Eesti Teadusagentuur
Estonian Research Council

Programmi RITA tegevuse 1 projekti „Masinõppe ja AI toega teenused“ lõpparuanne

Euroopa Liit
Euroopa Regionaalarengu Fond

Eesti tuleviku heaks

RITA

Uuringu tellis ja uuringut rahastab Eesti Teadusagentuur Euroopa Regionaalarengu Fondist toetatava programmi „Valdkondliku teadus- ja arendustegevuse tugevdamine“ (RITA) tegevuse 1 „Strateegilise TA tegevuse toetamine“ kaudu.
Uuring valmis Eest Vabariigi Majandus- ja Kommunikatsiooniministeeriumi eesmärkide elluviimiseks.

Raporti autorid:
Mihkel Solvak, TÜ CITIS, tehnoloogiauuringute kaasprofessor
Jaak Vilo, TÜ Arvutiteaduste instituut, bioinformaatika professor
Sulev Reisberg, STACC OÜ, TÜ ATI, terviseinformaatika teadur
Sirli Tamm, STACC OÜ, teadustöötaja
Marek Oja, STACC OÜ, teadur
Kadri Ligi, STACC OÜ, doktorant
Taavi Unt, TÜ CITIS, analüütik
Andres Võrk, TÜ CITIS, analüütik
Peeter Leets, TÜ CITIS, analüütik
Liina Kamm, Cybernetica AS, vanemteadur
Andre Ostrak, Cybernetica AS, nooremteadur
Hiroki Kaminaga, Cybernetica AS, programmeerija
Triin Siil, Cybernetica AS, privaatsustehnika konsultant
Tanel Tammet, TalTech, Tarkvarateaduse instituut, professor
Risto Vaarandi, TalTech, Tarkvarateaduse instituut, vanemteadur
Sven Nõmm, TalTech, Tarkvarateaduse instituut, vanemteadur
Toomas Lepik, TalTech, Tarkvarateaduse instituut, nooremteadur
Veiko Lember, TalTech, Ragnar Nurkse innovatsiooni ja valitsemise instituut, vanemteadur
Steven Nõmmik, TalTech Ragnar Nurkse innovatsiooni ja valitsemise instituut , nooremteadur
Colin van Noordt, TalTech Ragnar Nurkse innovatsiooni ja valitsemise instituut , doktorant
Martin Ebers, TÜ Õigusteaduskond, infotehnoloogiaõiguse kaasprofessor
Paloma Krõõt Tupay, TÜ Õigusteaduskond, riigiõiguse lektor
Gaabriel Tavits, TÜ Õigusteaduskond, sotsiaalõiguse professor
Tanel Kerikmäe, TalTech, Õiguse instituut, professor

Lisaks raporti autoritele osalesid tööpakettide ja projekti elluviimisel:
Sven Laur, STACC OÜ, projektijuht
Harry-Anton Talvik, STACC OÜ, projektijuht
Meri Liis Treimann, STACC OÜ, andmeteadlane
Raivo Kolde, TÜ Arvutiteaduse instituut, terviseinformaatika kaasprofessor
Anne Ott, TÜ Arvutiteaduse instituut, magistrant
Joonas Puura, TÜ Arvutiteaduse instituut, doktorant
Daniel Urda Muñoz, TÜ Arvutiteaduse instituut, külalisteadur
Baldur Kubo, Cybernetica AS, projektijuht
Robert Altmäe, Cybernetica AS, arendaja
Andres Jõgi, Cybernetica AS, turvainsener
Dan Bogdanov, Cybernetica AS, Infoturbeinstituudi direktor
Uta Kührt, TÜ Johan Skytte poliitikauuringute instituut, spetsialist
Tiina Jaksman, TÜ Johan Skytte poliitikauuringute instituut, projektijuht

Sisukord

1. Executive summary

This study was commissioned and funded by the Estonian Research Council using European Regional Development Fund programme 'Strengthening of sectoral R&D (RITA)' activity 1 'Support for strategic R&D'. The research was conducted to support the policy goals of the Estonian Ministry of Economic Affairs and Communications. The study period was 14.10.2019-28.02.2022.

RITA project MAITT (Estonian abbreviation for machine learning and AI supported services) sought to answer six broader research questions:

1. Which public sector services would be suitable for implementation of automated decision supports using AI?
2. Does the public IT architecture support implementation of AI solutions?
3. How would automated decision support with AI affect the general performance of the public sector?
4. Does the current legal framework enable implementation of automated decision support with AI?
5. How would implementation of AI-based decision support enable to take advantage of opportunities and avoid risks in the development of public e-services?
6. What are the lessons learned from adoption of MVPs that should be considered in future projects in the public sector?

This was done by developing machine learning applications in the form of minimum viable products in four separate domains with four separate potential use cases:

1. *Cyber security* – a horizontally cross usable ML/AI MVP example**.** A cyber security monitoring tool that is usable horizontally across different institutions who have to monitor cyber threats in house.
2. *Physical security* – a vertically usable ML/AI MVP example. A fire risk scoring tool that is usable for household level prevention, institution level event response planning and long-term infrastructure planning.
3. *eHealth* - internationally usable ML/AI MVP example. A medical data tool for mapping different data structures onto international standards potentially scalable across different country contexts
4. *Labor market* - latent service demand detecting ML/AI MVP. An unemployment risk detection tool that will estimate the latent demand for a state provided skill improvement service bundles among at risk population.

These four use cases gave input into a wider assessment for what type of state provided services a machine learning/AI component would bring the biggest societal benefit, as well as what would be the potential impact on public sector functioning, and specific effect on end users working on operational service provision level.

In parallel to the experimentation and development of the machine learning models and applications a legal assessment evaluated how conducive the current legal environment is for wide scale usage of ML/AI in service provision, what legal principles should be observed when adjusting the regulatory environment to this technology and how should Estonia respond to the current European wide initiative to regulate AI systems and use cases.

The result of the study shows a lot of promise for machine learning/AI applications usage in state service provision, but also many, mainly non-technical and legal hurdles to actual deployment and daily usage.

ML/AI methods would best fit for situations where decisions need to be made under uncertainty, i.e. without good knowledge on the likely outcome of the decision or the necessity to intervene with a service. Given the ability of ML/AI technology to produce precise predictions, the inherent uncertainty in decision making could be substantially lowered. Such situations are typically found when a service should prevent a societally or individually negative outcome and the intervention is designed to lower the risks for the at-risk population. The end result would be a more targeted narrow intervention that is both more efficient from the service owner organization's point of view, as well societally beneficial as the number of undesired events would be lowered and there would be less need for reactive ex post services that are designed to deal with the negative consequences like treatment of already developed diseases or providing support for people who become unemployed. Prevention is potentially also the most difficult use case to implement from a deployment point of view. But the high risks will bring high rewards if it works, as ML/AI would help to foster a safer society defined in a wider sense through better work security (labor market MVP), better physical security (fire risk MVP), better health (eHealth MVP) and better protection from cyber threats (cyber security MVP).

The experimentation with these applications resulted in the following lessons learned:

1. Empirical results show that excellent coverage with data allows to build ML/AI applications even for relatively difficult prevention use cases with high or very high prediction accuracy.
2. Though usually solving a narrow problem at the microlevel of operations, an ML/AI application output should be used vertically throughout the organization beyond the narrow deployment place, different aggregation options will allow for usage in managing operations and monitor service effectiveness and impact. If possible, a small additional investment in model building, given the initial investments into building a data dependent service support application, will allow for a continuous service impact evaluation, which is usable in scaling working services and changing or discontinuing service that don't bring the actual impact they were designed to provide.
3. High empirical accuracy comes with heavy cross-usage of data, which is currently still administratively cumbersome and in some specific use cases only possible in a scientific study and not for a deployed service support tool;
4. Access to different datasets is not comparable for all domains, but the organizational needs on an abstract level are very similar such as prevention in a low information environment, prediction needs that tend to be classification or regression problems, lack of information to what degree the service brings the desired effect. This paradox indicates that there is a need for a common way how to benchmark organizational problems against ML/AI capacity and the datasets that exist in Estonia upon which such a tools could be built. Ideally Statistics Estonia could be one of such institutions helping other organizations as it has an overview and access to data, as well the legal right to link and analyse most datasets.
5. Interviews with organisations for which the proposed AI/ML applications were developed confirmed a need to create "understandable AI/ML" as well as a need to create more knowhow to be able to properly procure and manage such systems. Though this is nothing new for any ICT development procurement an additional layer enters in the form the ability to be able to see which system exactly would be able to provide actionable information. Given that AI/ML technology is very good at predicting and less at explaining, model output or explanations that

don't provide a direct indication on how to intervene with a service, are bound to be less useful even though they might be highly accurate. A case in point is the unemployment prediction application where sociodemographic drivers of unemployment, that cannot be changed, could help to predict the outcome, but do not provide much actionable information for service intervention, whereas lack of skills as a factor would.

6. The general IT architecture does support building and deploying AI/ML tools as there is excellent coverage with data across many domains, database are easily linkable and data exchange services are abundant and relatively easy to develop. But ML/AI applications demand even higher level of data cross-usage (like a need to see both occurrences (1) and non-occurrences (0) of the phenomenon to be predicted) which highlight the need for smooth data exchange across administrative boundaries. Including a need to share information on the data generation process and distributional changes for almost all variables used in heavily data dependent AI/ML models in order to avoid or detect data mistakes that will start to skew partially independent automated ML processes and produce wrong outcomes for services that affect people's lives directly.
7. ML/AI supported services create the need for data science support as models need to be monitored, retrained and mistakes evaluated by someone who knows the method as well the domain. These skillsets go beyond standard IT support and would ideally be provided by someone who has a data engineer, applied statistics/mathematics as well as limited front and back-end development skills, i.e. is a data scientist. Currently this support capacity is not yet fully developed in most organizations.
8. ML/AI supported service application front-ends would have to be designed to enable a critical reading of the model output to demystify AI and allow for a meaningful interaction with the machine. The critical reading ability can be enhanced with smartly designed front-ends, well thought through usage procedures and instructions as well as training operational level service providers in what ML/AI does and what it does not do.

The legal analysis of ML/AI usage in service provision as well as specific use cases of the MVPs concluded:

1. Estonian needs to update recommendations on ML/AI systems development with recommendations on the ethical and legal requirements these systems need to follow in the public sector, as well as with the principles on what rights citizens have in communicating with e-government service providers entailed in the guiding principle of Estonian e-government (Eesti e-riigi harta).
2. Estonia should impose a requirement to conduct an algorithmic impact assessment on the implementation of any ML/AI system in specific use cases to limit or mitigate risks posed by such systems.
3. Estonia should refrain from implementing systems deemed to likely violate basic rights based on the algorithmic impact assessment. To ensure coherent and central evaluation criteria for algorithmic impact assessment a government level discussion on these principles is needed.
4. Given the pre-emptive nature of the Artificial Intelligence Act, Estonia does not need to adopt its own regulation on so called high-risk ML/AI system adoption, for the time being. Instead Estonia should make its positions heard on the implementation of AIA in EU law, as well as in the standardization work of standard setting organisations in Europe.
5. Current Estonian regulations do not set out concrete citizen rights for receiving explanations on ML/AI supported decisions specifically. Hence, the Estonian regulator should make the existing

rules clearer, or stipulate a concrete right of citizens to receive *ex post* explanations on ML/AI systems usage in the public sector.

6. Estonia should adopt regulations to protect citizens from ML/AI based discrimination. Given that EU regulations on discrimination do not hinder member states from adopting stronger anti-discrimination rules, Estonia has the freedom to consider various legal ways to inhibit ML/AI based discrimination in the public sector domain.

