

TARTU ÜLIKOOL  
LOODUS- JA TÄPPISTEADUSTE VALDKOND  
MATEMAATIKA JA STATISTIKA INSTITUUT  
MATEMAATILISE STATISTIKA ERIALA

HENDRIK PIIKOV  
**COVID-19 MÕJU KROONILISTE HAIGUSTE  
PÕDEJATELE EESTIS**  
BAKALAUREUSETÖÖ (9 EAP)

Juhendaja: PhD Raivo Kolde

TARTU 2023

# COVID-19 MÕJU KROONILISTE HAIGUSTE PÕDEJATELE

## EESTIS

Bakalaureusetöö

Hendrik Piikov

### Lühikokkuvõte

COVID-19 pandeemia on avaldanud mõju kogu maailma tervishoiusüsteemidele ja elanikkonnale, eriti krooniliste haigustega inimestele. Need patsiendid kuuluvad suurema riskiga rühma nii COVID-19 raske kulgemise kui ka nende põhihaiguste halvenemise osas. Käesoleva uurimistöö eesmärgiks on keskenduda kroonilisi haiguseid põdevatele inimestele ning hinnata nende ligipääsu tervishoiuteenustele COVID-19 pandeemia esimese aasta jooksul. Selleks kasutati elektroonilisi terviseandmeid, et ennustusmudelit treenida. Ennustusmudel kasutab gradientvõimendus algoritmi ehk järjestikus õppe algoritmi. Uurimistöö tulemusena selgus, et mudel hindab üsna hästi inimeste ligipääsu tervishoiuteenustele. Samas selgus ka, et mudelit on vaja kalibreerida iga teatud perioodi tagant ning et riskimudeleid saab kasutada järgmistel nakkushaiguste levikutel.

**CERCS teaduseriala:** B110 Bioinformaatika, meditsiiniinformaatika, biomaatemaatika, biomeetrika

**Märksõnad:** Kroonilised haigused, COVID-19, SARS-CoV-2

# THE IMPACT OF COVID-19 ON CHRONIC DISEASE PATIENTS IN ESTONIA

Bachelor thesis

Hendrik Piikov

### Abstract

The COVID-19 pandemic has impacted healthcare systems and populations

worldwide, particularly those with chronic diseases. These patients belong to a higher-risk group for both severe progression of COVID-19 and deterioration of their underlying conditions. The aim of this study is to focus on individuals with chronic diseases and assess their access to healthcare services during the first year of the COVID-19 pandemic. For this purpose, electronic health data was used to train a predictive model. The prediction model utilizes the gradient boosting algorithm, also known as a sequential learning algorithm. The results of the study revealed that the model assesses access to healthcare services quite well. However, it also emerged that the model needs recalibration at certain intervals, and that risk models can be utilized in subsequent infectious disease outbreaks.

**CERCS research specialisation:** B110 Bioinformatics, medical informatics, biomathematics, biometrics

**Key words:** Chronic diseases, COVID-19, SARS-CoV-2

# Sisukord

<b>Sissejuhatus</b>	<b>4</b>
<b>1 COVID-19 ja koroonaviirus</b>	<b>6</b>
<b>2 Andmed ja andmete töötlemine</b>	<b>8</b>
2.1 Andmed . . . . .	8
2.2 Kohort ja nende loomine . . . . .	9
<b>3 Ennustusprobleem</b>	<b>10</b>
3.1 Idee . . . . .	10
3.2 Riskimudel . . . . .	10
3.3 Ennustusmudeli kovariaadid . . . . .	12
3.4 Ennustusmudeli ROC ja AUC . . . . .	12
<b>4 Analüüs</b>	<b>14</b>
4.1 Tulemused periooditi . . . . .	15
<b>Kokkuvõte</b>	<b>23</b>
<b>Kasutatud kirjandus</b>	<b>26</b>

## Sissejuhatus

COVID-19, mida põhjustab uus koroonaviirus SARS-CoV-2, kujutas endast 2019. aasta lõpust alates globaalset tervisekriisi, mis avaldas olulist mõju maailma sotsiaalsetele, majanduslikele ja tervishoiu struktuuridele. Kogu maailm tunnistas enneolematut pingutust haiguse leviku tõkestamiseks ja ravivõimaluste laiendamiseks. Kuigi suur osa avalikkuse ja teadlaste tähelepanust keskendus COVID-19 otsesele mõjule, on selle pandeemia ulatuslikud tagajärjed tervishoiusüsteemile laiemad ja sügavamad.

COVID-19 pandeemia ajal olid paljud riigid, sealhulgas Eesti, silmitsi tervishoiusüsteemi piirangute ja väljakutsetega. Eesti tervishoiuametnikud ja meditsiinitöötajad tegid jõupingutusi, et tagada patsientidele vajalik ravi ja toetada ühiskonda kriisi ajal. Kuid nagu teisteski riikides, tõi pandeemia esile süsteemi kitsaskohad ja vajaduse leida lahendusi olemasolevatele ja tulevastele tervishoiuväljakutsetele. Paljud haiglad kohandasid oma tegevust, et keskenduda COVID-19 patsientidele. See tähendas mõnikord plaaniliste operatsioonide ja protseduuride edasilükkamist, et vabastada ressursse ja personali. Paljud tervishoiutöötajad töötasid pikki tunde ja olid suure stressi all, kuna nad pidid tegelema COVID-19 patsientide suureneva arvuga ja nende endi nakatumise ohuga. Mõningaid tervishoiuasutusi suleti või piirati nende tegevust, eriti pandeemia haripunktis. See mõjutas patsientide võimet saada arstiabi teiste haiguste või seisundite jaoks.

Üks suur grupp inimesi, keda koroonapandeemiast tingitud arstiabi kättsaadavuse puudumine mõjutab on kroonilised haiged. Need patsiendid kuuluvad suurema riskiga rühma nii COVID-19 raske kulgemise kui ka nende põhihaiguste halvenemise osas. Krooniline haigus on haigusseisund või haigus, mis on pikaajaline või püsiv. See tähendab, et haigus kestab tavaliselt vähemalt kolm kuud või kauem. Krooniliste haiguste hulka kuuluvad südame-veresoonkonna haigused, diabeet, kroonilised hingamisteede haigused, vähk jne. Kroonilised haigused võivad oluliselt mõjutada inimese elukvaliteeti, piirates tema võimet teha igapäevaseid tegevusi ja nautida

elu. Need haigused võivad nõuda regulaarset meditsiinilist jälgimist ja ravi. Kroonilised haigused on ka suur koormus tervishoiusüsteemidele ja võivad kaasa tuua märkimisväärseid otseseid ja kaudseid kulusid ühiskonnale, sealhulgas meditsiinilised kulud ja tootlikkuse kaotus. Tänapäeval on kroonilised haigused maailmas peamiseks enneaegseks surmapõhjuseks [1], seega peaks kroonilised haiged olema tervishoiul üks kõrgema prioriteediga patsiendirühmasid.

Uurimistöö eesmärk on keskenduda kroonilisi haiguseid põdevatele inimestele ning hinnata nende ligipääsu tervishoiuteenustele COVID-19 pandeemia ajal. Uurimistöös keskendutakse pandeemia esimesele aastale ning võrreldakse arstivisiite enne pandeemiat ja selle ajal. Andmestik on küsitud Haigekassalt RITA projekti CORIVA jaoks.

