

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Kristjan Luik 211809IAPM

Eesti metsaraie tuvastamine masinõppe meetoditega

Magistritöö

Juhendaja: Juhan-Peep Ernits

PhD

Tallinn 2025

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Kristjan Luik

03.04.2025

Annotatsioon

[ANNOTATSIOONI TEKST LÄHEB SIIA]

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 11 leheküljel, 5 peatükki, 3 joonist, 1 tabel.

Abstract
Thesis Title

[YOUR TEXT GOES HERE]

The thesis is in Estonian and contains 11 pages of text, 5 chapters, 3 figures, 1 table.

Lühendite ja mõistete sõnastik

API	Rakendusliides (<i>Application Programming Interface</i>)
CPU	Keskseade (<i>Central Processing Unit</i>)
IDE	Integreeritud programmeerimiskeskkond (<i>Integrated Development Environment</i>)
IOT	Asjade Internet (<i>Internet Of Things</i>)
VM	Virtuaalmasin (<i>Virtual Machine</i>)

Sisukord

1	Sissejuhatus.....	9
2	Valdkonna ülevaade	11
2.1	Metsandus.....	11
2.2	Copernicus ja EstHub	11
2.2.1	Sentinel.....	12
2.2.2	Lainepikkuste spekter	12
2.2.3	Koordinaatsüsteemid ja CRS	13
2.3	Masinõppe meetodite kasutus kaugseires	13
3	Lahendus	14
3.1	Töövahendid.....	14
3.2	Andmestiku loomine	15
3.2.1	Raie piirkonna andmete kogumine.....	15
3.3	Mudelite valik	17
3.4	Treenimis protsetuurid.....	17
4	Tulemuste analüüs	18
5	Kokkuvõte.....	19
	Kasutatud kirjandus	20
	Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.....	22
	Lisa 2 – Something	23
	Lisa 3 – Something Else	24

Jooniste loetelu

Joonis 1. Sateliidi andmete liikumine andmekeskuste vahel	12
Joonis 2. Näidis ühe lageraie päringust saadud ümbrus	16
Joonis 3. Andmestiku loomise töövoog	17

Tabelite loetelu

Tabel 1. Sentinel-2 MSI spektriribad ja nende kasutusvaldkonnad.....	13
--	----

1 Sissejuhatus

Antud magistritöö põhieesmärgiks on võrrelda masinõppe meetodeid ja tuua välja täpseim mudel, mis suudaks tuvastada metsaraiet satelliidipiltidelt. Mida aeg edasi seda rohkem on riigid hakanud mõistma kui tähtis on metsamajandus, metsade säilitamine ja hoidmine. Tehnoloogia pideva arenguga on hakatud ka otsima viise kuidas riik või kogukond saaksid paremat ülevaadet suurtest metsaga kaetud aladest. Metsa seireks kasutatakse peamiselt mehitamata õhusõidukeid (Unmanned Aerial Vehicles), maapealseid sensoreid, satelliidipildi töötlust, vabatahtlike kaasavaid rakendusi (Crowdsourcing Applications) [1].

Praegusel hetkel kasutatakse Eestis mõni aasta tagasi Keskkonnaagentuuri ja Tartu Ülikooli koostöös väljatöötatud statistika mudelit, mis raie tuvastamiseks kasutab suvasalu (Random forest) algoritmi [2] satelliidi piltidelt. Selle mudeli esmased tulemused olid paljulubavad, aga peale mõndaaegset kasutamist pole see ikkagi rahuldavaid tulemusi andnud ja mudeli kasutajad on sunnitud siiski manuaalseid viise kasutama.

Euroopa Liidu kaugseireprogrammt Copernicus võimaldab Eesti riigil koguda satelliidi pilte andmekeskusesse Esthub [3]. Lisaks muule infole, mida hallatakse Copernicus-es ja seeläbi ka Esthub-is, on kasutusel informatsioon mis tuleb erinevatelt Sentineli nime kandvatelt satelliitidelt [4]. Kuna Sentinel-2 on juba 2015. aastast töös olnud, sisaldab laia valikut valgusribasid ning on tiheda korduskülastus sagedusega [5], siis keskendub käesolev magistritöö peamiselt sellele satelliidi tüübile.