2. Sissejuhatus

Uuringu tellis ja uuringut rahastab Eesti Teadusagentuur Euroopa Regionaalarengu Fondist toetatava programmi „Valdkondliku teadus- ja arendustegevuse tugevdamine“ (RITA) tegevuse 1 „Strateegilise TA tegevuse toetamine“ kaudu. Uuring valmis Eesti Vabariigi Majandus- ja Kommunikatsiooniministeeriumi eesmärkide elluviimiseks. Uuring viidi läbi perioodil 14.10.2019 – 28.02.2022.
Antud raport võtab kokku masinõppe ja AI toega avalike teenuste uuringu sisuks olevate masinõppe rakenduste piloteerimise katsed koos võimaliku mõju hinnangu ning õigusliku hinnanguga. Raporti eesmärgiks on tuua välja kesksed õppetunnid, mida selliste rakenduste loomine andis, koos soovitustega, mida arvesse võtta kui tulevikus masinõpet Eesti riigi teenuste osutamisel suuremas mahus kasutama hakata. Uuringus käsitleti lähemalt kuut uurimisküsimust:

1. millised avaliku sektorid teenused oleksid kõige sobivamad masinõppe ja AI otsustustoe rakendamiseks;
2. kas riigi IT infrastruktuur on sobiv masinõppe ja AI laiemaks kasutamiseks;
3. mil viisil mõjutaks masinõppe/AI otsustustugede laiem kasutuselevõtt avaliku sektori toimimist;
4. kas praegune õiguskord toetab selliste tööriistade kasutuselevõttu;
5. mil viisil oleks masinõppe/AI toega teenuseid võimalik kasutusele võtta suurima kasu saamiseks samas tekkivaid riske minimeerides;
6. millised õppetunnid masinõppe/AI sisuga rakenduste katsetamisest on olulised tulevikus analoogsete projektide läbiviimisel avalikus sektoris.

Nendele küsimustele vastati kahel viisil.

Esiteks katsetati neljas omavahel mitteseotud valdkonnas masinõppepõhiste rakenduste väljatöötamist, alates vajalikele andmetele juurdepääsu taotlemisest kuni rakenduste võimalikele lõppkasutajatele kättesaadavaks tegemiseni. Neljaks valdkonnaks olid:

1. tööturg – loodi töötuse riskiskooride masinõppe ennustusmudel;
2. sisejulgeolek – loodi eluhoonete tulekahjude riskide masinõppe ennustusmudel;
3. eTervis – loodi andmete standardiseerimise lahendus ja terviseväljundite prognoosimise masinõppe mudelid;
4. küberturvalisus – loodi küberohtude tuvastamise masinõppe mudel.

Kuigi rakendused on väga erinevatest valdkondadest, on nad kõik üldisel kujul suunatud ühiskonnas olevate negatiivsete riskide paremale manageerimisele ja ennetamisele. Seega on antud projekti katusteemaks kuidas masinõppe abil saab suurendada turvalisuse taset ühiskonnas ning see turvalisus on defineeritud laialt, sisaldades nii sotsiaalse turvalisust (materiaalselt kindlustatud ja tervislikult elatud elu), füüsilist turvalisust ((tule)ohutu keskkond) kui ka virtuaalset turvalisust (kaitstus küberohtude eest). Mil viisil masinõpe seda turvalisust suurendada aitab on lähemalt välja toodud konkreetsete kasutuslugude juures. Kuigi tegemist on eksperimentaalsete katsetustega, ei ole see turvalisuse suurendamine utopistlik loosung, vaid lähtub konkreetselt Eesti andmestike peal saavutatud empiirilistest tulemustest, mis on kohati üllatavalt täpsed. Täpsed mudelid ise loomulikult turvalisuse tõusu efekti ei anna, nende mõju negatiivsete riskide ärahoidmisele avaldub vaid kui neid ka sihitult tegelike teenuste osutamisel laiaulatuslikult rakendama hakatakse.

Eksperimenteerimise tulemusena võib aga järeldada järgmist:

1. Tugevad alused masinõppe/AI massiivsemaks rakendamiseks teenuste osutamisel on olemas eelkõige väga hea andmetega kaetuse osas;
2. Kõik loodud prognoose pakkuvad mudelid saavutasid kõrge või väga kõrge täpsuse taseme ehk hoolimata valitud riskantsetest kasutuskohtadest on heade masinõppe mudelite ehitamine võimalik, kasutamine on aga eraldi probleem;
3. Masinõppe rakenduste väljundeid on soovi ja vajaduse korral võimalik vertikaalselt kasutada ehk kitsa ülesande rakendamiseks loodud mikrotaseme andmeid kasutav masinõppe/AI mudel toodab väljundit, mida on väikese lisainvesteeringuga võimalik kasutada valdkonnas vertikaalselt juhtimiseks ja teenuste efektiivsuse muutmiseks;
4. Masinõppe mudelite täpsus suureneb andmete ristkasutuse ja teisese kasutuse läbi, kuid selline kasutus on administratiivselt ja õiguslikult hetkel ikka veel oluliselt takistatud;
5. Andmetele ligipääsu protseduurid on olenevalt valdkonnast ja kasutusloost erinevad, samas kui organisatsioonide äriprobleemid, mida masinõppe/AI toega teenused lahendada aidata suudaksid, on abstraktsel tasemel üsna sarnased – ennetus vähese info olukorras, prognoosi vajadus klassifitseerimise või regressioonülesande kujul, teenuse mõju avaldumise ja hindamise puudumine. Teenuste arendamisest huvitatud organisatsioonidel oleks kasu kui nad saaksid oma probleemi sellisel kujul *benchmarkida* ja näha sellega seotud vertikaalseid kasutusviiside ideid ning eriti oluline - võimalikke andmestikke ja disaine, mis sellise probleemi lahendamiseks vajalikud on. Sellist benchmarkimise teenust võiks pakkuda ideaalis ESA juures olev masinõppe/AI toe soovituse tööriist või mingi ekspertsüsteem/eksperdid;
6. Intervjuud võimalike rakendavate asutustega kinnitasid vajadust seletatava masinõppe/AI järele ning organisatsioonide sisemise võimekuse järele olla kriitiline arenduse tellija. Targa tellija vajadus pole midagi uudset, kuid masinõppe juures lisandub vajadus saada aru ka loodava mudeli iseloomust, sest täpseim mudel pole alati seletavaim ning veel olulisemana ei pruugi see anda seletust, mis on otseselt kasutatav (*actionable*) ameti teenusega tehtaval sekkumiseks operatiivsel tasandil kuigi prognoos võib väga täpne olla;
7. Riigi IT arhitektuur on sobiv masinõppe suuremahulisemaks kasutamiseks eelkõige suuremahulise lingitavate andmetega kaetuse ning juba suures mahus toimiva standardiseeritud andmevahetuse näol, kuid masinõppe rakenduste jaoks vajalik veelgi suurem andmete ristkasutus (nn 0 ja 1 nägemise probleem vt osa 3.2.) toob kaasa veel suurema vajaduse andmete sujuva jagamise järele üle vastutusvaldkondade piiride, kaasa arvatud andmete genereerumise protsessi koordineerimise järele;
8. Teenuste omanikele tekib IT toe kõrval vajadus ka andmeteadusliku toe järele tagamaks masinõppe mudelite toimimist ja suure mõjuga kasutuskohtades potentsiaalselt mõjukate vigade vältimist. Hetkel ei ole see tugivõimekus veel olemas;
9. Masinõppe/AI tugede loomisel on lisaks mudeli täpsusele oluline luua kasutajaliidesed, mis lubavad mitteeksperdil aru saada, mida mudel teeb ja mida ta kindlasti teha ei suuda ehk demüstifitseerida masinõpe/AI. Seda tuleks teha nii kasutajaliidese targa disaini, kasutamise protseduuride läbimõtlemise kui ka reakasutajale masinõppe põhitõdede koolitamise kaudu.

Teiseks viidi eraldi läbi õiguslik analüüs masinõppe/AI kasutamise võimalustest tänase seadusandluse alusel üldiselt ning osaliselt rakenduste põhistest küsimustest tulenevalt neljas valdkonnas eraldi. Õigusliku analüüsi üldosa, mis käsitleb Eesti praeguse regulatiivset olukorda ja annab soovitusi, ning spetsiifiline MVP-de alased soovitused on toodud raporti osas 6. Õigusliku analüüsi tulemusena järeldati järgmist:

1. Eesti peaks täiendama juhiseid tehisintellekti süsteemide loomiseks lisades sinna teavet eetiliste aspektide ja õiguslike nõuete kohta tehisintellekti süsteemide kasutamisel avalikus halduses ning võtta arvesse Eesti e-riigi hartat, milles on loetletud inimeste õigused suhtlemisel e-riigi ametiasutustega;
2. Eesti võiks kehtestada algoritmilise mõjuhinnangu nõude, mis aitaks tuvastada ja leevendada võimalikke ühiskondlikke riske ja ära hoida kahju enne tehisintellekti süsteemi kasutusele võtmist;
3. Tehisintellekti süsteemide keelustamise asemel seadusandlikul tasandil võiks Eesti loobuda selliste tehisintellekti süsteemide kasutamisest, mille mõjuhinnang näitab, et süsteem tõenäoliselt rikub põhiõigusi. Võimalikud ohud, mis võivad tuleneda tehisintellekti kasutamisest avaliku halduse poolt, võivad olla väga erineva intensiivsusega. Seetõttu tuleks valitsuse tasandil algatada arutelu võimalike piirangute üle, et tagada ühtsed hindamiskriteeriumid ja nende ühtlane kohaldamine;
4. Eesti jaoks tähendab AIA (*Artificial Intelligence Act*) ennetav mõju, et esialgu on mõistlik hoiduda kõrge riskiga tehisintellekti süsteeme käsitlevate asjakohaste õigusaktide vastuvõtmisest. Selle asemel on soovitatav, et Eesti tagaks oma seisukohtade kuuldavaks tegemise AIA õigusloomeprotsessis ELi tasandil[1] ja võimalusel ka juba Euroopa standardiorganisatsioonide poolt tehtava standardimistöö käigus;
5. Eesti õigusaktid ei sisalda konkreetseid sätteid kodaniku õiguse kohta nõuda selgitusi tehisintellektipõhiste otsuste kohta. Seega võiks Eesti seadusandja läbipaistvuse tagamiseks kaaluda võimalust konkretiseerida olemasolevaid sätteid ja/või kehtestada eraldi selgesõnaline õigus saada tagantjärele selgitusi avalikus sektoris kasutatavate tehisintellekti süsteemide kohta;
6. Eesti peaks vastu võtma täiendavad seadusmuudatused kaitsmaks kodanikke masinõppel põhinevate diskrimineerivate otsuste eest. Kuna ELi diskrimineerimisvastane õigus ei takista liikmesriikidel sätestada rangemaid eeskirju inimeste kaitseks diskrimineerimise eest, on Eesti seadusandja vaba kaaluma erinevaid õiguslikke võimalusi, et ära hoida diskrimineerimist tehisintellekti kasutamisel avalikus sektoris.

Käesolev raport on ülesehitatud järgmiselt. Sissejuhatuse järgne osa kolm toob välja kus ja kuidas võiks avalikus sektoris masinõppe/AI toega teenuseid kasutada. Neljas osa annab detailse ülevaate loodud masinõppe rakendustest iga valdkonna kaupa eraldi. Viies osa käsitleb lühidalt võimalikku mõju avaliku sektori toimimisele. Kuues osa võtab kokku erinevatest osadest välja kasvanud soovitused koos lühikeste põhjendustega. Õigusliku hinnangu kokkuvõte on toodud lisas.

3. AI/ML kasutus avalikus sektoris

Masinõppe all on siin mõeldud algoritme või süsteeme, mille võimekus nähtusi seletada, prognoosida ning teatud funktsioone täita kasvab lisanduva kogemusega, AI või tehisintellekt on siin katustermin võtmaks kokku intelligentseid infosüsteeme või tarkvara, mis reeglina masinõppe abil teatud väljundit

[1] Eesti on juba oma seisukohas AIA kohta juhtinud tähelepanu sellele, et halduskulude ja muude väljaminekute kasv võib olla komisjoni hinnangust oluliselt suurem; Seletuskiri Vabariigi Valitsuse otsuse juurde EL tehisintellekti määruse eelnõu kohta 8-9, 11-12. Lisaks on Eesti märkinud, et tema arvates ei ole kõik AIAs sätestatud kohustused proportsionaalsed. Nagu seisukohas märgitud, on mõningaid kohustusi nende praeguses sõnastuses raske või võimatu täita, näiteks kohustus omada veatuid ja igakülgseid koolitus-, valideerimis- ja katseandmeid.

Euroopa Liit
Euroopa
Regionaalarengu Fond
Eesti
tuleviku heaks
RITA

produtseerivad. Kuna piiri tõmbamine algoritmi(de) ja süsteemi/tarkvaralahenduste vahel on kohati keeruline, on neid termineid allpool kasutatud enamasti koos.

3.1. Otsustamine suurema ebakindluse olukorras

Masinõppe ja AI meetodite suurim lisandväärtus tuleb nende võimekusest tuvastada suuremahulistes andmestikes meid huvitavat väljundit/nähtust põhjustavaid komplekssеid mittelineaarseid seoseid oluliselt paremini kui seda suudavad klassikalise statistilise analüüsi tehnikad (James et al. 2013). Teisisõnu on nende alusel võimalik esmapilgul kaosest korrastatud mustreid leida ehk vähendada otsustamise juures otsuse tulemuse osas vaikimisi alati esinevat ebakindlust (Gilboa 2009).