Uurimistöö koosneb neljast peatükist. Esimeses peatükis keskendutakse COVID-19 ja selle levikule. Teises peatükis käsitletakse andmeid ja andmete töötlemist. Kolmandas keskndutakse ennustusprobleemile ja riskimudelile. Viimases peatükis on välja toodud tulemused ja nende interpreteerimine.

# 1 COVID-19 ja koroonaviirus

COVID-19 on haigus, mille põhjustajaks on koroonaviiruse tüüp nimega SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). See viirus kuulub laiemalt koroonaviiruste perekonda, mida on teada olevat mitmeid tüüpe, sealhulgas SARS-CoV, MERS-CoV ning mitmed teised, mis põhjustavad inimestel vähemtõsiseid hingamisteede infektsioone [2].

Koroonaviirused on suur viiruste perekond, mida iseloomustab nende ümbritsev "koroonalaadne" ehk kroonitaoline väljumus elektronmikroskoobi all. Nimi "koroonaviirus" pärineb ladina sõnast 'corona', mis tähendab 'kroon' või 'halo'. Need viirused on zoonootilised, mis tähendab, et need võivad levida loomadelt inimestele. SARS-CoV-2 on mõeldud eriti inimestele, kuigi algne reservuaar on tõenäoliselt nahkhiired [3].

COVID-19 haigus levib peamiselt inimeselt inimesele läbi hingamisteede piiskade, mis eralduvad nakatunud inimese köhimisel, aevastamisel või rääkimisel. Sümptomid ilmnevad tavaliselt 2-14 päeva pärast nakatumist ja võivad hõlmata palavikku, köha, hingamisraskusi, väsimust ja maitse- või lõhnataju kaotust. Kuigi paljudel inimestel on haigus kerge, võib see põhjustada raskemaid tüsistusi, nagu kopsupõletik, äge hingamisteede distressi sündroom (ARDS) ja mõnel juhul surm [4].

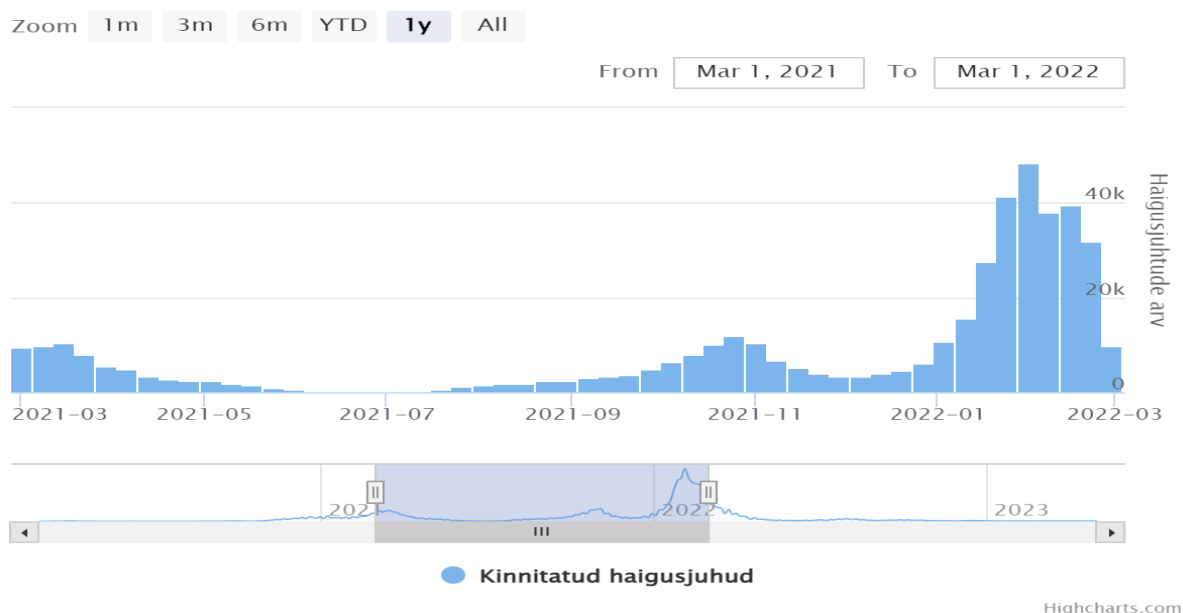
COVID-19 erineb teistest koroonaviirustest mitmel olulisel viisil. Kõigepealt, SARS-CoV-2 viirus on võrreldes SARS-CoV ja MERS-CoV viirustega inimestel palju nakavam, mis on osaliselt seletatav selle võimega levida ka asümptomaatilistelt inimestelt. Teiseks, SARS-CoV-2 põhjustab oluliselt madalamat suremust kui SARS-CoV või MERS-CoV, kuigi suurema leviku tõttu on COVID-19 põhjustanud rohkem surmajuhtumeid kui need haigused kokku [5].

SARS-CoV-2 genoomi analüüs on näidanud, et see viirus on lähedane sugulane SARS-CoV-le, millest sai nime. Viiruse S (spike) valk, mis on oluline inimeste rakkude nakatamiseks, on SARS-CoV-2 puhul muutunud, mis võib osaliselt seletada

selle suuremat nakkavust võrreldes SARS-CoV-ga [6].

WHO [4] andmetel on teadaolevalt COVID-19 viirusesse nakatanud üle 760 miljoni inimese, millest 6,9 miljonil korral on nakatumine lõppenud surmaga. Viirus levib üle terve maailma. Nakatumiste juhtumite poolest on suurimad piirkonnad Euroopa ja Ameerika, vastavalt 275 miljonit ja 193 miljonit nakatumise juhtumit. Eestis on teadaolevalt üle 600 tuhande nakatumise juhtu, millest 2,9 tuhandel korral on nakatumine lõppenud surmaga.

### Uued haigusjuhud



*Joonis 1.* Uued haigusjuhud vaadeldava perioodi (01.03.2021-01.03.2022) kohta Kooronakaart [7] andmetel.

Nagu jooniselt 1 näha, siis COVID-19 koroonaviirus levib lainetena. Esimene laine sai alguse 2021. aasta alguses, teine laine algas oktoobris 2021 ja viimane ning ühtlasi ka suurim viiruse laine sai alguse 2022. aasta alguses.

COVID-19 pandeemia on esile tõstnud koroonaviiruste põhjustatud haiguste mõju ja ohtu inimestele. On oluline, et jätkaksime nende viiruste, nende leviku ja nende vastu võitlemiseks mõeldud meetmete uurimist.

## 2 Andmed ja andmete töötlemine

Käesolevas peatükis toome välja kõige olulisemad uurimistöö läbiviimisega seotud metodoloogilised detailid.

### 2.1 Andmed

Käesolevas uurimistöös on andmestikuks Haigkassa projekti CORIVA (COVID-19 haigusjuhtumite analüüs ja riskirühmade väljaselgitamine Eestis) raames kogutud andmed. Andmestiku ajavahemikuks on 2020. aasta veebruar kuni 2021. aasta veebruar ja selles on esindatud kõik positiivse SARS-CoV-2 testi või COVID-19 diagnoosi saanud isikute terviseandmed. Lisaks on iga isiku kohta ka 4 kontrollisikut, kes andmestikku lisamise hetkel ei olnud COVID-19-sse nakatunud. Kontrollisikute lisamine käis juhuslikult, s.t kontrollisikuid ei sobitatud andmestikku erinevate tunnuste poolest. Kokku on andmestikus 386 557 patsiendi terviseandmed, kuid planeeritud 1:4 suhe ei jäänud püsima, sest hilisemal perioodil leidis selliseid kontrollisikuid, kes said positiivse COVID-19 diagnoosi. Väljastatud terviseandmed sisaldasid

- 1) patsientide raviandmeid, mis sisaldavad patsiendi sugu, sünniaastat, diagnoose, arstivisiitide kuupäevaid ja -tüüpe ning patsiendile osutatud protseduure.
- 2) patsiendile välja kirjutatud retsepte perioodil veebruar 2017 kuni november 2021, mis sisaldab patsiendi diagnoosi, väljakirjutatud ravimit, kuni 3 ravimi toimeainet, ettenähtud ravimi kogust ning ravimi tarvitamise juhendit.