Sellest tulenevalt ona üheks alam eesmärgiks luua Python programm, mis hõlbustaks satelliidi piltide allalaadimist ja töötlemist. Peale andmete kogumist on plaan läbi viia tänapäevaste masinõppe mudelite võrdlus raiete tuvastamiseks. Raiet hinnatakse piksli põhise täpsusega üle pildi. Hiljuti on tehtud mitmeid uuringuid selles valdkonnas, kus kasutatakse ka suvasalu, XGBoost ja U-Net'il põhinevaid mudeli arhitektuure [6], [7]. Mõlemas uurimistöös on ka mudelite võrdlus välja toodud, aga need keskenduvad erinevatel suundadel. Esimese puhul ehitatakse mudelid kasutades rohkem pilte läbi aja, et mudel

saaks paremini tuvastada muutust. Teise puhul keskendutakse erinevate lainepikkuste kombineerimisele, et tabada muutusi.

Peale mudelite treenimist samadelt lähteandmetelt on antud magistritöös välja toodud tulemuste mõõtmine. Piksli tasemel täpsuse mõõtmiseks kasutatakse Intersection over Union - kattuvuse hinnang, Dice Coefficient - meetrika mis on põhimõtteliselt segmenteerimise F1 Score [8], [9]. Nende tulemuste abil saab teha võrdluse erinevate tuvastusmudelite vahel, et leida neist täpseim.

2 Valdkonna ülevaade

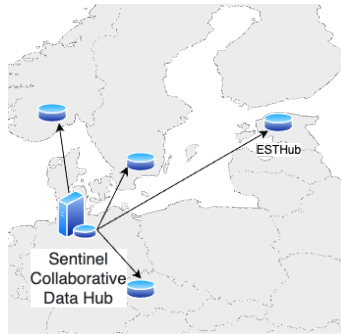
2.1 Metsandus

Metsad omavad olulist rolli nii ühiskonna igapäevaelus kui ka planeedi heaolus. Alates mööblis kasutatavast puidust kuni paberini, millele kirjutame. Lisaks neile nähtavatele toodetele sisaldavad paljud ravimid, kosmeetika ja pesuvahendid metsadest saadud kõrvalsaadusi. Rohkem kui 1,6 miljardit inimest sõltub metsadest toidu ja kütuse saamiseks ning umbes 70 miljonit, sealhulgas paljud põlisrahvad, peavad metsi oma koduks [10]. Metsad varustavad meid hapnikuga, pakuvad varjualust, töökohti, puhast vett ja toitu, olles seega inimkonna ellujäämiseks hädavajalikud. Kuna nii paljude inimeste elu sõltub metsadest, on metsade saatus otseselt seotud ka meie endi tulevikuga. [11]

2.2 Copernicus ja EstHub

Copernicus on üks osa Euroopa kosmoseprogrammist (EUS), mis tegeleb planeedi jälgimisega. Copernicus programmi raames, lisaks maa pealse info kogumisele, on loodud mitmeid satelliite, mis koguvad informatsiooni kosmosesest. See info on kõigile kättesaadav tasuta. Selle programmiga seotud satelliit kutsutakse **Sentineliks**. [12]

EstHub on Eesti riiklik satelliitandmete keskus, mis kogub ja integreerib mitmekesiseid georuumilisi andmeid automatiseeritud protsesside kaudu. Andmekogumine hõlmab kõrge resolutsiooniga satelliitkaardrite allalaadimist ja standardiseerimist erinevatest allikatest. EstHubi eesmärk on koguda kokku satelliidi andmed mis katavad Eesti territooriumi. [13]



Joonis 1. Satelliidi andmete liikumine andmekeskuste vahel

2.2.1 Sentinel

Sentinel-1 on radaripõhine satelliit, mis võimaldab jälgida maapinna vajumist, struktuuride kahjustusi ning geohazarde nagu maavärinad ja maalihked. Samuti on see ideaalne mere- ja Arktika seireks, sealhulgas laevade jälgimiseks ning naftareostuse tuvastamiseks. [14]