Samas on selge, et mustrid saavad esineda vaid olukordades, kus meid huvitavat nähtust genereerib mingi mittejuhuslik protsess või mehhanism. Seda mehhanismi või protsessi suudavad need meetodid replikeerida vaid selles ulatuses, mis sisaldub meie vaadeldud andmetes. Mingi osa on aga andmete poolt katmata ehk nähtust oma kogu komplekssuses me reeglina vaadelda ei suuda ning lõpuks on olenevalt nähtusest mingi osa temast ka juhuslik. Olemasolevates andmetes sisalduva mehhanismi identifitseerimiseks ehitatud liialt lihtsustatud mudel, andmetega mitte kaetud mehhanismi osa ning juhuslikkuse komponent summeeruvad kokku ennustusveaks. Otsustustoe eesmärgiks on otsustamise juures seda viga, ehk ebakindlust otsuse tulemi suhtes, täpsema prognoosiga vähendada.

Masinõpe/AI pakub sellise võimaluse, sest suudab komplekssemate mudelite abil vähendada seda vea osa: 1) mis tuleneb liiga lihtsustatud mudeli kasutamisest; 2) ja seda osa, mis tuleneb meie võimetusest klassikaliste statistiliste tehnikatega kasutada kogu olemasolevat andmete hulka, mis antud nähtuse mehhanismi leidmiseks kasulik olla võiks. Kokkuvõttes peaks seega otsuse kvaliteet paranema, sest otsuse aluseks on päris elu komplekssusele täpsemini sobituv mudel. See masinõpe/AI võimekus viitab ka otsesetele kasutuskohtadele, mis lubaksid kõige paremini nende meetodite olemuslikke eeliseid riigi teenuste arendamisel arvesse võtta.

Suhteliselt suurem kasu võiks tõusta olukordades, kus:

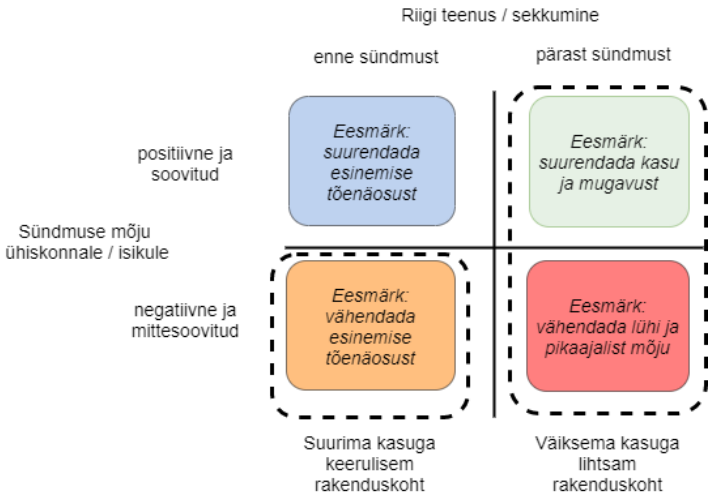
1) teenuse osutamisel on vaja langetada mingi otsuseid teatud sekkumise läbiviimise osas ***suurema ebakindluse*** olukorras. Mida suurem on teadmatus otsuse õigsuse ja selle tulemi osas, seda suhteliselt suurema kasu tooks masinõpe/AI rakendamine ebakindluse vähendamisel teenuse osutamisel;
2) nähtuse tekke mehhanismist suur osa **on *kaetud vaadeldud andmetega***, mille seest mehhanismi identifitseerida saab. Kusjuures need andmete korje ei pruugi olla/ei ole reeglina otsustaja otsese kontrolli all;
3) otsused on ***osaliselt standardiseeritavad*** ehk korduva mehhanismi tuvastamisel oleks tagajärjeks sarnane otsus;
4) madalaima tasandi otsuse aluseks olevad ***parameetrid on vertikaalselt kasutatavad*** kogu organisatsiooni või valdkonna juhtimiseks;
5) otsustel on ***olulised sotsiaalmajanduslikud tagajärjed,*** kas oluliste mõjudega väljundi tõttu, mille suhtes otsus langetatakse, või suure arvu kodanike tõttu, kelle suhtes otsust langetatakse.

Need kriteeriumid viitavad otseselt, mis tüüpi teenustele masinõppe/AI tuge kas vaja oleks ja/või milliste teenuste puhul olulist kasu oodata võiks.

3.2. Halbade riskide ennetav kontrollimine - ennetuseks loodud teenused

Hea näide masinõppe/AI ulatuslikuma kasutuselevõtu eelistest, aga ka kaasnevatest probleemidest, on riigi poolt viimasel ajal prioriteediks tõstetud digitaalsete teenuste sündmuspõhiseks ja pro-aktiivseks viimise initsiatiiv (Sündmusteenuse analüüs 2020; Digiühiskonna arengukava 2030). Kuna selliste teenuste ehitamiseks on vaja niikuinii integreerida kodanikke puudutavaid andmeid üle erinevate domeenide, oleks masinõppe elementide kasutamine teenuse teatud osades juba tehtud integreerimise pingutusele veel lisandväärtust andev. Keskne probleem, mis selliste teenuste juures lahendada tuleb, on kuidas muidu lahus domeene koos valitseda ja kuidas tagada teenuse toimimist kui teenuse omanik ja andmedoonorid ei kattu.

Joonis 1 toob üldistatud kujul ära võimalikud kasutuskohad masinõppe/AI toega teenustele lähtudes sündmusteenuse põhimõttest.



Joonis 1. Masinõppe/AI toega teenuste rakendamise kohad.

Sündmused ise väga laialt jaotatud kaheks:

a) sotsiaalselt soovitud/kasulikeks (näiteks lapse saamine, õppima asumine, juhtimisõiguse omandamine jne) ja;
b) mittesoovitud/kahjulikeks (näiteks liiklusõnnetusse sattumine, enneaegne surm, kooselu lahutamine, kuriteo ohvriks langemine, töötuks jäämine jne).

Riigi ja ühiskonna eesmärk sotsiaalselt soovitud ning kasulike sündmuste puhul on:

1) suurendada eri meetmetega nende esinemise tõenäosust;
2) maksimeerida nende realiseerumise järel ühiskondliku kasu ja mugavus.

Negatiivsete ehk mittesoovitud ja ühiskonnale kahjulike sündmuste puhul on eesmärk vastupidine:

1) minimeerida nende esinemise tõenäosus;
2) minimeerida nende realiseerumise järel negatiivne lühi- ja pikaajaline mõju:

Pidades silmas kõiki õiguslikke ja eetilisi piire riigi sekkumise osas inimeste ellu on selle lihtsustatud skeemi alusel riiklike teenuste eesmärk eelkõige kasvatada sekkumiste kaudu turvalisust ja heaolu, maksimeerida võimalik ühiskondlik kasu ja minimeerida võimalik kahju.

Teenustega sekkumise kohad on kas enne (*ex ante*) või pärast (*ex post*) vastavat sündmust. Hetkel piloteerimis- ja arendusstaadiumis digitaalsed sündmusteenused on valdavalt *ex post* disainiga, kus mingi vaadeldud päästik automatiseerib teatud teenuste kimbu proaktiivse pakkumise inimesele (Juhised määruse /…/2019).

Reeglina on *ex post* teenuse disaini või lähenemise eeliseks vajadus näha vaid väljundi esinemise fakti (lapse sündi või sellele eelnevaid fakte terviseloos, abiellumist, liiklusõnnetust, surma, töötust jne), sest mitteesinemine ei tekita vajadust teenuse järele. Taoliste reageerivate teenuste puhul on seega ebakindluse määr väiksem kui *ex ante* teenuste puhul.

Ex ante teenuste puhul on olukord keerulisem, sest toimunud sündmuse asemel on latentne tõenäosus selle esinemiseks, mis võib, kuid ei pruugi teatud aja jooksul tegelikuks sündmuseks realiseeruda. Selliste ennetuseks loodud teenuste tegevus on suunatud ühiskonnale kahjuliku ja kalli sündmuse nagu enneaegne surm, haigestumine, tulekahju, töötus, ärahoidmisele või nende esinemise riski märgatavale vähendamisele. Selle võrra võimendub ka vajadus neid riske kõige paremini ette hinnata suutva masinõppe/AI mudelite prognoosivõime järele.

Samas tekib sellises kasutuskohas üks oluline piirang kuna masinõppe/AI toe rakendamiseks on vajalik näha kogu relevantse populatsiooni, kelle seas sündmuse esinemine ja mitte-esinemine võimalik on. Ehk töötuse riski saab prognoosida vaid juhul kui kasutame nii töötute (1) kui töötavate inimeste (0) infot koos. Eluhoonete tulekahjuriski saab skoorida vaid kui näeme põlengute (1) ja mittepõlengute (0) infot. Haiguse kujunemise riski saab hinnata vaid kui näeme selle haiguse diagnoosiga (1) ja mittediagnoosiga (0) inimesi.

Riskide hindamisele ülesehitatud teenuse sisuks olev sekkumine on samas olemuselt kitsas ehk ideaalses olukorras toimub see vaid selles grupis, kus latentne tõenäosus on piisavalt suur sündmuse realiseerumiseks. Seega kitsas grupis võimalikult täpse sekkumise ehk efektiivse ja mõõdetud ennetustegevuse jaoks, on paradoksaalselt esmalt vaja näha ja analüüsida kogu kodanike populatsiooni infot, kus sündmus üldse teoreetiliselt esineda võiks.

Kokkuvõttes, lähtudes ülaltoodud masinõppe suurema võimekuse kriteeriumitest ning joonis 1 skeemist, saab väita, et masinõppe mudelite suurim lisandväärtus avaldub eelkõige just *ex ante* ennetavate teenuste puhul. Ehk teisisõnu olukordades, mida iseloomustab suhteliselt suurem ebakindlus otsustamise juures, ning olukordades, kus riigiasutus peab opereerima veel mitte toimunud sündmustega, vaid nende võimaliku esinemise ennustatud tõenäosusega. Selliseid olukordi kohtame eelkõige negatiivsete tagajärgede ennetamiseks loodud tegevuste juures. Võib ka väita, et tegevused inimesele negatiivse sündmuse ärahoidmiseks on riigi seisukohast õigustatum ja põhjendatavam eesmärk kui positiivse sündmuse esilekutsumiseks sekkumine. Seega on negatiivsete riskide efektiivsem haldamine ja ennetamine üks paljulubavamaid masinõppe/AI toega teenuste kasutuskohti. Antud uuringu raames läbiviidud masinõppe/AI mudelite arendustest on kõik loodud ennetava tegevuse täpsemaks sihitamiseks või riskide haldamiseks.

3.3. Tagajärgede täpsem haldamine - sündmuste järgsed teenused

Teine potentsiaalselt oluline kasutuskoht masinõppe/AI toega teenuste jaoks on sündmuste järgne riigi teenuse pakkumine (vt Joonis 1).

Kasutuskoht tuleneb asjaolust, et valitud sündmusi kogenud inimeste seas esineb omakorda sellises koguses heterogeensust, mille tagajärjeks on pakutavate teenuste suur erinevus nende eesmärgi saavutamise efektiivuses. Näiteks töötuks jäämise sündmuse, mis on üks sündmusteenuse võimalikke kandidaate (Sündmusteenuse analüüs 2020), järel pakub Töötukassa tuge ligi 70 000 isikule aastas eesmärgiga need inimesed võimalikult kiirelt ja sujuvalt uuesti hõivesse aidata. Töötuna registreerimise järel on isikutel erinev latentne tõenäosus teatud perioodi jooksul tagasi hõivesse liikuda. Töötukassa juba rakendab hetkel just sellisel juhul masinõpet otsustustoes OTT, mille eesmärk on tööturuteenuste sihitud suunamine vastavalt isiku riskisegmenti kuulumisele. Kuna erinevad teenused on olenevalt teenuse tarbija karakteristikutest omakorda erineva efektiivsusega kiire hõivesse tagasi liikumise võimaldamisel, on vajadus identifitseerida, milline sündmuse järgne sekkumine erinevatele gruppidele kõige efektiivsema mõjuga on. Seega esineb ka sündmuse järel pakutavate teenuste osutamise hetkel vajadus langetada otsus, milles on tagajärje osas teatud ebakindlus.