OHDSI Andmed asuvad PostgreSQL andmebaasis. PostgreSQL aitab OHDSI projektis terviseandmete haldamist, analüüsimist ja modelleerimist ning pakub ka avatud lähtekoodiga ja laiendatavat andmebaasi keskkonda. Bakalaureusetöök vajalikud andmed olid viidud The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) kujule. Observational Health Data Sciences and Informatics (OHDSI) kohaselt on OMOP CDM spetsiaalselt välja töötatud sel-

leks, et erinevatest allikatst terviseandmed ühtlustada ning muuta need omavahel võrreldavaks [8]. OMOP CDM abil on võimalik teha mitmesugusid analüüse (nt epidemioloogilisi, farmakoepidemioloogilisi jne) kasutades suuri andmebaase. OMOP CDM toetab rahvusvahelist koostööd, võimaldades eri riikide ja süsteemide andmebaase ühendada ja analüüsida. Kuna OMOP CDM standardiseerib andmeid, on võimalik tagada parem andmete anonüümsus ja turvalisus, vähendades individuaalsete isikute tuvastamise riski. OHDSI toob välja, et tänu standardiseeritud andmete kujule on võimalik andmetele ligipääsu omamata defineerida analüüsi sammud, mida saab seejärel jookсутada serveris, kus andmeid hoitakse. Käesoleva töö autor on saanud õiguse kasutada andmeid bakalaureusetöö kirjutamise raames.

## 2.2 Kohort ja nende loomine

Kohort on grupp inimesi, kellel on ühine omadus, tavaliselt vanus [8]. Kohortide loomine toimus läbi veebirakenduse OHDSI ATLAS.

OHDSI ATLAS on rahvusvaheline teaduskogukond, mis on pühendunud terviseandmete analüüsimisele. ATLAS on üks OHDSI poolt välja töötatud platvormidest, mille eesmärk on toetada terviseandmete analüüsi OMOP CDM põhjal.

Kohortide loomiseks läbi veebipõhise kasutajaliidese ATLAS tuleb esmalt mõistete hulk defineerida. Seejärel andes ette erinevaid kriteeriumeid, saab päringut aina spetsiifilisemaks muuta (näiteks anda ette teatud ajavahemik, millal inimene arstivisiidil käis, kas inimene haigestus või mitte, kas patsiendile määrati mingi retsept jne). Pärast seda on ka võimalik kohort genereerida, viies see OMOP CDM kujule, mille saab hiljem SQL päringuna välja võtta. Kohort koosneb neljast näitajast: kohort ID, patsiendi ID, kohorti sisenemise aeg ja sealt väljumise aeg.

## 3 Ennustusprobleem

### 3.1 Idee

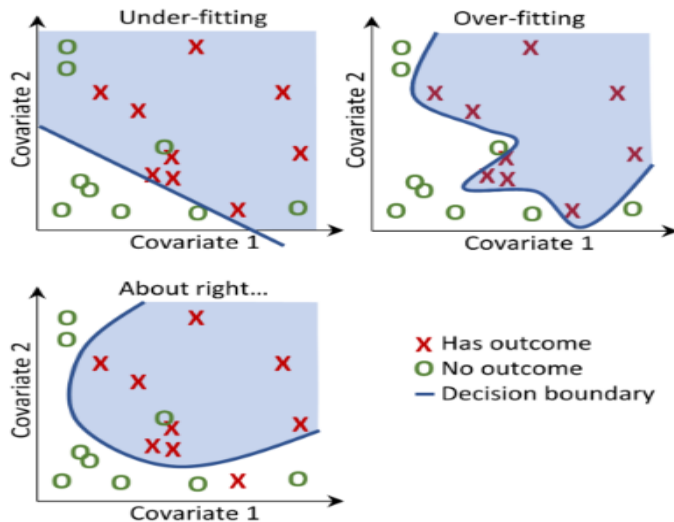
OHDSI Patient-Level Prediction (PLP) ennustusprobleem kasutab tervishoiu andmestikku, et selle abil ennustada tulemust individuaalsele patsiendile. Selleks arendatakse mudel, mis ennustab ajalooliste terviseandmete pealt kas indiviid võib haigestuda või mitte. See aitab tuvastada inimesi, kellel võib olla teatud haigusseisundi oht, võimaldades ennetatavat meditsiinilist sekkumist ja personaalset tervishoiuteenust.

### 3.2 Riskimudel

Riskimudel annab hinnangu riskile, arvutades välja protsendi tulemi tekkimisele. Kui protsent ületab ette määratud riskilävendit, siis mudel hindab, et tulem tekib. Kui protsent jääb alla ette määratud riskilävendit, siis mudel hindab tulemi mitte-tekkimist [9]. Riskimudelid on laialdaselt kasutuses, näiteks finantssektoris võib riskimudel prognoosida laenuvõtja võimet laenu tagasi maksta, kindlustussektoris võivad riskimudelid hinnata kindlustuslepinguga seotud kahjude tõenäosust, tervishoius võivad riskimudelid hinnata patsiendi tõenäosust haigestuda teatud haigusesse jne. Riskimudelite lävend tuleb hoolikalt valida, sest valepositiivsed ja valenegatiivsed hinnangud võivad teha suurt kahju, näiteks valepositiivse hinnangu tulemusena määratud ravikuur võib tekitada rohkem kahju kui kasu.

Riskimudeli üheks väga tähtsaks osaks on otsustuspiir. See jaotab patsiendid juhtumiteks ja mitte-juhtumiteks. Otsustuspiir on hüperpind, mis eraldab erinevaid klasse omaduste ruumis. Ülesobitumine toimub siis, kui mudel toimib suurepäraselt treeningandmetel, kuid ei pruugi hästi generaliseeruda uutele, nägemata andmetele. Alasobitumine toimub siis, kui mudel on liiga lihtne ja seetõttu paigutatakse palju patsiente valesti. Mudel on ideaalne, kui leitakse ülasobitumise ja alasobitumise

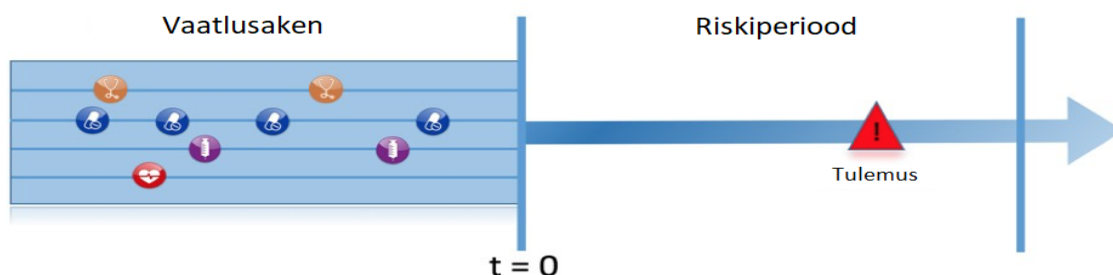
vaheline otsustuspiir.