Sentinel-2 missioon koosneb kahest identsest satelliidist, Sentinel-2B (käivitatud 2017) ja Sentinel-2C (käivitatud 2024), mis töötavad koos, et pakkuda kõrge eraldusvõimega multispektraalseid pilte Maa pindadest, rannikualadest ja siseveekogudest iga viie päeva järel. Need andmed toetavad rakendusi põllumajanduses, metsanduses ja maakatte klassifitseerimisel. [15]

Sentinel-3 on Euroopa Maa seire satelliitmissioon, mille eesmärk on mõõta merepinna topograafiat, mere ja maa pinnatemperatuure ning ookeani ja maa pinnavärvi suure täpsusega. Neid andmeid kasutatakse ookeani prognoosisüsteemides, keskkonnaseires ja kliimaseires. [16]

Sentinel-5P on esimene Copernicuse missioon, mis on pühendatud atmosfääri seirele. See kannab tipptasemel **Tropomi** instrumenti, mis kaardistab mitmeid gaase nagu lämmastikdioksiid, osoon, formaldehüüd, vääveldioksiid, metaan, vingugaas ja aerosoolid - kõik need mõjutavad meie hingatavat õhku, tervist ja kliimat. [17]

2.2.2 Lainepikkuste spekter

Spektriribad on satelliitandmete analüüsimisel üliolulised, sest need võimaldavad eristada maapinna erinevaid omadusi, lähtudes elektromagnetilise spektri konkreetsetest lainepikkustest. Näiteks Sentinel-2 MSI instrumendi 13 spektririba hõlmavad nähtavat valgust,

lähedast infrapunat ja lühilaine infrapunat, võimaldades detailset maastiku klassifitseerimist, sealhulgas metsade, veekogude ja muu loodusliku keskkonna eristamist. Iga ribaga seondub kindel lainepikkuse vahemik, mida spetsiifiliste filtrite abil eraldatakse. [18]

Riba	Resolutsioon	Kasutus
B01	$60m\ px^{-1}$	Aerosool
B02	$10m\ px^{-1}$	Sinine
B03	$10m\ px^{-1}$	Roheline
B04	$10m\ px^{-1}$	Punane
B05	$20m\ px^{-1}$	Vegetatsiooni klassifitseerimine
B06	$20m\ px^{-1}$	Vegetatsiooni klassifitseerimine
B07	$20m\ px^{-1}$	Vegetatsiooni klassifitseerimine
B08	$10m\ px^{-1}$	Lähiinfrapunariba on hea rannajoonte ja biomassisisalduse kaardistamiseks
B8A	$20m\ px^{-1}$	Kitsam lähedane infrapunane
B09	$60m\ px^{-1}$	Veeaur tuvastus
B10	$60m\ px^{-1}$	Pilvede tuvastus
B11	$20m\ px^{-1}$	Lühilaine infrapunane 1
B12	$20m\ px^{-1}$	Lühilaine infrapunane 2

Tabel 1. Sentinel-2 MSI spektriribad ja nende kasutusvaldkonnad

2.2.3 Koordinaatsüsteemid ja CRS

Koordinaatsüsteem on meetod, mille abil määratletakse ja kirjeldatakse punktide asukohti maastikul, kasutades koordinaate. Selles kontekstis eristatakse kahte tüüpi: geograafilised koordinaatsüsteemid, mis kasutavad laiuse ja pikkuse väärtusi, ning projekteeritud koordinaatsüsteemid, mis teisendavad geograafilised koordinaadid lameda kaardi koordinaatideks, kasutades matemaatilisi projektsioone. CRS ehk koordinaatide viite süsteem määratleb reeglid ja parameetrid, mille alusel need koordinaadid seonduvad reaalse maastikuga. [19]

2.3 Masinõppe meetodite kasutus kaugseires

smt smt smt smt smt

3 Lahendus

3.1 Töövahendid

Python

Python on üldotstarbeline programmeerimiskeel, mida kasutatakse laialdaselt andmeteaduse, masinõppe ja ruumandmete analüüsi ülesanneteks tänu oma lihtsusele ja mitmekülgsusele.