Ex post teenuste puhul oleks rakendusjuhud eelkõige olukordades, kus teenuse osutamisel on ametnikul/teenistujal või teenuse pakkujal valik sekkumise sisu osas ning ebakindluse olukord selle tulemuse osas, näiteks milline kitsam sekkumine (soovitus A, B või C) annab tõenäoliselt parima tulemuse antud olukorras. Siia alla langevad kõik olukorrad, kus täpsem pakutud teenus/nõuanne on teenistuja valitav ja tegevuse tulemus ei ole determinismlikult määratud või teenistuja sekkumine näiteks seadusest tulenevalt alati kohustuslik. Näiteks, töötukindlustushüvitisele kvalifitseerub isik automaatselt töötukindlustusseadusest tulenevate spetsiifilise tingimuste alusel, seega määramise otsus on sisuliselt deterministlik – hüvitis saadakse mingiks perioodiks mingis suuruses või mitte. Aga spetsiifilise tööturuteenuse isikule kättesaadavaks tegemine, eesmärgiga aidata ta tagasi hõivesse deterministlik otsus ei ole ning tark otsustustugi aitaks seda otsust teha viisil, mis annaks efektiivseima tulemuse minevikuandmetest õpitu alusel.

3.4. Täpsem osutatavate teenuste mõju hindamine

Kolmas masinõppe/AI toe rakendamise koht on üldine horisontaalselt kasutatav teenuste mõju hindamine masinõppe abil. Kuna teenusele masinõppe toe lisamine on võimalik vaid juhtudel, kus äriprotsessi osad on andmetega kaetud ja vaid läbi selle, et liidetakse teenuse poolt mõjutatud väljundi kohta käivad erinevad andmekogud, oleks loogiline lõplik samm lisada sinna juurde ka otsene teenuse mõjuhinnangu automatiseerimine. See on minimaalne lisainvesteering masinõppe toega teenuse arenduses, kuid annab olulise sisendi tööprotsessi, aga ka teenust teotavasse mudeli enda edasiarendamisse, sest:

1) masinõppe/AI abil peaks amet/riik suutma adresseerida paremini seda osa sündmuste esinemisest, mis ei ole juhuslikkusest tulenev või mis tuleneb mudelile teadaolevatest faktoritest (tunnustest millega väljund modelleeritakse). Juhul kui masinõppe/AI soovituse järgne tegevus on efektiivne ehk amet/riik suudab inimese riskitaseme toetavate tegevustega alla tuua, teeb ta paradoksaalselt tulevikus teenuse all oleva mudeli katki, sest sekkutakse nende faktorite muutmiseks, mis näiteks tulekahjuriski ülesse viivad. Üle aja on seega oluline aru saada, kuidas läbi sekkumise mudeli võimekus muutub ehk peab olema viisi võimalikult automaatselt tuvastata millal masinõppe/AI toega teenuse efektiivne osutamine selle sama mudeli katki teeb. Mudel läheb katki kui modelleeritavat väljundit hakkab tekitama järjest rohkem kas pelgalt juhus või siis see osa mehhanismist, mis andmetega kaetud ei ole. Iseenesest oleks see hüpoteetiline olukord, kus masinõpe/AI on oma ülesande täitnud, samas võib seda olenevalt olukorrast esineda ka väga kiirelt paari aasta jooksul. Näiteks Päästeametit huvitav eluhoonete tulekahjude langus ajas on olnud suhteliselt kiire ja nüüd ollakse sattunud platoole, kus edasine langus enam nii kiire pole.

See on sotsiaalmajanduslike olude paranemise ja Päästeameti tegevuse kombineeritud efekt. Masinõppe toega teenuse puhul aga oleks sel juhul täheldatud mudeli ennustustäpsuse langemist ajas, sest tulekahju tekkepõhjused, mida Päästeameti ennetamisega adresseerib, selle tulemusena ajas nõrgenevad;

2) tekib ametile oluline teenuse kvaliteedi ja efektiivsuse hindamise tööriist. Eestis rakendatakse riiklike teenuste mõju hindamist ulatuslikult läbi vastavate uuringute tellimise, kuid väga harvad on juhud kus see oleks masinõppe abil automatiseeritud. Lisaks esineb reeglina just ennetuse puhul tihti teadmatus, kuidas saaks üldse taolise tegevuse mõju hinnata.[2] Ainuke meile teadaolev juht on Töötukassa OTT projekti raames loodud rakendus MALLE (*Maschine learning in labor economics*) ehk mõjuhinnangu rakendus, mis automatiseerib tööturuteenuse mõju hindamise tööturule tagasi liikumise ja töötasu muutuse osas, kasutades kvaasi-eksperimentaalset sobitamist, mis on masinõppe abil automatiseeritud. Nii saavutatakse masinõppe toele ülemineku ühe kaasneva tulemina parem andmepõhine teenuse juhtimise tööriist;

3) tekib võimalus testida tulevikus täielikult automatiseeritud sihitatud riigipoolset teenust kujul, kus:
    a. hinnatakse näiteks mingi isiku või muu ühiku risk mingi mittesoovitud tulemuse suhtes (näiteks isiku nõusolekul tema töötuks jäämise risk);
    b. valitakse automaatselt sellise profiiliga isikutel seni suurimat mõju avaldanud teenus (koolituse/ümberõppe tüüp);
    c. soovitatakse talle seda teenust, koos päriselu andmete põhjal arvutatud põhjusliku mõjuga valitud väljunditele (palju muutub tuleviku sissetulek);
    d. vaadeldakse pidevalt prognoositud riski realiseerumist/mitterealiseerumist tulevikus (näiteks töötamine või mittetöötamine x-kuu möödudes) ja tagatakse masinõppe kaudu selle riski sündmuseks muutumise või mittemuutumise alusel pidevalt uute riskihinnangute kohendamine ning uute täpsustatud mõjuhinnangute loomine uuteks täpsemateks soovitusteks.

Me ei saa hetkel öelda midagi lõplikku selliste keeruliste teenuste täieliku automatiseerimise katsetuse õiguslike või eetiliste aspektide kohta, sest ilmselgelt peavad need olema pigem *opt-in* tüüpi mugavusteenused, kuid lähtudes riigi poolt tehtavate sündmusteenuste arendustest oleks selline eksperimenteerimine loogiline jätk sujuva digiriigi kogemuse arendamisel.

Mõju hindamise sisseviimine on kindlasti keerulisem suure mõjuga harvade sündmuste ennetuseks mõeldud tegevuste puhul. Näiteks tuleõnnetuste ennetuse puhul on üheks sekkumiseks kodukülastus, mille tulemusena tekib nn klassikaline mõju hindamise probleem (*fundamental problem of evaluation*) (Loi & Rodriguez 2012), kus sama majapidamist ei vaadelda koos ja ilma külastuseta. Antud näite puhul on lisaks keeruline rakendada klassikalisi ja lihtsasti tõlgendatavaid kvaasi-eksperimentaalseid kausaalse mõju hindamise meetodeid kuna väljundsündmus on omakorda äärmiselt harv. Samas on teaduslikke meetodeid, mida on võimalik rakendada ka sellistel erijuhtudel, näiteks meditsiinivaldkonnas välja arendatud väga harva esinevate haiguste tekkemehhanismi ja ravi efektiivsuse hindamiseks (Balzer et al 2016, Hu & Gu 2021) või tööstuses loodud nn ennustava hoolduse (*predictive maintenance*) mudelitel

[2] Vt näiteks: https://www.politsei.ee/files/Ennetus/politsei-ja-piirivalveameti-ennetust-kontseptsioon-sept-2018-.pdf?40da87a884

harvaesinevate rikete tuvastamiseks, mis põhinevad kõik andmekaevel ja masinõppel (Yadav 2017). Neid mudeleid saab omakorda masinõppe abil automatiseerida.

3.5. Vertikaalsed kasutusviisid

Vertikaalsed kasutusviisid tähendavad madalaima operatiivtasandi töö jaoks arvutatud parameetrite või muude mõõdikute erinevatele asutusesisestele või poliitikavaldkonna valitsemise tasanditele üles agregeerimist ning seal monitoorimis- ja juhtimisinstrumendina kasutamist.

MAITT projektis piloteeritud rakenduste puhul otseselt vertikaalseid kasutusviise ellu ei viidud, kuid ka siin on selgeid näiteid, kuidas kitsast ülesannet täitev AI võiks piisava täpsuse ja hea disaini korral vertikaalselt kasutatav olla. Kuigi teenustele masinõppe/AI toe loomine on võimalik vaid juhatuse ja kesktaseme juhtide ja teenuseomanike soovi korral ehk soov tuleb ülalt alla, võiks teoorias masinõppe/AI rakenduste suurim kasu tekkida kui operatiivtasandil tekkiv väärtus alt üles agregeeritakse ja eri tasanditel läbivalt kasutusse võetakse. Ka siin on Töötukassa OTT lahendus üks näide, mis on kitsalt võttes masinõppe abil indiviidide riskide leidmisel põhinev kliendikonsultandi tööriist ehk operatiivsel tasandil kasutatav asutusesisene andmepõhine teenus. OTT riskiskooridele on aga loodud erinevad agregeerimise reeglid analüütiliste ja juhtimisotsuste langetamiseks kuni neljal tasandil:

1) esiteks kasutab konsultant riskiskoori otseselt tööturuteenust saavate klientidele nõu, abi ja teenuse osutamisel;
2) teiseks kasutab konsultant riskiskoore enda kliendiportfelli tasemele agregeeritult enda töö efektiivistamiseks ja enesejuhtimiseks;
3) kolmandaks kasutab büroojuht riskiskoori büroo tasemele agregeerituna konsultantide koormuste juhtimiseks;
4) neljandaks agregeeritakse seda maakonna, regiooni ja riigi tasemele teenuste, poliitika ja tööturu dünaamika analüüsiks ja juhtimiseks.

Selline asutusesisene vertikaalne kasutusviis näitab kuidas on kitsalt võttes ühest numbrist – isiku riskiskoorist – võimalik välja võtta maksimaalne kasu. See aitab ka ülevalt alla tulnud initsiatiivi ja innovaatilise idee operatiivsele tasandile maha müüa ja veenda asutuse erinevaid tasandeid selles, et masinõppe/AI teenuse ehitamine ei ole keeruline ja kallis PR projekt, vaid horisontaalselt ja vertikaalselt tulemust toov arendus. ML/AI väljundite agregeerimine selle juures ei nõua sisuliselt lisaarenduse vajadusi kuna on tehtav täielikult ilma masinõppeta juba kasutuses olevate või riiulist ostetavate ärianalüütika tarkvaradega.

Näide võimalikust vertikaalsest kasutusviisist antud projektis väljatöötatud lahendustest oleks Päästeametile (PÄA) arendatud eluhoonete tulekahjuriski rakenduse väljundi eri viisidel kasutamine. Ennetuse kõrval on PÄA esmane ülesanne toimunud sündmusele kiire reageerimine. Latentne tulekahju tõenäosus realiseerub põlenguna sellistel asjaolude kombineerumisel viisil, mis ületab teatud kriitilise piiri millele tuleb selle järel reageerida. Ennetustöö eesmärk on sihitud kodukülastuste programmiga õnnetuse esinemise tõenäosust keskmiselt allapoole tuua.[3] Samas on tulekahju risk dünaamiline, ta kõigub keskpikalt (sesoonselt) ja lühiajaliselt (ööpäevaselt ja nädalapäeviti) teatud taseme ümber (vt Chhetri et al. 2018, Wang et al. 2014, Taylor 2011, Rohde 2010), päästekomando asukoht ja mehitatus on aga spetsiifilises ruumis ning ajas staatiline. Halbade riskide skoore agregeerides tekib võimalus lühiajaliste

[3] https://www.rescue.ee/et/kodu-tuleohutusalane-noustamine

kõrgete riskitasemete esinemise alusel näiteks päästekomandode 1-minuti valmisoleku meeskondade mehitatust suurendada või vähendada, arvestades, et valvekomandod veedavad rohkem aega ootel olles kui sündmustele reageerides. Ajas ja ruumist keskpikalt agregeerides oleks komandosid võimalik ajutiselt eelpositsioneerida vähendamaks reageerimiskiirust piirkondadesse, kus sündmus tõenäoliselt esineb. Pisut analoogsel kujul on hetkel hankimisel politsei väljakutsete prognoosimise mudel[4], mille väljundit tulevikus piisava empiirilise täpsuse saavutamisel just operatiivtasandil reageerimiseks kasutatakse.