Joonis 2. Joonistel [8] on kujutatud alasobitamist (under-fitting), ülesobitamist (over-fitting) ja korrektset sobitamist otsustuspiiri abil.

Käesolevas bakalaureusetöös kasutati gradientvõimendust. Gradient Boosting Machine (GBM) ehk gradientvõimendi on populaarne järjestikuse õppe algoritmi põhine masinõppe mudel, mida kasutatakse nii regressiooni- kui ka klassifikatsiooniprobleemides. Gradientvõimendi mudeli põhiidee on kombineerida mitu nõrka ennustajat (tavaliselt otsustuspuud) tugevaks ennustajaks. Gradientvõimendi eripära seisneb selles, et see optimeerib järk-järgult ehitatavaid nõrku õppijaid, tavaliselt otsustuspuud, vähendades järk-järgult viga, mis on määratud gradientidega. [10]

Riskimudeli koostamiseks on vaja defineerida populatsioon, kellel mudelit plaanitakse rakendada, valim ehk milliste inimeste andmetel mudel treenitakse, tulem ehk mis sündmust ennustatakse ning mis ajaperioodil inimesi jälgitakse.



Joonis 3. Joonisel [8] on ennustusprobleem.

Vaatlusakna pikkuseks on 365 päeva ehk 1 aasta. Vaatlusaken koondab endasse kõiki neid patsiente, kes käisid arstivisiidil ning kes ei olnud ise COVID-19-sse nakatunud. Mudeli treenimisel/õpetamisel kasutati nende patsientide raviandmeid ja välja kirjutatud retsepte. Ajahetk  $t=0$  algab märtsist 2020 (umbes siis kuulutati välja ka koroonapandeemia) ja riskiperioodi pikkuseks on 3 kuud. Mudel üritab ennustada kas patsiendil tekib tulem või mitte.

### 3.3 Ennustusmudeli kovariaadid

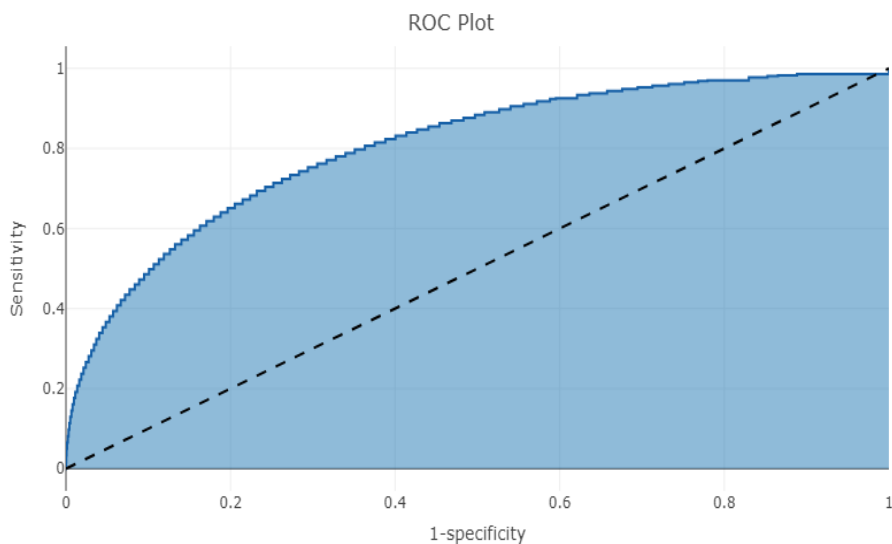
Antud uurimistöona valmis mudel, mis ennustab inimese arstivisiiti. Mudel kasutab ennustamisel mitut erinevat kovariaati. Kovariaatidest on kõige suuremateks mõjutajateks arstivisiit viimase aasta jooksul (0.703), apteegis käimine viimase aasta jooksul (0.523) ja need patsiendid, kellele on välja kirjutatud südame ja veresoonkonna rohud (0.406). See tähendab, et neil on suurem tõenäosus minna arsti juurde. Kõige väiksemateks mõjutajateks on meessugu (-0.219), või kui on võõrkeha silmas viimase aasta jooksul olnud (-0.017). Näidis

### 3.4 Ennustusmudeli ROC ja AUC

ROC kõver (Receiver Operating Characteristic curve) ja AUC (Area Under the Curve) on statistilised mõõdikud, mida kasutatakse sageli masinõppe ja meditsiiniliste uuringute valdkondades, et hinnata klassifitseerimismudelite tulemuslikkust.

Kuna tegemist on binaarse tulemiga, kasutame ROC kõverat. ROC kõver on graafiline kujutis, mis kujutab tõelise positiivse määra ('True Positive Rate ehk TPR) ja valepositiivse määra ('False Positive Rate ehk FPR) suhet erinevatel läveväärtustel. TPR määrab, kui palju tegelikest positiivsetest juhtudest ennustatakse õigesti, samal ajal kui FPR määrab, kui palju tegelikest negatiivsetest juhtudest ennustatakse valesti. [11]

AUC on ROC kõvera alune pindala. AUC väärtus jääb vahemikku 0 kuni 1. Suurem AUC tähendab, et mudelil on parem võime eristada positiivseid ja negatiivseid klasse. Ideaalse mudeli korral on  $AUC = 1$ , samas kui juhusliku klassifikaatori puhul on  $AUC = 0,5$ . [12]



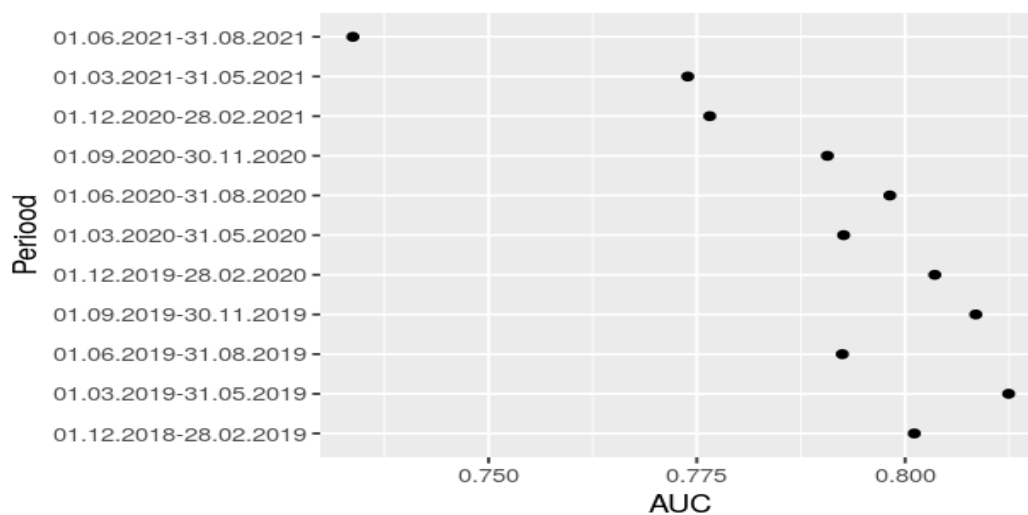
*Joonis 4.* Mudeli ROC kõver.

Tegemist on hea mudeliga, sest ROC kõver ei ole diagonaalse sirge lähedal. See Ennustusmudeli AUC väärtuseks sai 0.808, mis näitab, et mudel eristab tulemit mitte-tulemist üsna hästi.

