittelytiede

Kielet:
suomi
englanti
ruotsi (heikko)

Esittely

Minulla on kokemusta ohjelmistotekniikasta, datatieteestä ja koneoppimisesta 229 opintopisteen edestä. Yleinen keskiarvoni on 3,76/5. Ohjelmistotekniikan, datatieteen ja ML-kurssien keskiarvoni on 4,18/5. Olen työskennellyt osana ryhmää yhteensä kuudessa eri ohjelmisto- tai dataan liittyvässä projektissa, ja olen oppinut ryhmän jäsenten välisen kommunikaation ja avun tärkeyden. Arvostan uusia oppimismahdollisuuksia ja jatkuvaa taitojen kehittämistä. Valmistumisen jälkeen näen itseni työskentelemässä tiimissä data scientistin tai data engineerin roolissa.

Koulutus

Syys 2023 - Jatkuu

Maisterintutkinto datatieteestä – Helsingin yliopisto

Odotettu valmistuminen: syksy 2025

Syys 2019 - Jouluku 2023

Kanditutkinto tietojenkäsittelytieteestä – Helsingin yliopisto

Taidot

- Python
- SQL
- Scrum ja ketterä ohjelmistokehitys
- Jatkuva integraatio
- Frontend & Backend
- End-to-end koneoppiminen (MLOps) ja Kubernetes
- Data-analyysi
- JavaScript/TypeScript
- R (heikko)

Kokemus

Loka 2023 - Jouluku 2023

Introduction to Machine Learning: Group Project – Yliopistokurssi

https://samusyrjanen.github.io/CV_Samu_Syrjanen/IML_project_report

Tässä projektissa tehtävämme oli ennustaa eri molekyylien kyllästymishöyrynpaine niiden ominaisuuksien perusteella. Tätä varten kokeilimme useita erilaisia koneoppimismalleja ja eri tekniikoita löytääksemme ja kouluttaaksemme parhaan mahdollisen mallin tehtävään. Projektiraportti sisältää esittelyn muun muassa seuraavista: exploratory data analysis, principal component analysis, käyttämämme ML-mallit, hyperparametrien optimointi, feature selection ja johtopäätökset.

Projektiin liittyen järjestettiin Kaggle-kilpailu ([linkki kilpailuun](#)), jossa kukin ryhmä julkaisi tuloksensa. Ryhmämme suoriutui keskimääräisesti. Ryhmämme R-squared arvo oli 0,6729, kun taas voittajaryhmän arvo oli 0,7275.

Keskeiset taidot:

- Exploratory data analysis (EDA)
- Feature selection
- Hyperparametrien optimointi
- Decision trees
- Random forest
- Gradient boosting

- Datan prosessointi ja käsittely
- Linear regression
- Python
- Dokumentaatio

Tammi 2023 - Touko 2023

Ohjelmistotuotannon ryhmäprojekti: Berry Picker Tracker – Yliopistokurssi

<https://marjanpoimijat.github.io/>

Kännykkäsovellus, joka seuraa käyttäjän kulkemaa reittiä ehkäisten metsiin eksymistä. Käyttäjät näkevät muiden seurattujen käyttäjien kulkemat reitit. Jos nettiyhteys katkeaa, viimeisin sijainti näkyy muille käyttäjille. Kattavampi esittely tiimin [github-sivulla](#).

Kurssin tavoitteena oli simuloida työelämää. Ryhmämme jatkoi sovelluksen kehittämistä siitä mihin edellinen ryhmä oli jäänyt. Koimme paljon ongelmia liittyen riippuvuuksien yhteensopivuuteen ja työkaluihin, kuten React Native, Expo, karttapalvelut ja välimuisti.

Suurimmat saavutuksemme:

- Sovelluksen käyttöönotto yliopiston palvelimilla
- Toimiva testiversio oikealla kännykällä (emulaattorin lisäksi)
- Koodin refactorointi: koodi on helposti luettavaa ja ylläpidettävää (laatu)
- Alustusscripti joka mahdollistaa nopean aloituksen uudelle kehittäjälle
- Dokumentaation päivitys
- Riippuvuuksien päivitys
- Käyttäjiin liittyvä data on kryptattua ja turvassa
- Kyky lisätä ja seurata useita käyttäjiä samanaikaisesti (+ asetukset ja UI siihen liittyen)
- Karttapalvelua pystyy vaihtamaan
- Kieltä pystyy vaihtamaan
- Käyttöliittymän uudistus
- Muita quality of life -ominaisuuksia

Keskeiset taidot:

- Full stack -kehitys
- Python (for backend)
- TypeScript (for frontend)
- Ryhmänjohtotaidot
- Asiakkaan tarpeiden ymmärtäminen ja toteutus
- Scrum ja ketterä ohjelmistokehitys
- Ohjelmistokehitys tiimissä
- Testaus
- Ohjelmistojen arkkitehtuuri
- Dokumentaatio

Touko 2023 - Kesä 2023

Tekstidatan ryhmittelyn yksilöprojekti: Sortext – Yliopistokurssi

<https://github.com/samusyrikanen/sortext>

Ohjelma ryhmittelee samantyyppisiä tekstejä ryhmiin. Se on tehty itse ilman ulkopuolisten kirjastojen tarjoamia valmiita algoritmeja. Ohjelma käyttää vanhoja BBC uutisartikkeleita ja käyttäjän syötteenä antamia tekstipätkiä ryhmien luontiin. Samankaltaisuuksien tunnistamiseen käytetään TF-IDF taulukoita ja K-means clustering -algoritmia.

Keskeiset taidot:

- Python
- Valvomaton koneoppiminen (K-means ja TF-IDF matriisit)
- Datan prosessointi ja käsittely
- Testaus
- Ohjelmiston arkkitehtuuri
- Dokumentaatio

[Näytä lisää kokemuksia \(vain englanniksi\)...](#)