Kokkuvõttes, masinõppe/AI toega teenuste arendamisel on loodava toe suurim praktiline kasu võimalik saavutada kui ei keskenduta puhtalt nn operatiivtasandile ehk kodanikuga otseses kontaktis olevatel kasutuskohtadele, vaid peetakse juba arendamise faasis silmas võimalusi suurendada loodava toe abil rakendusasutuste teenuste äriprotsesside efektiivsust ning kulutõhusust.

3.6. Masinõppe rakendamise valmisolek

Teenuste kasutuskohtade ettepanekute kõrval hinnati ka valmisolekut selliste teenuste kasutusele võtmiseks. Lühidalt on seda käsitletud iga MVP kirjelduse kasutusviiside ettepanekute juures, siin käsitleme tekkinud üldisemaid järeldusi.

Kitsalt tehniliselt ei ole masinõppe/AI mudelit sisaldava teenus oluliselt erinev muust suure IKT komponendiga teenusest. Väga minimalistliku arhitektuuri näide on Töötukassa OTT, mis kasutab ka muude teenuste jaoks loodud olemasolevat andmejärve, kuhu kirjutatakse uuesti tagasi ka mudeli väljund. Masinõppe mudeli treenimiseks on vabavara R-i testserver ja riskide arvutamiseks *live* server. Toe kasutajad näevad väljundeid olemasolevatesse infosüsteemidesse integreeritud vaadetes. Testserveri mälu ja arvutusvõimuse vajadus on suur mudeli ümbertreenimise ajal, *live*-serveri tehnilised vajadused on oluliselt väiksemad. Suuremate andmemahtude, näiteks pildimaterjalide töötlemine, ja selliste võimalike kasutuskohtade juures, kus on vaja väljund arvutada võimalikult reaalajaliselt ja kuvada seda iseseisva rakenduse vaadetes, on tehnilised nõuded oluliselt keerulisemad. Kuid need on kitsamad tehnilised detailid, kus on oluline organisatsiooni IT toe võimekus ja seda antud projektis ei hinnatud.

Riigi IT arhitektuuri masinõppe laiemat rakendamist ei takista, hajusarhitektuuriga omavahel ühendatud andmekogud lubavad printsiibis andmeid ristkasutada ning on olemas andmevahetuse jaoks väljatöötatud teenused. Samas identifitseeriti viis võimalikku aspekti, mis ei ole uudsed, kuid mis võimenduvad juhul kui masinõppe/AI toega teenuseid suuremas mahus kui seni kasutusele võtta soovitakse. Need toovad endaga tagajärgi asutuste IKT toele ja teenuste pakkumist toetavate infosüsteemide kasutajaliideste disainile:

1) andmetele juurdepääs, andmete ristkasutus ja selle seos hilisema täpse kasutuslooga;
2) andmete genereerumise ja kvaliteedikontrolli protsessi koordineerimine;
3) tugi andmeteadlase poolt;
4) mahukate arvutuste tegemine;
5) lugemisoskust toetav disain.

Järgnevalt kirjeldame neid kõiki põhjalikumalt.

[4] https://riigihanked.riik.ee/rhr-web/#/procurement/3667776/general-info

3.6.1. Juurdepääs andmetele ja ristkasutus

Esimene ja ilmselt olulisim on sujuv juurdepääs masinõppe/AI rakenduse loomiseks vajalikele andmetele. Rakenduste loomisel prooviti erinevaid andmetele ligipääsu ja linkimise lähenemisi saamaks aru, mil viisil oleks tulevikus masinõppe tuge lihtsam välja töötada ja millised lähenemised ilmselt nii kergelt rakendatavad ei ole.

Prooviti nelja erinevat viisi:

1) sensitiivsete registri- ja organisatsioonide andmete töötlemine ESA *sandboxis*;
2) organisatsiooni mittesensitiivsete andmete liidestamine avalike (registri)andmetega kolmanda osapoole juures ja sensitiivsete andmetega ESA *sandboxis*;
3) sensitiivsete registrite pseudonüümitud kujul väljaküsimine, liidestamine ja töötlemine kolmanda osapoole juures;
4) organisatsiooni sensitiivsete andmete töötlemine organisatsiooni sees.

ESA *sandbox*

Esimene lähenemise võib kokku võtta kui Statistikaameti (ESA) teadlase töökoha kasutamise nn *sandbox'ina*. Seda lähenemist kasutati töötuse riskide mudeldamiseks.

Töötuse riskiskooride rakenduse loomiseks taotleti alusandmetele juurdepääs ESAs. Kuna tegemist oli tundlike isikuandmetega ei võimaldatud neile juurdepääsu üle VPN tunneli, vaid ainult füüsiliselt ESA turvalisel teadlase töökohal. Seega kogu andmete linkimine, analüüs ja mudeli ehitus ning treenimine toimus ESA kontrolli all olevas keskkonnas. Selle lähenemine eeliseks on juurdepääs andmetele, mida ESA juba niikuinii riikliku statistikakorje raames korjab ning ka andmetele, mida on ESA-le erinevatel seotud eesmärkidel varem antud või antakse teatud regulaarsusega – näiteks Töötukassa tööturu teenuste saamise andmed. Teadustöö või uurimistöö eesmärgil on sellistes tingimustes võimalik pääseda aeganõudvatest juurdepääsulubade taotlemisest. Masinõppe mudeli väljatöötamise järel tõsteti ESA nõusolekul nende keskkonnast välja vaid mudelobjekt ilma sensitiivsete andmeteta. Mudelobjekti saab kasutada eraldi rakenduses juba teenuse osutamisel, kliendigrupi segmenteerimisel ja meetme sihitamisel või indiviiditasandi teenuse osutamise otsustustoena. Viimane on küll võimalik vaid juhul kui isik annab nõusoleku rakendusele tema isiklikku profiili ehk siis konkreetselt tema isikuandmeid liidestada. Oluline on rõhutada, et sellisel lähenemise sensitiivsed sisendandmed ega isikutele arvutatud isikustatud riskiskoorid ESAst kunagi välja ei liigu, vaid isikustatud skoorimine toimub vaid siis, kui selleks saadakse isiku nõusolekul uuesti tema sensitiivsed andmed. Teisisõnu korratakse siis juba väljaspool ESA keskkonda isiku riskiskoori leidmine tema enda antud nõusoleku ja andmete alusel.

Sellise lähenemise eelised on:

1. kiire ligipääs andmetele ja suhteline vabadus erisuguste andmete linkimisel ja analüüsil juba olemasoleva seadusandluse raames paika pandud piiride sees;
2. juba paigas ESA andmekaitse protseduuridest lähtumine, näiteks ühesed pseudo- ja anonümiseerimise reeglid ning andmete juurdepääsureeglid. Mis kõik kokku tähendavad, et protsess on andmete kaitse seisukohalt kontrollitud ja standardiseeritud.

Sellise lähenemise miinused on:

1. ESA teadlase töökoha tehnoloogiline piiratus, mis takistab väga suurte andmete ja arvutusmahukate mudelite väljatöötamist, kuid see on lahendatav lihtsalt mälumahu ja arvutusvõimsuse suurendamine ning protseduuride lihtsustamise kaudu ning projekti jooksul parandati MKMi, ESA ja RMITi initsiatiivil juba oluliselt töökohade riistvara võimekust;

2. *sandboxi* faasi järel masinõppe mudeli tootestamise kohmakus kuna mudelit tuleks regulaarselt ümbertreenida ESAs, objekti regulaarselt väljastada ning lahendada kohatine ebakindlus kuidas tagada mudelis kasutatud ebaregulaarselt uuendatavate andmete uuendamine juhul kui neid andmeid ei anta ESAle kohustuslikus korras regulaarselt.

Kokkuvõttes on sellise lähenemise skaleerimiseks vajalik teha ESAst sisuliselt masinõppe teenuste osutamise platvorm, kus mudelid andmete peal välja töötatakse, testitakse ning neid regulaarselt üle treenitakse ja mudelobjekt näiteks API kaudu masinõppe/AI otsustustoega teenuse omanikuks olevale asutusele kättesaadavaks tehakse.

Asutuse enda andmed koos avalike registriandmetega + ESA *sandbox*

Teise lähenemise puhul küsiti mittesensitiivsed andmed omanikelt otse välja ning prooviti seejärel sensitiivsete andmetega pilti parandada. Esimene samm oli katsetada mudeli väljatöötamist mittetundlike mikroandmete ja avalike registriandmete kombineerimise kaudu. Sellisel juhul piisab vaid teenust ise kasutava asutuse nõusolekust ning mudeli väljatöötamine ja kasutusse võtmine ei oma erilisi tehnilisi ega õiguslikke takistusi. Teiseks prooviti ka sensitiivsete isikuandmete liidestamist ESA *sandbox'is,* nägemaks kuivõrd palju see masinõppe mudelit täpsemaks teeb. Kuigi viimane on empiiriline asjaolu, sai selle katsetuse käigus ka selgeks piirid, kus hetkel kehtiv õiguskord andmete kasutamist lubab. Kuigi uuringu raames saab sellisel kujul ESAs jällegi andmeid kasutada, ei tohi teisene kasutus minna kaugemale kui uurimistöö läbiviimine (vt osa 4.2.5). Antud kasutusloo puhul tähendaks see, et tulekahju riskiskoori mudel on välja töötatud teisel eesmärgil kogutud delikaatsete isikuandmete abil ja seepärast ei tohiks Päästeamet seda mudelit oma haldustegevuse läbiviimiseks tulevikus ESA andmetel kasutada.

Jääb õhku küsimus, mida ikkagi mudeli kasutamise all mõeldakse ja tundub, et see vajab iga kasutusloo puhul eraldi täpsustamist. Näiteks kasutuslugu, kus asutuse enda andmete kõrval on ka teisel otstarbel kogutud sensitiivseid isikuandmeid kasutatud majapidamise tulekahjuriski arvutamiseks ja siis selle abil identifitseeritakse konkreetsed kõrge riskiga majad mida külastada, ei ole Andmekaitse inspektsiooni (AKI) arvates seaduslik.

Samas, mudeli väljatöötamine, mis on olemuslikult uurimistöö ja mille tulemust võib kuvada: a) kas tavalise staatilise raportina või b) tihedama regulaarsusega dünaamiliselt muutuva analüüsirakendusena, on võimalik, kuna tegemist on uuringu abil antava praktilist laadi abiga riigiasutusele tema igapäevatöös. Seega, kas uurimistööle järgneva sammuna sellise sensitiivseid andmeid kasutanud uurimistöö tulemusena tekkiva mudelobjekti kättesaadavaks tegemine elanikkonnale, näiteks iseenese elamise riskiskoorimiseks, oleks õigustatud, on ikkagi ebaselge. Ei ole ühest vastust, vaid see sõltub iga kord detailsest kasutusloost ehk mida täpselt kasutatakse (analüüsi, automatiseeritud abstraktset mudelobjekti, mudeli arvutatud umbisikulist või isikustatud väljundit, jne) ja mis eesmärgil täpselt (meetme planeerimiseks või näiteks külastatava majapidamise valimiseks, jne).

Pseudonüümitud ja liidetud registriandmed

Kolmandal juhul küsiti andmed pseudonüümitud kujul registri omanikult või töötlejalt eetikakomiteede lubade alusel välja ja töödeldi kolmanda osapoole juures. Protseduuriliselt on see keeruline ja aeganõudev ning dubleeriv protsess (vt osa 4.3.3). Selle läbimise järel on samas võimalik luua avalikult kasutavaid tööriistu, kus kasutaja otse pseudonüümitud mikroandmetele ligi ei peagi pääsema, kuid mis annavad detailseid haiguste kujunemise prognoose. Kuna sellisel kujul tööriistade jaoks ei ole vaja allolevaid masinõppe mudeleid uute isikuandmete abil väga tihedalt ümber hinnata, oleks nende kasutuselevõtt

võimalik ka juba hetkel uurimismeeskonna poolt muudes rahvusvahelistes projektides loodud algoritmide rakendamise infrastruktuuri abil.

Analüüs näitas, et nende andmete abil on võimalik luua ka selgelt indiviidi tasandi ennetuse või ravi jaoks kasutavaid tööriistu. Sellistel kasutusjuhtudel on aga hetkel osaliseks takistuseks meditsiiniseadmete tootmise/tootjavastutuse küsimus, sellele võimalike lahenduste pakkumine jäi projekti skoobist välja.