Jupyter Notebookid

Jupyter Notebookid pakuvad interaktiivset keskkonda, kus saab koodi kirjutada, käivitada ja dokumenteerida ühes kohas. Need võimaldavad dünaamilist andmeanalüüsi ja tulemuste visuaalset esitlust, muutes uurimisprotsessi läbipaistvaks ja korduvaks.

Pandas

Pandas on andmetöötlusteek, mis pakub paindlikke ja efektiivseid andmestruktuure tabelipõhise andmetöötluse jaoks. See lihtsustab andmete puhastamist, analüüsi ja manipuleerimist.

GeoPandas

GeoPandas laiendab Pandase võimalusi, lisades tuge georuumilistele andmetele. See võimaldab lugeda, analüüsida ja visualiseerida ruumandmeid ning teostada geomeetrilisi operatsioone nagu lõikumine ja ühendamine.

Rasterio

Rasterio on Pythoni teek, mis keskendub rasterandmete lugemisele ja töötlemisele tuginedes GDAL-ile. See võimaldab ruumandmete analüüsi ning laseb rasterfailidega töötada efektiivselt ja intuitiivselt.

QGIS

QGIS on tasuta ja avatud lähtekoodiga töölaua GIS-tarkvara, mis võimaldab kasutajatel andmeid visuaalselt analüüsida, redigeerida ja kaardistada. See toetab mitmeid andmeformate ja pakub laialdasi geoprotsessimise võimalusi, olles populaarne nii akadeemilises kui ka professionaalses keskkonnas.

PostGIS

PostGIS on PostgreSQL andmebaasi laiendus, mis lisab geospaatialsete andmete töötlemise funktsionaalsuse. See võimaldab keerukaid ruumioperatsioone ja on oluline tööriist suurte geospaatialsete andmekogude haldamisel ning analüüsil.

3.2 Andmestiku loomine

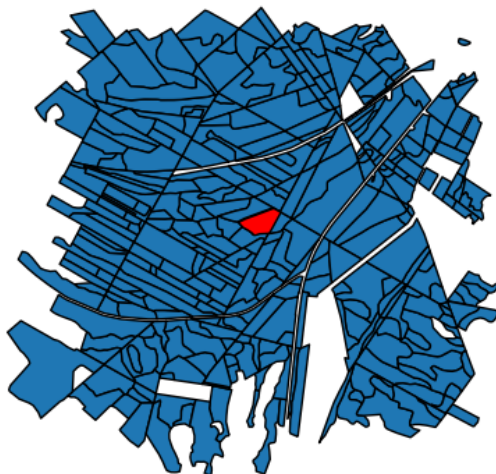
Selles peatükis käsitletakse andmestiku loomis protsessi, sealhulgas andmete kogumist, töötlemist ja maskimist. Andmestiku loomine on oluline samm igasuguste andmete analüüsi ja masinõppe projektide puhul. Andmete kvaliteet ja sobivus mõjutavad otseselt mudeli täpsust ja usaldusväärsust. Nagu muudes valdkondades kehtib ka informaatikas Pareto printsiip, mille kohaselt 80% probleemidest tuleneb 20% põhjustest. Seega on andmestiku loomine ja töötlemine äärmiselt oluline etapp, mis võib määrata kogu projekti edasise käigu.

3.2.1 Raie piirkonna andmete kogumine

Metsateatis on dokument, mille kaudu metsaomanik esitab Keskkonnaametile kavandatavate raietööde või oluliste metsakahjustuste kohta teabe. Keskkonnaamet kontrollib esitatud teatiste nõuetekohasust ning veendub, et kavandatav raie vastab kehtivatele õigusaktidele. Metsateatised menetletakse ja säilitatakse riiklikus metsaregistris. Peale edukat menetlemist või raie töö alustada 10 päeva peale otsust ja kuni 24 kuu jooksul. [20] Metsateatised on avalikud ja neid saab vaadata riiklikus metsaregistris.