## 4 Analüüs

Kohortide loomiseks, töötlemiseks ja analüüsimiseks kasutati arenduskeskkonda RStudio. Rstudio ühendamiseks PostgreSQL andmebaasiga kasutati DatabaseConnector paketti. DatabaseConnectori abil on võimalik teha andmebaasis SQL päringuid, mida saab analüüsimisel kasutada. Kohortid loodi paketi PatientLevelPrediction funktsioonide abil.

Analüüsiks koostati 11 erinevat kohorti. Iga kohort on koostatud erineva perioodi kohta. Esimene ajaperiood on 01.12.2018-28.02.2019, teine ajaperiood on 01.03.2019-31.05.2019 jne. Iga ajaperioodi pikkuseks on 3 kuud. Viimane ajaperiood on 01.06.2021-31.08.2021.



Joonis 5. AUC iga perioodi kohta.

Joonisel 5. on välja toodud iga perioodi kohta arvutatud AUC väärtus. Kuna ennustusmudel on treenitud perioodi 01.03.2019 - 01.03.2020 pealt, siis eeldatavasti on sellel perioodil kõige kõrgem AUC.

Nagu jooniselt näha, on kõige täpsemaks perioodiks 01.03.2019 - 31.05.2019, mille korral on AUC väärtuseks 0.8124. Antud periood kuulub ühtlasi ka ennustusmudeli õppimise ajaperioodi vahemikku. Samas ka teised perioodid, mis kuuluvad mudeli õppimise ajaperioodi, on saavutanud kõrge AUC väärtuse.

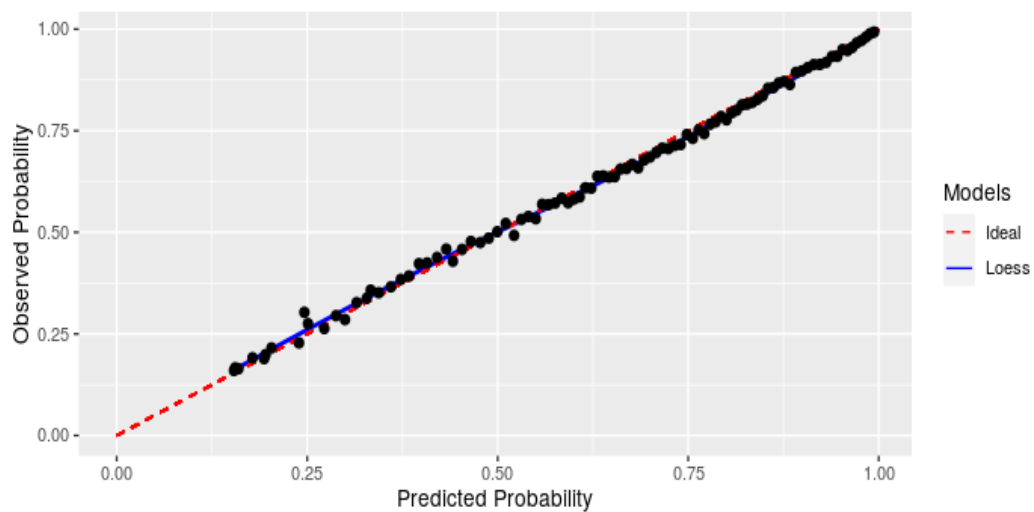
Kõige halvemini ennustab mudel 1.06.2021 - 31.08.2021 ehk viimase perioodi kohta, mille korral on ROC kõvera aluse pindala ehk AUC väärtuseks 0.7337. On näha ka, et mida aeg edasi, seda halvemaks lähevad ennustused (AUC väärtus järk-järgult langeb). Seega kindlasti on vaja mudelit uuesti treenida/kalibreerida, et vaadata hilisemaid perioode.

## 4.1 Tulemused periooditi

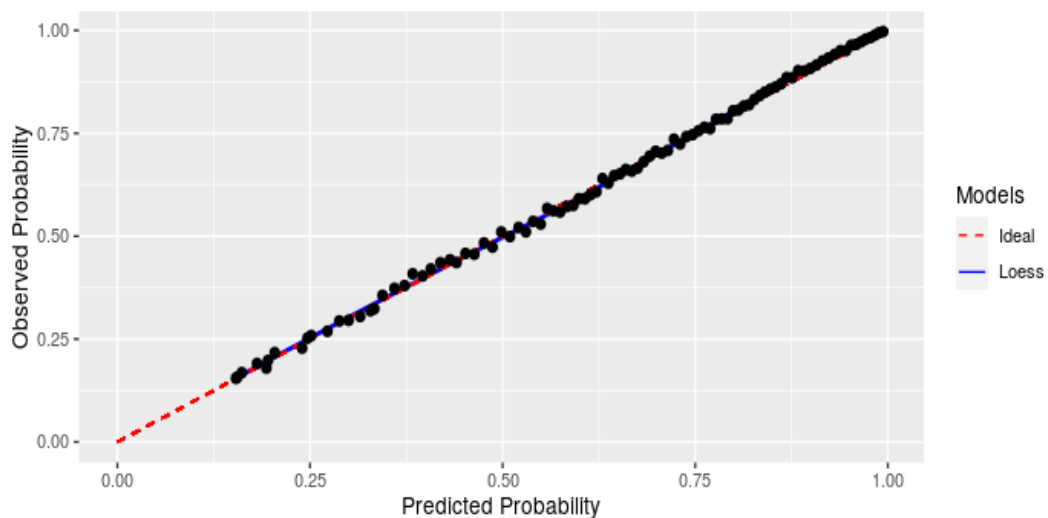
Järgmisena uuriti suhet ennustuse ja tegelikkuse vahel. Selleks vaadati kalibreerimiskõverat.

Mudeli sobivust reaalandmestikule on võimalik kirjeldada kalibratsiooniga. Kalibratsioon näitab ennustatud riskide sobivust vaadeldavate riskidega [8]. Kalibreerimiskõver on graafik, mille x-teljel on kantud andmepunktide ennustatud risk ning y-teljele on kantud samade andmepunktide vaadeldud keskmine esinemistõenäosus [8].

Ideaalselt kalibreeritud mudeli kalibratsioonikõvera punktid asetsevad  $x = y$  joonel. Kui punktid asuvad  $x=y$  joonest kõrgemal, siis ennustatav hinnang osutus madalamaks tegelikest vaadeldavatest andmepunktidest. Kui punktid asuvad  $x=y$  joonest madalamal, siis ennustatav hinnang osutus kõrgemaks tegelikest vaadeldavatest andmepunktidest.