Asutuse enda sensitiivsed andmed

Viimaks kasutati küberohtude automaatse filtreerimise ja prioretiseerimise rakenduse väljatöötamisel lähenemist, kus kesksed sisendandmed mudeli treenimise faasis olid asutuse enda sensitiivsed andmed. Nendele ligipääsu tagamine ja andmete kaitse rakenduse loomise faasis on tagatav protseduuriliste ja tehniliste meetmetega asutuse enda sees ja otseseid juurdepääsu küsimusi ei teki. Ka hilisem tootestamine asutuse sees on sellisel juhul suhteliselt kergesti tehtav (vt osa 4.4) k una erinevalt teistest katsetatud rakendustest on siin tegemist rakendusega, mis on potentsiaalselt samal kujul kasutatav kõikides vastava suurusega organisatsioonides, hoolimata nende tegevusvaldkonnast. Pigem tekib küsimus, kas vastavad organisatsioonid on masinõppe rakenduse treenimise ja seadistamise faasis valmis andma võimalikele välistele osapooltele limiteeritud juurdepääsu ülimalt sensitiivsetele andmetele kuni sõnumisaladust sisaldavate andmeväljade tasemeni välja. Kuidas sel juhul tagada, et mudeli treenimisel kaasataks ainult absoluutselt vajalikud andmeväljad, vajab eraldi lahendust ja järelevalvet? Sellise spetsiifilise kasutusloo juures on skaleerimiseks vajalik luua organisatsiooni küberturbe spetsialistidele täpne arusaam, mida ja miks peab masinõppe algoritm „nägema“, et ta saaks väljundi korrektseks prognoosimiseks kohanduda. Samuti, mil viisil tagada toorandmete eeltöötlus selle spetsialisti poolt kujule, mis kaitseb organisatsiooni protsesse ja isikuid, aga võimaldab samas mudel ehitada, seadistada ja tööle rakendada.

Kokkuvõttes võib andmete taotlemise ja töötlemise protsesside alusel järeldada:

1. ESA on *sanbox'ina* kasutatav ja oleks masinõppe teenuse toe jaoks lahendus kui masinõppe/AI mudeli uuendamine ja treenimine ei pea toimuma pidevalt, vaid piisab näiteks kvartaalsest hindamisest koos mudelobjekti väljastamisega teenuse omanikule. Selleks oleks vaja anda ESA-le suurem võimekus ja volitus olla selliste tugede loomise platvorm. See on kooskõlas ESA missiooniga andmeid väärindada ja riigi põhifunktsioonide täitmiseks statistilise toe pakkumisega. Samas läheb see juba oluliselt kaugemale nende senisest statistika ja analüüsi pakkumisest;
2. Andmete teisese kasutuse keeld raskendab oluliselt masinõppe/AI tugede kasutuselevõttu, sellest on teatud kasutusviisidega võimalik mööda minna, kuid peab arvestama, et kehtib üldine printsiip, kus rikkalikum andmestik annab täpsema empiirilise tulemuse. Teisese kasutuse keerulisus toob seega endaga kaasa ebatäpsemaid tulemusi. Mis olukorras kaalub täpsuse vajadus ülesse teisese kasutamise problemaatilisuse on vajalik otsustada juhtumipõhiselt;
3. Olukordades, kus andmetele ligipääs piiratud ei ole, on masinõppe/AI tugede ehitamine piiratud vaid loodavate mudelite empiirilise täpsusega ning läbimõeldud kasutusviisi leidmisega;
4. Terviseandmete kasutamine teenuste ehitamisel lihtsustuks oluliselt kui ühtlustataks nende taotlemise protsessi ja kaotaks dubleerivate nõusolekute küsimise vajadus.

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

3.6.2. Andmete genereerumise ja kvaliteedikontrolli koordineerimise vajadus

Oodatult kinnitati üle tõsiasi, et rohkem ja paremad andmed annavad täpsema tulemuse. Arvestades projektis katsetatud rakendusvaldkondi, kus kas sekkutakse otse inimeste ellu või antakse soovitusi millel on reaalsed elulised tagajärjed, on just mudelite empiiriline täpsus oluliselt suurema kaaluga kui võimalikes muudes kasutusvaldkondades.

Täpsus saavutatakse eri andmete ristkasutusega. Eeldades hetkeks, et andmete juurdepääsu ja liidestamise probleemid on lahendatud, toob reaalne teenuse rakendamine välja suurema andmehalduse koordineerimise vajaduse kui seni erinevate relatsiooniliste infosüsteemide puhul.

Suuremahulisel ristkasutamisel tekib lisavajadus mitte lihtsalt tagada samade tunnuste vahetamine hoolimata andme uuenduste või äriprotsessi muudatustest, vaid ka nende jagatud tunnuste jaotuslike erinevuste testprotseduuride järele. See on vajalik vältimaks masinõppe mudelite varjatud vigade teket olukorras, kus neid kasutatakse näiteks indiviide mõjutavate otsuste tegemiseks ja kus mudel treenitakse tunnustel, mille jaotused ei ole alusandmete loomuliku muutuse, vaid tehniliste muutuste artefaktid.

Kui masinõppe toega teenus on suhteliselt reaalajaline, tekib omakorda vajadus eri domeenides korjatavate andmestike genereerumise protsesside koordineerimiseks. Eriti terav on see olukorras, kus olulised alusandmed korjatakse kellegi teise kui teenuse enda omanikasutuse poolt. Kolmanda osapoole andmete teisene kasutus muutub oluliseks teenuse tõrgeteta toimimise riskikohaks. Põhjalikum kirjeldus selle potentsiaalselt suure mõjuga niššiprobleemist on toodud lisas 1.

3.6.3. Tugi andmeteadlase poolt

Masinõppe/AI toega teenuse oluline osa on andmetele ehitatud masinõppe mudel ja selle kasutuslugu. Muus osas vajab sellise teenuse püsti hoidmine samasugust baas IKT tuge nagu ka ilma masinõppeta teenus. Seega lisandub standardsele infosüsteemi toele masinõppe mudeli monitoorimise, seadistamise ja perioodilise ümbertreenimise vajadus. Mudeli väljundi integreerimine olemasolevatesse infosüsteemidesse on juba arendustöö, kuid mudeli sisendandmete töötlemise skriptide, mudeli enda järelvalve ja selle väljundi API loomise töölõik vajab andmeteaduse kompetentsiga inimest. Ideaalis on tegemist andmeinseneri, rakendusstatistiku/matemaatiku ja limiteeritud kasutajaliidese arendaja oskusi katva ametikohaga ehk andmeteadlasega. Masinõppe/AI toega teenuste omanikud vajavad tulevikus kas majasisest või välist andmeteaduse kompetentsiga tuge ja riigi eesmärk võiks olla sellise kompetentsi tekitamine organisatsioonide IT- või analüüsiosakondade juurde.

Arvestades selliste oskustega inimeste järel valitsevat suurt nõudlust tööturul ning nende oskuste taset, oleks mõistlik püüda seda kompetentsi tsentraliseerida, sest igal organisatsioonil, kellel on sellise toega teenus ei ole tingimata võimekust või ka igapäevast vajadust hoida andmeteadlast palgal. Tasuks kaaluda võimalusi luua selline võimekus kas ESA või miks mitte RIA juurde ja võimaldada spetsialiseeruda andmeteaduses domeenipõhiselt, kuna viimane on mudelite hoolduse juures oluline. Ideaalis võiks mudelit hooldada see kes ta ehitas, aga sellist olukorda ei võimalda hetkel organisatsioonide andmeteaduse põhine võimekus.

Tsentraliseeritud kujul toe pakkume oleks ühe võimaliku variandina mõeldav kui keskse asutuse juures oleks valdkonnale spetsialiseerunud andmeteadlased. Näiteks on ESAs eraldi majandus- ja keskkonnastatistika osakond ning rahvastiku- ja sotsiaalstatistika osakond kuna nende valdkondade andmehõive, andmetüübid, standardid ja mudelid on väga erinevad. Analoogselt võiks olla sellisel kujul tsentraliseeritud sotsiaalvaldkonna andmeteaduse tugi, milles töötav andmeteadlane vastutaks mitme tugeva masinõppe ja AI komponendiga teenuse mudeli elutsükli eest erinevates asutustest korraga

(Töötukassas, Sotsiaalkindlustusametis, jne). Analoogseid rolle võib olla ka muudes domeenides, kus sellised teenused tulevikus kasutusse tuleks näiteks siseturvalisuse valdkond (Politsei, Päästeamet, TTJA, jne), jne.

3.6.4. Suuremahuliste arvutuste vajadus

Masinõppe/AI toega teenuste puhul on just mudeli väljatöötamise ja treenimise ajal reeglina vajalik suur salvestusmaht, mälumaht ja protsessorite tuumade arv, mida ei ole mõistlik teenust omaval organisatsioonil ise püsti hoida või omada. Suuremahulised, näiteks ka paralleelarvutust kasutava mudeli hindamisel oleks seega mõistlik pakkuda keskset infrastruktuuri tuge arvutuskeskuse (HPC) kujul. Siin projektis kasutati Tartu Ülikooli HPC-d terviseandmete töötlemisel ja mudelite loomisel. ESA teadlaste töökoht hetkel midagi sellist ei võimalda. Seega juhul kui sellise toega teenuseid tulevikus rohkem arendatakse oleks mõistlik, vastavate lubade ja õigusliku korra piires, kasutada juba riigi toel loodud arvutuskeskusi. Arvestades, et osaliselt loodud HPC-d juba pakuvad teenust ka Riigipilvele, tasuks kaaluda masinõppe/AI toega teenuste arvutusvajaduse rahuldamiseks riigipilve kaudu HPC-de riistavaraga. Olenevalt andmestikest eeldab see eraldi turvaprotseduure ja HPC-de vastavat andmeturbe sertifitseerituse taset.

3.6.5. Masinõppe/AI kriitilise lugemisoskuse vajadus ja toetav disain

Kasutajate tagasiside tõi selgelt välja masinõppe/AI rakenduste väljundite võimalikult ühese mõistmise vajaduse lõppkasutajate seas. See vajadus võimendub veelgi kui rakendusi kasutatakse nn operatiivtasandil otseselt kodanikuga suhtlemiseks/vahetuks teenuse osutamiseks ning langetatav otsus peab olema seletatav. Seletatavus tähendab siin, et leitud efektid ja produtseeritud väljundi väärtuse vastav tase, peab olema põhjendatud. Näiteks peab saama üheselt seletada miks hindab masinõppe mudel indiviidi riski teatud tasemel olevaks. Mudeli kriitiline lugemisoskus läheb aga seletatavusest kaugemale ja tähendab algelisel tasemel arusaamist, kuidas valitud teenuse mootoriks olev masinõppe algoritm tehniliselt oma tulemuseni jõuab, mis andmeid ta näeb ja mis dimensioone nii andmetest kinni püütakse ning kuidas tekivad mudeli vead (Jaton 2021). Erinevate kasutajaliidese seadistuste testimiseks teenuse eluea jooksul, lähtudes senisest teaduslikust kirjandusest, on inimese ja masinõppe/AI koostöös täheldatud kohati planeeritud ja tegeliku kasutuse erinevust. See ilmneb viisil, mida on keeruline tuvastada teenuse igapäevases kasutuses inimese poolt, näiteks, kas kasutaja järgib masina soovitust, sest ta nõustub sellega või sest tal on mugavam soovituse järgi käituda (Peeters 2020; Bailey ja Barley 2020). Seega on oluline tekitada teenuse osutajatel adekvaatne arusaam mida masinõpe suudab ja mida mitte.

Laialt üldistatult on tegemist siiski reeglina keskmistavate tehnikatega, mis otsivad andmetes teatud juhtude vahel jagatud mittelineaarseid tüüpmustreid, mille alusel prognoose väljastada. Ideaalne kasutaja peaks seega suutma mõista, et näiteks masinõppega leitud töötuse risk isikul A või eluhoonete põlengu riskiskoor ruudus F või haiguse K kujunemise riskitase tuleneb sellest, et inimesed on oma kitsatest parameetrites ja suurt hulka minevikuandmeid arvesse võttes väga sarnased minevikus midagi sellist kogenud isikutele või ruutudele. Eriti ennetuseks loodud mudelite puhul tähendaks see näiteks ka teenistujapoolset suutlikkust aru saada, et masinõppepõhise teenuse skoor ühe isiku puhul võib olla üle või alahinnatud kuna masinõpe pole teatud parameetreid näinud, mida aga tema isiku kohta teab/vaatleb. Kriitiline lugemisoskus siin tähendaks näiteks seda, et kui mudel ütleb, et inimese töötusrisk on väike, sest tal on hea haridus, varasem enam vähem korralik tööajalugu, aga vestluses temaga ilmneb, et tal on näiteks sõltuvusprobleem või terviseprobleem, mille kohta masinal infot pole, siis ta ei nõustu masinaga ja langetab teistsuguse otsuse kliendi kohta.