Koostöös Keskkonnaametiga (envir) saadi andmed metsateatistest, mis sisaldavad teavet nii metsateatise esitamise kuupäeva, metsateatise menetlemise kuupäeva, metsateatise kehtivuse alguskuupäeva kui ka metsateatise kehtivuse lõppkuupäeva kohta. Kuna riigimetsade

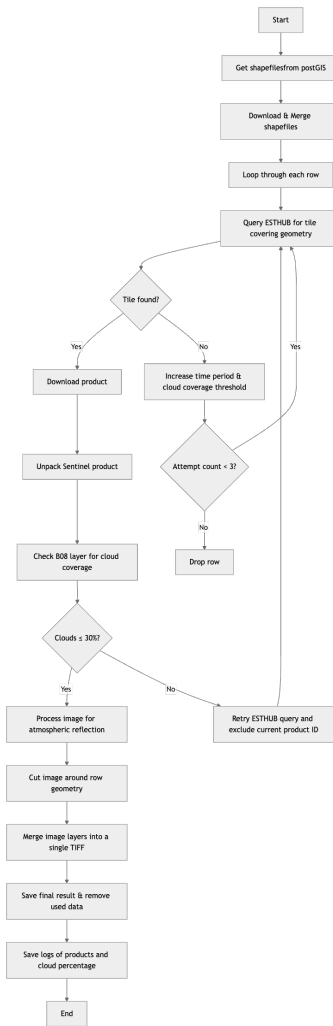
teatised on täpsemas seisukorras siis sai võetud need raieteatised aluseks. Seoses sellega et ühe lõigu peal võib olla väga väike kogus metsa, sai teatiste pärimine ümber ehitatud sedasi, et ühe metsa raie ümber kogutakse peale raie toimumist kokku ka kõik teised piirkonnad millel on teada kas on mets või raie. Piirkonniti pärimine sai teostatud kasutades PostGISi liidestust Posgresi andmebaasiga. Iga raie sisaldab ka endas geomeetria veergu, mis esitab polügooni kujul selle asukohta.



Joonis 2. Näidis ühe lageraie päringust saadud ümbrus

Polügoon on geomeetriline kujund, mis määratleb kindla ala, ühendades üksteisega punktid, et moodustada suletud piirjoon. Andmetöötluste ja ruumiandmete analüüsi kontekstis kasutatakse polügoone, et täpselt määratleda geograafilisi alasid. [21]

See on terve andmetöötlus workflow!!!



Joonis 3. Andmestiku loomise töövoog

3.3 Mudelite valik

Detail which machine learning algorithms you are using (e.g., Support Vector Machine, Random Forest, CNNs) and justify your choice based on the literature. Model Architecture: Provide a detailed description of the model architecture (e.g., input layers, hidden layers, output layer) including hyperparameters, kernel types (if SVM), or network design (if deep learning).

3.4 Treenimis protsetuurid

Describe the training process, including parameter tuning (grid search, cross-validation), and optimization methods.