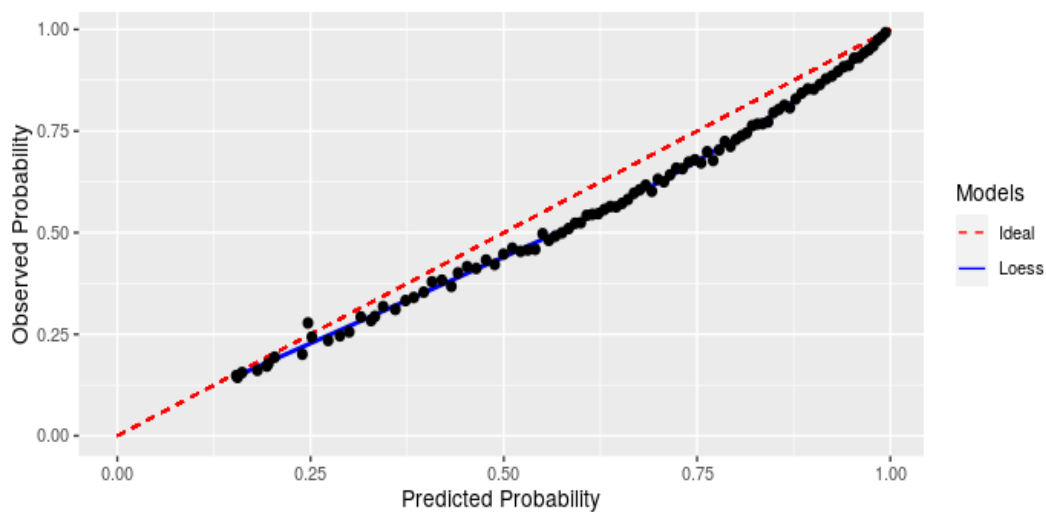


*Joonis 6.* Kalibreerimiskõver perioodil 01.12.2018-28.02.2019



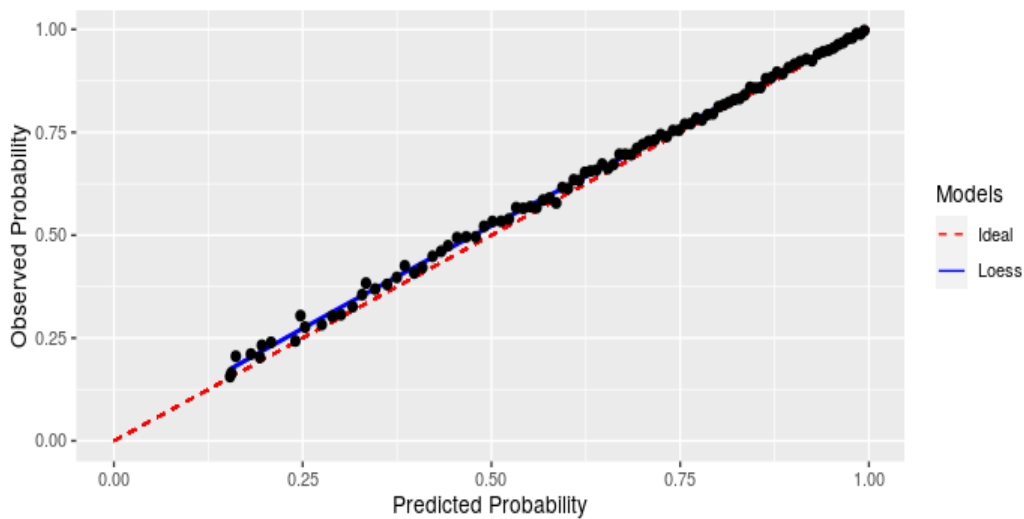
*Joonis 7.* Kalibreerimiskõver perioodil 01.03.2019-31.05.2019

Joonistelt 6 ja 7 on näha, et ennustatud punktide hinnang on sarnane vaadeldavate andmetega. Seda seetõttu, et tegemist on koroonaeelse perioodiga ning seetõttu ei olnud ka arstiabi kättesaadavus mõjutatud. Mudel on peaaegu ideaalselt kalibreeritud, sest erinevus on minimaalne.



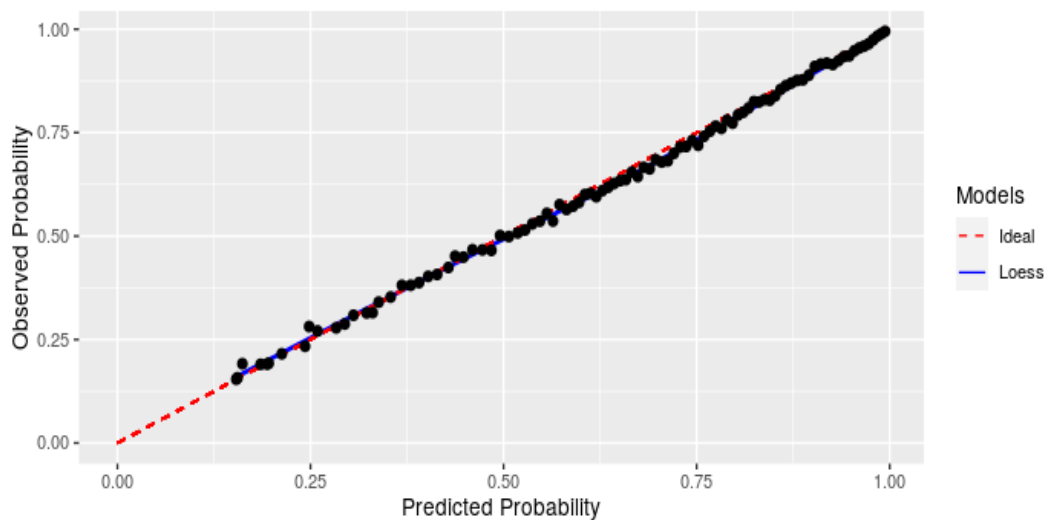
Joonis 8. Kalibreerimiskõver perioodil 01.06.2019-31.08.2019

Joonisel 8 on näha, et ennustatud punktide hinnang on kõrgem kui tegelik vaadeldavate punktide skoor. See võib olla tingitud sellest, et suvisel perioodil inimesed ei ole nõus arstile minema. Mudel vajab lisa kalibreerimist suveperioodile.



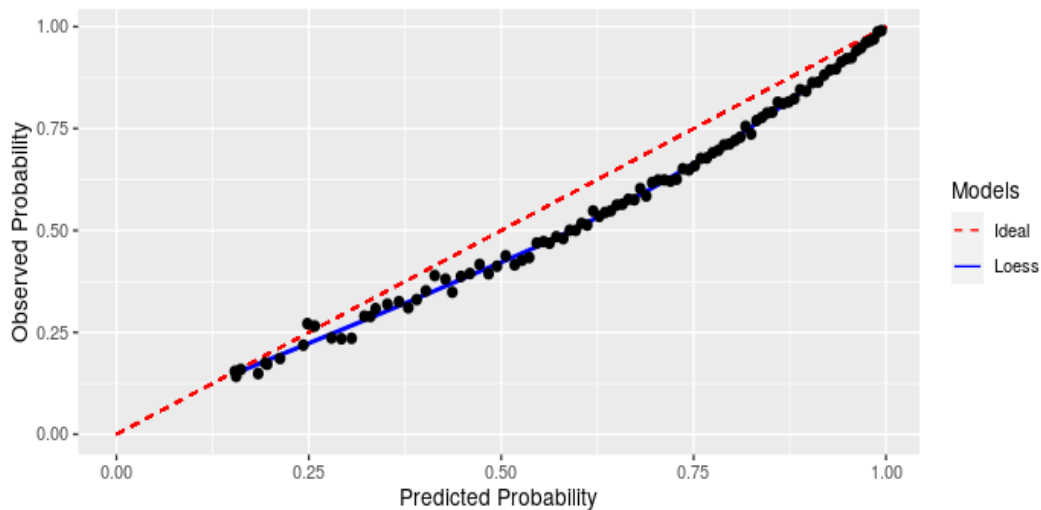
Joonis 9. Kalibreerimiskõver perioodil 01.09.2019-30.11.2019

Joonisel 9 on tegemist koroonaaegse perioodiga ning ennustatavate ja tegelike punktide asukohad on sarnased. Mudel on peaaegu ideaalselt kalibreeritud.