Oluline on masinõppel põhineva teenuse *front-* ja *back-end* lahenduste puhul pakkuda võimalikult palju mudeli lugemisevõimet toetavat diagnostikat, kaasa arvatud tagasisidet mudeli väljundile ning erinevaid hinnangute usalduspiire kui kasutusviis ja rakendatud ML/AI meetod seda lubab. See tähendab et ML/AI toega teenuste arendustöödes tuleks rakenduste *front-end* disainida kujul, mis soodustab omakorda masinõppe lugemisoskuse suurendamist. Kuigi *front-endi* disainis on oluline prioretiseerida kasutajamugavust ja tööprotsessi lihtsust, tuleks kriitilisel lugemisoskuse huvides luua sinna juurde erinevaid lihtsaid analüütilisi vaateid nagu näiteks mudeli kindlus prognoositud väljundi täpsuse osas. Üks viis selleks oleks hakata masinõppe/AI toega teenuste juures tegema süstemaatilisi infosüsteemi vaadete A/B testimist ja läbi selle kasutajakogemuse ja kriitilise lugemisoskuse optimaalne tasakaal leida. Juhul kui ei ole tegemist iseseisva rakendusega, tuleks siiski ka olemasolevatesse infosüsteemidesse ML/AI väljundiotsa integreerimisel selliseid kriitilist lugemisoskust toetavaid vaateid luua.

Lisaks on oluline demonstreerida näidisjuhtudega kuidas masinõppe mudel indiviidi andmeid kohtleb, näiteks kui masinas on juhumets peaks kasutaja teadma, et juhumetsas leitakse andmetest otsustuspuud, kuidas need tegelike kodanike andmete peal luuakse, kuidas nende pealt kombineeritakse kokku metsa ehk mis on üldised rakendatud reeglid antud lahenduses, ja kuidas sealt valitakse prognoos, mida tema näiteks kliendi kohta näeb. Lisaks mõistma, mis andmeid masin näeb, saamaks aru, mida see masin ilmselgelt arvesse ei oska võtta.

Kriitilise lugemisoskuse, seda tekitavate koolituste ning seda toetava *front-end* rakenduste süstemaatiline koostoime aitab maha võtta hirme masinõppe ja AI ees, demüstifitseerida nende peal jooksvaid teenuseid ning kaugemas perspektiivis luua suuremat usaldust selle tehnoloogia vastu. Kuna ML/AI on populaarne teema, aga erialaspetsialistidest väljaspool sisulist teadmist vähe, annaks selliste teenuste ehitamise üldprintsiipide rakendamine ka selge signaali, et riik võtab keeruliste tehnoloogiate juurutamisega kaasneda võivaid tagajärgi tõsiselt ja on agiilne vigade paranduses. On spekuleeritud, et just madala usaldustasemega kodanike seas võib teenuste automatiseeritus kaasa tuua isegi neutraalsema ja usaldavama suhtumise teenusesse kuna infosüsteemi automaatset otsust tajutakse erapooletuna (Miller ja Keiser 2020). Keeruline on näha, et see Eestis tingimata vajalik on, sest usaldus riigi vastu on kõrge ning usaldus digiriigi ja e-teenuste vastu veelgi kõrgem ning see tundub ka suhteliselt robustne arvestades erinevate šokkide nagu ID kaardi ROCA kriisi vähest mõju. Sellest hoolimata tasub masinõppe ja AI teenuste usalduse küsimust tõsiselt võtta, sest erinevates Euroopa riikides on potentsiaalselt suure mõjuga masinõppepõhised rakendused kriitika ja vastusseisu tõttu skandaalidesse sattunud ja deaktiveeritud – näiteks Austrias AMS nimelisest süsteemist loobumine tööturuteenuste osutamiseks[5] või Hollandi lastetoetuse petujuhtumite algoritmilisest tuvastamisest loobumine.[6]

4. MVP-de ülevaade

Järgnev osa annab detailse ülevaate projekti raames loodud AI/masinõppe rakendamisest neljas erinevas domeenis – tööturu ja tööjõu analüüs, eluhoonete tuleohutus, suuremahuliste terviseandmete analüüs ning küberohtude tuvastamine. Iga domeeni puhul tuuakse ära esialgne probleemipüstitus, kirjeldatakse andmekogud, mille abil sellele vastama hakatakse ja kuidas saadi või ei saadud andmetele juurdepääs. Lisaks antakse ülevaade kuidas rakendused loodi ja mida need teevad ning pakutakse välja erinevaid kasutusviise.

[5] https://www.derstandard.de/story/2000108890110/warum-das-ams-keine-ki-auf-oesterreichische-buerger-loslassen-sollte
[6] https://algorithmwatch.org/en/syri-netherlands-algorithm/

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

4.1. Töötuse riski prognoosimine

4.1.1. Probleemipüstitus

Eesti Töötukassa (TK) vastutab töötuskindlustuse korraldamise ning tööpoliitika elluviimise eest. Tööpoliitika kesksed eesmärgid on tagada tööealistele kõrge hõive määr ning töötuse ja pikaajalise töötuse ennetus. TK põhilisteks klientideks on registreeritud töötud ehk inimesed kelle töötuse risk on juba realiseerunud ning mille isiklike ja sotsiaalsete tagajärgedega tuleb tegeleda. Eeldatavalt oleks riski realiseerumise ennetus ühiskonnale vähem kulukas, aga ka indiviidi heaolule soodsam. TK on loonud teenuste paketi, mis on suunatud kõrgendatud töötuse riskiga, kuid hetkel tööturul hõivatud isikutele just sellise ennetuse eesmärgiga. Teenuse sisuks on erinevate oskuste juurde- või ümberõppe võimaluste pakkumine, kraadi või ametiõppesse astumise toetamine, muutunud tööpraktikatega kohanemise toetamine ning täiendav keeleõpe. Teenuste paketi eesmärk on adresseerida seda oskuste puudujäägi osa, mis antud isikul töötuse riski suurendab, aga ka seda osa, mis takistab teda paremale töökohale või palgale liikumast. Teenusele on määratud ka kvalifitseerumistingimused, mis on seotud töötava inimese praeguse sissetuleku, vanuse või tervisliku olukorraga ning keeleoskuse või kvalifikatsiooni olemasolu/puudumisega.

Karjäärinõustamise teenuse pakkumise puhul on teada selle sihtrühma suurus populatsioonis ehk kui suur osakaal vastab kvalifitseerimistingimustele. Kuid TK-l on raskendatud sihtrühmani jõudmine, sest indiviidtasandi andmeid nende kohta ei omata. Selle tõttu peab sihtrühma kuuluv isik hetkel esmalt ise Töötukassani jõudma ja teenust küsima, mille järel tekkib võimalus hinnata, kas antud isik ka tegelikult teenusele kvalifitseerub.

Ennetustegevus, mis aitaks vähendada töötuse tagajärgedega tegelemise mahtu, on seega paradoksaalselt keerulisem neljal põhjusel:

1) **Infopuudus sihtrühma osas** - Töötukassal tekib teadlikkus isiku sihtrühma kuulumisest või mittekuulumisest alles siis kui isik antud teenuse kohta huvi üles näitab. Puudub sarnane andmepõhine pilti indiviiditasandil ja riskiprofiilide detailne vaade nagu Töötukassal on olemas registreeritud töötute kohta neile teenuste osutamisel;
2) **Sihtrühma vajadus teenust aktiivselt küsida** - ennetuse töötamiseks peab suurema töötusriskiga isik seda endale teadvustama ja aktiivselt probleemile lahendust otsima, kuid paradoksaalselt on see grupp kõige vähem tõenäolisem aktiivselt ümberõppe teenuseid ise otsima. Kahjuks on teada, et enesetäiendusega tegeleb kõige rohkem see segment, kellel on kõige madalam töötusrisk. Näiteks on kõrgharidusega inimeste täiskasvanute koolitusel osalemise määr rohkem kui kaks korda suurem põhiharidusega inimeste määrast[7];
3) **Riski ulatuse valideerimise keerulisus** - ennetusteenuse sihtrühmaks on suurema töötuse riskiga isikud, range abimeetmele kvalifitseerumise tingimused on defineeritud parima teadaoleva info alusel, kuid neid ei ole pidevalt jooksvalt valideeritud riski ja selle realiseerumisel tekkinud sündmuste vastu ehk latentne töötuse riski jaotus võib olla selline, kus meetme praegused tingimused on kas liiga kitsad, laiad või väga erineva riskitasemega isikuid ühte kategooriasse grupeerivad;
4) **Teenuse mõju hindamise staatilisus** - pakutava teenuse tegelik mõju töötuse riski realiseerumise ärahoidmiseks ei ole seni hinnatud, nii töötavatele kui töötutele suunatud täiskasvanute täienduskoolituse mõju hinnati staatilise analüüsiga viimati 2015 aastal (Võrk et al. 2015). Samas

[7] https://andmed.stat.ee/et/stat/sotsiaalelu__haridus__taiskasvanute-koolitus__taiskasvanute-haridus/HTT38

on andmete olemasolu korral võimalik hindamine automatiseerida, mis lubab teenuse omanikel selle efektiivsust monitoorida ja vajadusel sisu muuta.

Antud probleemi ühe võimaliku lahendusena testiti projekti raames masinõppel põhineva rakenduse loomist, mis:

1) leiab automatiseeritult kogu Eesti tööturul hõivatute töötuse riski ja segmenteerib populatsiooni riskitasemete alusel;
2) leiab individuaalse profiili riskiskoori ja pakub selgituse, miks on sellise profiiliga isikul töötuse risk sellisel tasemel;
3) automatiseerib kvaasieksperimentaalsete meetodite abil ennetusteenuse mõju hindamise: a) hõive staatusele; b) isiku sissetuleku muutusele.

4.1.2. Riskimudeli väljundi definitsioon ja ülevaade kasutatud andmetest

Mudeli väljund defineeriti kahe erineva andmekogu alusel, üks Töötamise register TÖR ja teine Töötukassa arvel olevate töötute andmestik. Töötuse riskiskoorimise rakenduse ehitamiseks ja masinõppemudeli treenimiseks defineeriti huvipakkuv väljund järgmiselt:

1) kas hetkel *t* töötav inimene on järgneva 360 päeva jooksul **rohkem kui 180 päeva töötuna arvel**;
2) kas hetkel *t* töötav inimene on järgneva 360 päeva jooksul **vähem kui 180 päeva töötav**.

Kaks erinevat definitsiooni olid vajalikud kuna registreeritud töötamise lõpp võib päädida, kas töötuna arvele tulekuga (väljund 1) või muul põhjusel töötamise lõpetamisega (väljund 2), näiteks õppima minemise, Eestist lahkumise või lihtsalt inaktiivsusega. Definitsioonist lähtuvalt on tegemist nn klassifitseerimisülesandega, kus on vaja sisendtunnuste alusel klassifitseerida isik tulevikus töötavaks (0) või töö kaotavaks (1). Täpne sündmuse ennustus ei ole aga antud rakenduse eesmärk, vaid eesmärgiks on töötuse latentse riskiskoori leidmine, mis olenevalt riski tasemest võib, kuid ei pruugi, teatud perioodil reaalse töö kaotusega päädida. Eesmärgiks oli seatud just riskiskoori ja nende tingimuslike jaotuste leidmine, mis annab võimaluse populatsiooni erinevate riskide lõikepunktidega segmenteerida ning neid segmente edasiseks ennetavaks tegevuseks kasutada.

Eesti kohta tehtud uuringud hõivest töötusesse liikumiste kohta on analüüsinud nii ettevõtte tasandi tegurite olulisust kui ka inimeste karakteristikuid. Masso, Eamets ja Philips (2005, 2007) analüüsisid ettevõtte suuruse, vanuse, tegevusala ja asukoha olulisust töökohtade loomisele ja kadumisele. Meriküll (2011) selgitas töötusse liikumist teguritega nagu sugu, vanus, rahvus, eesti keele oskus, perekonnaseis, haridus ning elukoht. Analüüsid on näidanud, et nii ettevõtte kui inimese tunnused on olulised, sest töökohtade kaotust mõjutavad šokid tabavad ettevõtteid erinevalt ning mõjutavad ka ettevõtetes töötavaid inimesi erinevalt, sest nende töötajate tootlikkus ja seega kasulikkus ettevõtte jaoks on erinev sõltuvalt muutustest majanduskeskkonnas ja tootmistehnoloogias. Seega valiti ka antud masinõppe mudeli ehitamise jaoks tunnuseid kahe keskse dimensiooni, inimene ja tema tööandja, katmiseks.