4 Tulemuste analüüs

third

5 Kokkuvõte

summ3

Kasutatud kirjandus

- [1] Loretta Cheung, Jonathan Mason ja Meaghan Parker-Forney. „Perimeter Defense: 4 Technologies for Detecting and Preventing Illegal Logging“ (Fri, 11/06/2015 - 11:16). URL: <https://www.wri.org/insights/perimeter-defense-4-technologies-detecting-and-preventing-illegal-logging> (vaadatud 01.03.2025).
- [2] Tartu Ülikooli teadlased töötasid välja statistilised meetodid Eesti metsaressursi hindamiseks kaugseireandmete põhjal | Tartu Ülikool. R, 11.09.2020 - 11:05. URL: <https://ut.ee/et/sisu/tartu-ulikooli-teadlased-tootasid-valja-statistilised-meetodid-eesti-metsaressursi-hindamiseks> (vaadatud 11.01.2025).
- [3] Maa-amet. Riiklik satelliidiandmete keskus ESTHub. URL: <https://geoportaal.maaamet.ee/est/ruumiandmed/riiklik-satelliidiandmete-keskus-esthub-p443.html> (vaadatud 26.02.2025).
- [4] Infrastructure Overview | Copernicus. URL: <https://www.copernicus.eu/en/about-copernicus/infrastructure-overview> (vaadatud 01.03.2025).
- [5] Sentinel-2 - an Overview | ScienceDirect Topics. URL: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/sentinel-2> (vaadatud 01.03.2025).
- [6] K. Isaienkov *et al.* „Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem with Sentinel-2“. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), lk. 364–376. doi: 10.1109/JSTARS.2020.3034186.
- [7] N.S. Podoprigoorova *et al.* „Recognition of Forest Damage from Sentinel-2 Satellite Images Using U-Net, RandomForest and XGBoost“. Teoses: Proceedings of the 2024 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering, REEPE 2024. 2024. doi: 10.1109/REEPE60449.2024.10479810.
- [8] Intersection over Union (IoU): Definition, Calculation, Code. URL: <https://www.v7labs.com/blog/intersection-over-union-guide> (vaadatud 26.02.2025).
- [9] Understanding DICE COEFFICIENT. URL: <https://kaggle.com/code/yerramvarun/understanding-dice-coefficient> (vaadatud 26.02.2025).
- [10] A. Karsenty. „Underlying Causes of the Rapid Expansion of Illegal Exploitation of Tropical Timber“. *International Forestry Review* 5.3 (1. september 2003), lk. 236–239. issn: 1465-5489. doi: 10.1505/IFOR.5.3.236.19136. URL: <http://www.ingentaconnect.com/content/10.1505/IFOR.5.3.236.19136> (vaadatud 04.03.2025).
- [11] WWF - The Importance of Forests. URL: https://wwf.panda.org/discover/our_focus/forests_practice/importance_forests/ (vaadatud 04.03.2025).

- [12] *About Copernicus | Copernicus*. URL: <https://www.copernicus.eu/en/about-copernicus> (vaadatud 04.03.2025).
- [13] Maa-amet. *National Satellite Data Centre ESTHub*. URL: <https://geoportaal.maaamet.ee/eng/spatial-data/national-satellite-data-centre-esthub-p654.html> (vaadatud 10.03.2025).
- [14] *S1 Applications*. URL: <https://sentiwiki.copernicus.eu/web/s1-applications> (vaadatud 04.03.2025).
- [15] *S2 Applications*. URL: <https://sentiwiki.copernicus.eu/web/s2-applications> (vaadatud 04.03.2025).
- [16] *S3 Mission*. URL: <https://sentiwiki.copernicus.eu/web/s3-mission> (vaadatud 04.03.2025).
- [17] *S5P Applications*. URL: <https://sentiwiki.copernicus.eu/web/s5p-applications> (vaadatud 04.03.2025).
- [18] *S2 Mission*. URL: <https://sentiwiki.copernicus.eu/web/s2-mission> (vaadatud 25.03.2025).
- [19] *8. Coordinate Reference Systems — QGIS Documentation Documentation*. URL: https://docs.qgis.org/3.40/en/docs/gentle_gis_introduction/coordinate_reference_systems.html (vaadatud 26.03.2025).
- [20] *Metsateatis ja metsaregister | Keskkonnaamet*. URL: <https://keskkonnaamet.ee/elusloodus-looduskaitse/metsandus/metsateatis-ja-metsaregister> (vaadatud 02.04.2025).
- [21] *What Is a Location Polygon?* URL: <https://www.narrative.io/knowledge-base/concepts/data-terms/what-is-a-location-polygon> (vaadatud 02.04.2025).

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Kristjan Luik

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Eesti metsaraie tuvastamine masinõppe meetoditega”, mille juhendaja on Juhan-Peep Ernits
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

03.04.2025

¹Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtjaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2 – Something

```
<!DOCTYPE html>  
<html>  
<body>  
  
<h1>Example Title </h1>  
  
<p>Some text here </p>  
  
</body>  
</html>
```

Lisa 3 – Something Else

Pythagorean theorem

$$x^n + y^n = z^n \tag{1}$$

Normal distribution

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{2}$$