*Joonis 10.* Kalibreerimiskõver perioodil 01.12.2019-28.02.2020

Joonisel 10 on tegemist koroonaeelse perioodiga ning ennustatavate ja tegelike punktide asukohad on sarnased. Inimeste käitumismustris ei ole märgata muutusi. Mudel on peaaegu ideaalselt kalibreeritud.

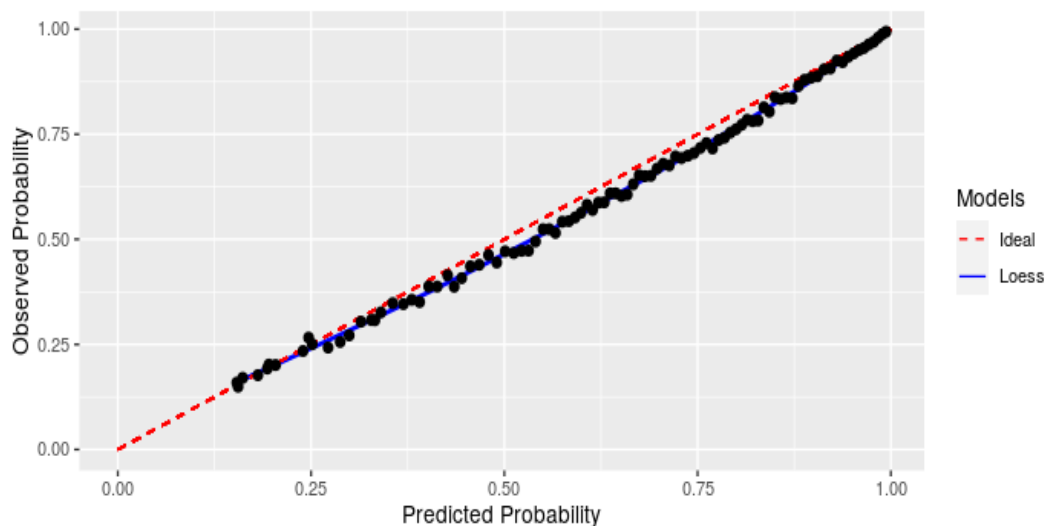


*Joonis 11.* Kalibreerimiskõver perioodil 01.03.2020-31.05.2020

Joonisel 11 on näha, et punktid asuvad  $x=y$  teljest madalamal, ehk ennustatav hinnang osutus kõrgemaks tegelikust vaadeldavatest andmepunktidest.

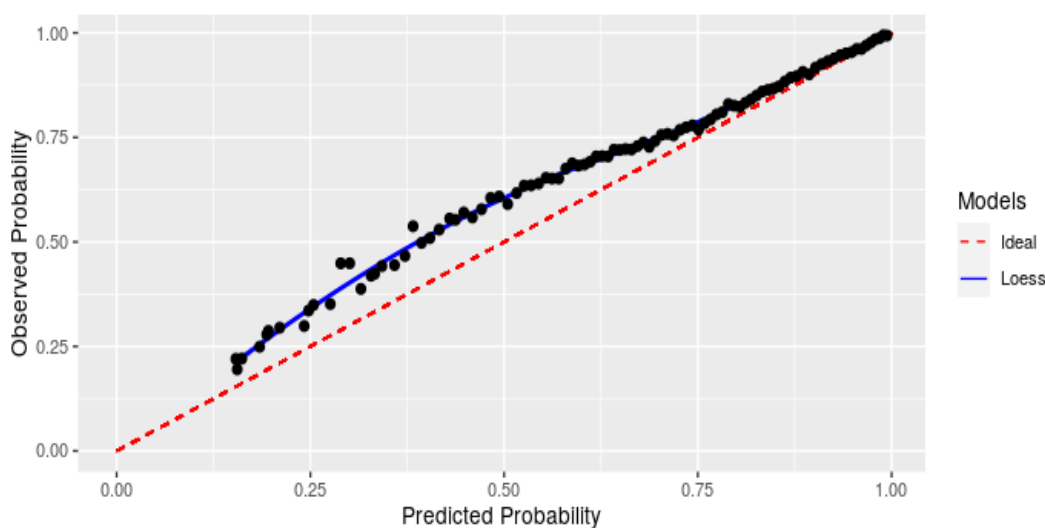
Esimesed piirangud seati 2020. aasta märtsis ning see võib olla põhjuseks, miks

toimus suur muutus võrreldes eelmise aasta sama perioodiga. Inimesed ei osanud ilmselt uute pandeemia reeglitega kohanduda ning arsti poole ei söandanud pöörduda.



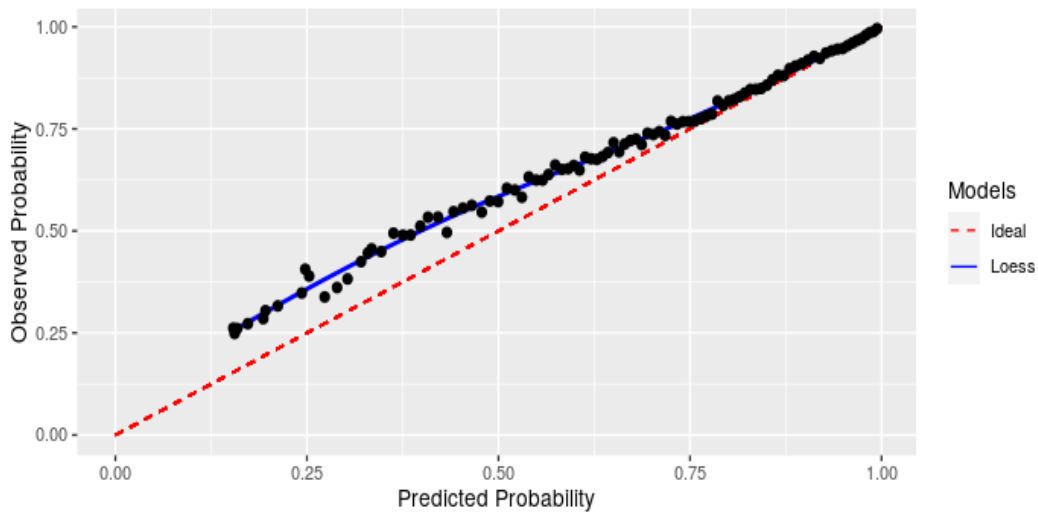
Joonis 12. Kalibreerimiskõver perioodil 01.06.2020-31.08.2020

Joonisel 12 on tegemist suvise perioodiga ehk tegelik arstivisiitide arv peakski justkui madalam olema. Kuigi kuna tegemist on koroonapandeemia perioodiga, siis tegelikud vaadeldavad andmepunktid on lähemal  $x=y$  teljele kui eelneval aastal.



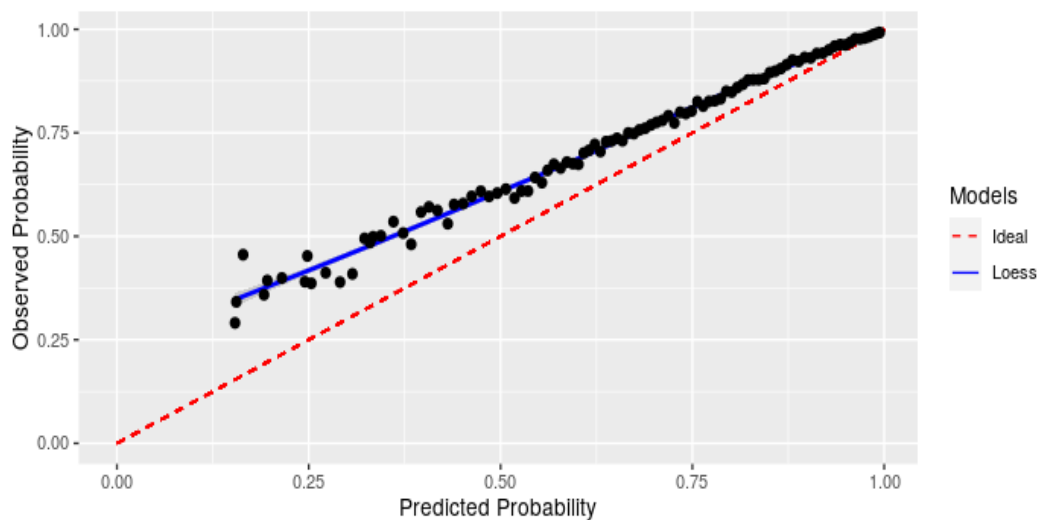
Joonis 13. Kalibreerimiskõver perioodil 01.09.2020-30.11.2020

Joonisel 13 on näha, et arstivisiitide arv on oluliselt tõusnud, seda ilmselt seetõttu, et antud perioodil oli uute COVID-19 nakatanute arv ööpäevas tõusnud üle 100 (Enam kui 100 uut nakatanud ööpäevas iga päev) [13] ning inimesed on teadlikumad COVID-19 sümptomitest ning selle potentsiaalsetest tagajärgedest. Seetõttu on ka rohkem arstivisiite isegi juhul, kui ennustusmudeli hinnang tulemile on madal.



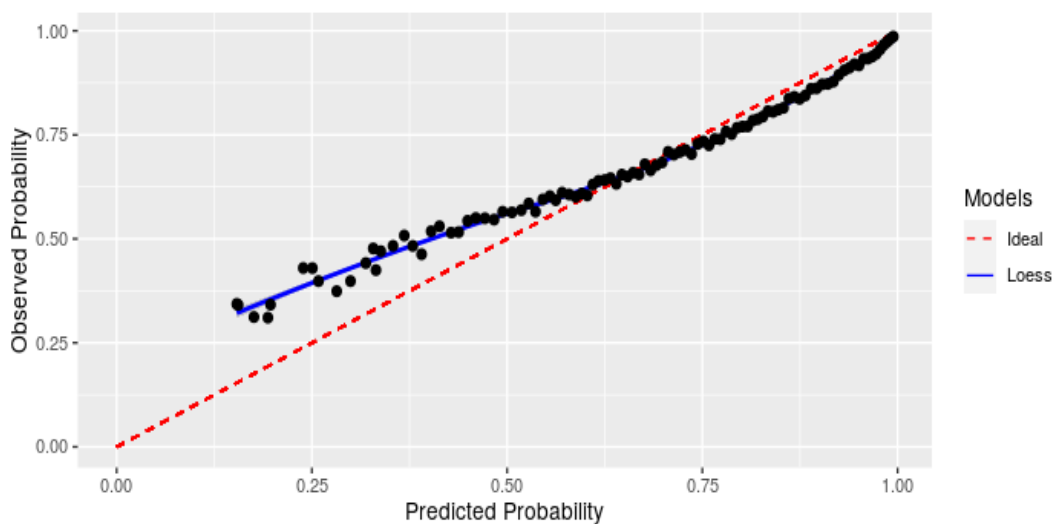
*Joonis 14.* Kalibreerimiskõver perioodil 01.12.2020-28.02.2021

Joonisel 14 on näha sarnast kulgu joonisega 13. Inimesed on teadlikumad ning käivad rohkem arsti juures.



Joonis 15. Kalibreerimiskõver perioodil 01.03.2021-31.05.2021

Joonisel 15 on näha, et inimesed käisid arstivisiidil oluliselt rohkem kui mudel ennustas. Põhjuseks võib olla see, et 2021. aastal halvenes koroonapandeemia olukord ning nakatumisnäitaja tõusis üle 1000. 2021 aasta märtsi keskpaigas jõudis kätte hetk, mil Eesti oli suhtelise nakatusnäitaja poolest maailmas esimene. [14]



Joonis 16. Suhe ennustuse ja tegelikkuse vahel perioodil 01.06.2021-31.08.2021

Kuna tegemist on suvise perioodiga, siis eeldatavasti on tegelik vaadeldav skoor madalam kui ennustatud risk. Tegelikkuses on need, kellele ennustatakse madala-

mat risk, käinud arstivisiidil rohkem kui mudel neile ennustas ning need, kellele ennustatakse suuremat riski, käivad arstivisiidil vähem kui ennustatud.

## Kokkuvõte

COVID-19 pandeemia, mis algas 2019. aasta lõpus, on avaldanud mõju tervishoiusüsteemidele üle kogu maailma, pannes proovile nende võime reageerida hädaolukordadele ja samal ajal pakkuda tavapärasest meditsiinilist abi. Käesoleva uurimistöö eesmärk oli hinnata arstiabi kättesaadavust kroonilisi haigusi põdevate inimeste seas esimese koroonapandeemia aasta jooksul. Selleks päriti Haigekassalt luba ligipääsu terviseandmetele perioodil veebruar 2017 kuni november 2021.

Algoritmiks, mille abil ennustusmudel loodi, osutus gradientvõimendus. Gradientvõimendi mudeli põhiidee on kombineerida mitu nõrka ennustajat (tavaliselt otsustuspuud) tugevaks ennustajaks neid järk-järgult optimeerides.

Loodud mudeli tulemused andsid teada, et ennustusmudeli kasutamine on üsna kasulik, sest tulemused erinevad reaalsusest üsna vähe. Mudeli AUC väärtus püsis enamasti 0.75-0.81 vahel, mis näitab, et mudel eristab tulemit mitte-tulemist üsna hästi. Küll aga viimase perioodi AUC väärtus langes hüppeliselt ning selleks, et mudel püsiks täpne, on vaja mudeli kalibreeritust pidevalt jälgida ja vajadusel uuendada.

Kuigi avalik huvi uurida ja ennustada COVID-19 haigestumist on vähenenud kui mitte kadunud, siis antud uurimistöö lähenemismetoodikat saab kasutada ka järgmistel viirushaiguste levikutel.

## Kasutatud kirjadus

- [1] <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>
- [2] Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-733.
- [3] Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579(7798):270-273.
- [4] World Health Organization. (2020). Coronavirus disease (COVID-19) Pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [5] Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA.* 2020;323(13):1239-1242.
- [6] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Velesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* 2020;181(2):281-292.
- [7] <https://koroonaakaart.ee/et> (21.08.2023)
- [8] OHDSI. The Book of OHDSI: Observational Health Data Sciences and Informatics. OHDSI, 2021.
- [9] Harrell Jr, F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis.* Springer.
- [10] <https://hastie.su.domains/Papers/ESLII.pdf> (21.08.2023)
- [11] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
- [12] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145-1159.

[13] <https://www.terviseamet.ee/et/uudised/covid-19-blogi-28-oktoober-oopaevagalisandus-125-positiivset-testi> (21.08.2023)

[14] <https://ourworldindata.org/covid-cases> (21.08.2023)

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Hendrik Piikov,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Covid-19 mõju krooniliste haiguste põdejatele Eestis**, mille juhendaja on Raivo Kolde, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Hendrik Piikov

23.08.2023