Tabel 1 annab ülevaate, millised andmekogud mudeli ehitamisel kasutamist leidsid. Töötuse riski leidmiseks kasutati inimese sotsiodemograafilist tausta, piirkonna infot, tööajalugu, töötuse ajalugu ning tema tööandjate karakteristikuid. Nende dimensioonide kaupa leiti rida sünteetilisi tunnuseid – näiteks isiku sissetuleku keskmised muutused vaadeldud perioodil, tema sissetuleku jaotus erinevate tööandjate vahel, tööandjate arv ning nendelt saadud töötasude keskmised suurused ja nende suhted Eesti keskmisesse, aga samuti varasemate töötuse perioodide kirjeldused ehk kõik mis iseloomustab isiku

tööajaloo stabiilsust või ebastabiilsust. Lisaks võeti arvesse isiku varasemaid Töötukassa poolt pakutud teenuste kasutamisi juhul kui isik oli töötuna arveloleku perioodidel neid vastavalt kasutanud. Lõpuks liidestati mõju hindamise eesmärgil ka töötust ennetavate teenuste tarbimise ajalugu.

Tabel 1. Töötuse riski mudeldamiseks kasutatud andmekogud.

Andmekogu	Kasutatavad tunnused	Periood
Rahvastikuregister	Isikukood, sugu, emakeel, maakond, haridustase	2015 - 2019
EMTA Tulu- ja sotsiaalmaksu deklaratsioon (TSD) Lisa 1	Isikukood, tööandja ID, sissetuleku liik, sissetuleku suurus	2013 - 2020
Töötamise register TÖR	Isikukood, lepingu liik, tööandja ID, lepingu algus- ja lõpukuupäev	2015 - 2019
Töötukassa registriandmed Statistikaametis	Isikukood, töötuse perioodid, tööotsimise perioodid, töötuse lõpetamise põhjus, tööotsimise lõpetamise põhjus	2012 - 2019
Töötukassa registriandmed OECDle Statistikaametis	Isikukood, teenuste liik, teenuste algus ja lõpp	2012 - 2019
Statistikaametist tööandja tunnused	Tööandja ID, tööandja EMTAK	2015 - 2019
Äriregister	Tööandja ID; Majandusaasta aruandest ettevõtte põhivara; Majandusaasta aruandest ettevõtte käive	2015 - 2019

Alusandmete koosseisust tulenevalt on mudelite hindamiseks kasutatud ajavahemikul jaanuar 2015 kuni detsember 2018 töötanud isikuid. Treeningperioodi 2018. aasta lõpuga piiramine on võimaldanud ka 2018. aastal töötanud isikuid ühe aasta jagu jälgida, mis on vajalik väljundtunnuste korrektseks defineerimiseks. Lisaks on selgitavate tunnuste defineerimisel kasutatud TSD andmeid kuni 2 aastat ja Töötukassa andmeid kuni 3 aastat tagasiulatuvalt. Mudeli ehitamise aluseks olnud andmestiku suurus on 24,5 miljonit kirjet, milles sisaldub kogu Eesti tööealise populatsiooni igakuine töötamise info antud perioodil. Modelleeritavatel väljundtunnustel väärtuse 1 esinemine oli andmestikus haruldane, vastavalt 4,0% ja 1,1% rohkem kui 180 päeva töötuna arveloleku ja vähem kui 180 päeva töötamise kohta.

4.1.3. Kasutatud meetodid

Riskiskoori leidmine

Mudelid mõlema väljundi kohta loodi 12 kalendrikuu kohta eraldi lubamaks mudelitel võtta arvesse töötuse tugevalt sesoonset iseloomu. Kokku hinnati seega 24 masinõppe mudelit - jaanuari töötuse mudel väljund 1 ja 2, veebruari töötuse mudel väljund 1 ja 2, jne. Alusandmestikust 70% võeti treeningandmestikuks ning 20% testandmestikuks, lõplikuks valideerimiseks jäeti 10%. Sellist disaini rakendati järgnevate masinõppe algoritmide perekondadel:

1) logistiline regressioonimudel (logit)
2) regulariseeritud logistilise regressiooni mudelid (*glmnet*)
3) juhumets (*random forest*)
4) *gradient boosting* (XGBoost)

Otsustuspuude põhiste meetodite (juhumetsad ja XGBoost) piiratuseks on väljundtunnuse ühe grupi (vaadeldud töötus=1) väike arv ehk kuna töötus on Eestis madal, on tegemist andmestikus harva esineva

sündmusega. Parema tulemuse saavutamiseks kasutati erinevaid tasakaalustamise meetodeid, kus kas alavalitakse suurema grupi liikmeid ehk töötuse mitteesinemisi ning kaasatakse kõik töötuste esinemised; või vastupidi, kaasatakse enamus töötuse mitteesinemisi ja ülevalitakse töötuse esinemisi ehk kaasatakse neid andmestikus mitu korda. Testiti ka samaaegse ala- ja ülevalimise erinevaid kombinatsioone.

Erinevate algoritmide proovimisel lähtuti kolmest omavahel seotud kriteeriumist – mudeli empiiriline täpsus, mudelobjekti suurus ja mudeli arvutussressursi vajadus, milleks on näiteks hindamiseks kuluv aeg fikseeritud kettaruumi, aktiivmälu mahu ja protsessorite arvu juures ning ka prognoosimiseks vajalikku infot sisaldava mudelobjekti suurus, mis varieerub valitud meetodist sõltuvalt väga suurel määral. Kõikide mudelite puhul, välja arvatud logit mudel, kasutati paralleelarvutust.

Kui *logit*-i ja *glmnet*-i AUC väärtused on TÖR väljundi korral keskmiselt 0,804 ja TK väljundi korral 0,729, siis *XGBoost* mudeli AUC näitajad on vastavalt 0,818 ja 0,758, seega keskmiselt vastavalt 0,014 ja 0,029 ühikut suuremad. Logistilise regressiooni mudelitest annab täpsema tulemuse ka juhumetsa mudel, kuigi see jääb omakorda alla *XGBoost* algoritmile. Juhumetsa mudel on oluliselt suuremast ressursivajadusest, eelkõige mudeliobjekti suurusest, tingituna hinnatud küll vaid jaanuarikuupõhistel andmetel. Suvekuude, eelkõige juuli ja augusti, märgatavalt kõrgem AUC väärtus on tingitud põhiliselt noortest, kes teevad suvevaheajal ajutist tööd.

Tabel 2 toob ära sobitatud mudelite mudelobjektide suurused ja mudelite treenimiseks kulunud aja. Joonist ja tabelit võrreldes ilmneb, et kõige parem tasakaal empiirilise täpsuse, mudeliobjekti suuruse ja mudeli sobitamise juures annab *gradient boosting*.

Tabel 2. Töötuse prognoosimise masinõppe mudelite objektide võrdlus

Mudeli klass	Objekti suurus (MB)	Treeninguks kulunud aeg (minutid)
logit	0,75*	27
glmnet	0,06*	65***
*Juhumets***	1437,5	7***
XGBoost	13,7	3***

** mudeli rakendamiseks piisab tekstifailist, mis sisaldab infot tunnuste, kordajate ja standardvigade osas, mistõttu saab lõplik objekt olla palju kompaktsem*
*** hinnatud vaid 1 kuu (jaanuari) jaoks*
**** paralleelarvutust kasutades, arvutusressursiga kuni 128GB RAM ja 8 tuuma*

Kuigi logistiline regressioon on oma täpsuse poolest puupõhistest meetoditest mõnevõrra kehvem, on selle lihtsamast tõlgendamisest ja tõenäosusskooride suuremast kalibreeritusest tulenevalt otsustatud siiski selle mudeliklassi kasuks. Samuti on logistilise regressiooni mudeliobjekti lihtne Statistikaameti turvalisest töökohast välja tõsta ja seda väljaspool Statistikaametit rakendada - sisuliselt piisab vaid tunnuste koefitsientidest. See tagab ühtlasi parema andmete turvalisuse, sest ei ole vaja eraldi veenduda, et keerulist struktuuri omava mudeliobjektiga tuleks kaasa treeningvaatluste andmeid.

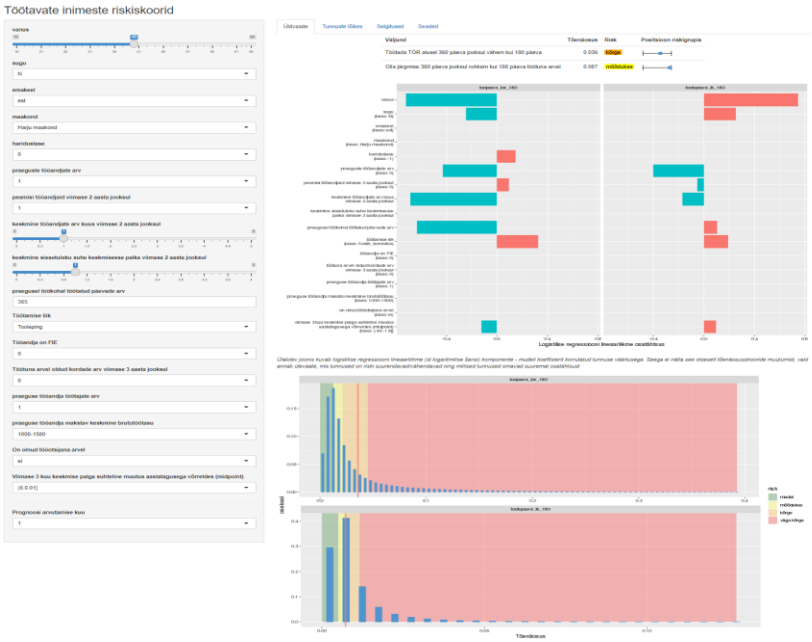
Ülevaade mõjuhindamisel rakendatud meetoditest koos nende sobivusega on toodud lisas 2.1. Rakenduse ehitamise detailne protsess on koos valitud lähenemise plusside ja miinustega toodud lisas 2.2.

4.1.4. Tulemuste kirjeldus koos kasutusviiside ettepanekutega

Rakenduse analüütilised vaated

Töötuse riskiskoori mudeli kasutajaliidese **üldvaade** on toodud Joonisel 2. Vasakus menüüribas saab kasutaja valida teda huvitava inimese profiili parameetrid, mille järel kuvatakse paremal antud profiili riskitaseme detailsem vaade, kus kasutaja näeb väljund 1 ja 2 kaupa:

1) täpset riskiskoori skaalal 0-1
2) isiku riskisegmenti skaalal madal, mõõdukas, kõrge või väga kõrge
3) ülevaadet millised isikut iseloomustavad tunnused skoori suurendavad või vähendavad ja kui palju ehk mudeli skoori seletust
4) üldist riskiskooride jaotust ning isiku suhtelist asukohta sellel



Joonis 2. Töötava inimese töötuse riskiskoori mudeli kasutajaliidese vaade alamlehel *Üldvaade.*

Rakenduses on realiseeritud lisaks vaated **tunnuste lõikes**, mis näitab riskide vaateid tunnuste kaupa. Tegemist on analüütilise vaatega, kus olevad joonised kuvavad tunnuste kaupa antud isiku tõenäosust uuritava tunnuse eri väärtuste korral. Näiteks 'vanuse' joonise korral on kõik ülejäänud tunnused fikseeritud ning vaadeldakse isiku tõenäosusi eri hüpoteetiliste vanuse väärtuste korral. Sellisel viisil saab mudeli tuvastatuid seoseid kasutajatele paremini seletada demonstreerides milliste hüpoteetiliste väärtuste korral oleks isiku risk kas suurem või väiksem. Lisaks mudeli seletamise ja mudeli väljundi korrektse tõlgendamise juures annab see täpseid viiteid milliste parameetrite muutus isiku riski alla tooks. Loomulikult ei saa selleks olla vanuse või muu sotsiodemograafilise näitaja muutus, vaid ikkagi nende faktorite adresseerimine, mis on muudetavad, kuid mõistmaks miks inimese risk on sellisel tasemel nagu ta on, on ka mittemuudetavate parameetrite hüpoteetiliste tasemete mõju näitamine informatiivne ja lubab paremini mingit sekkuvat teenust pakkuda.

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

Rakenduse vaade **selgitused** näitab erinevaid mudeli sisendparameetrite kodeerimise selgitusi ning osade tunnuste väärtuste taset. See on eelkõige vajalik mudeli tulemuste korrektseks lugemiseks.

Rakenduse vaade **seaded** lubab kasutajal valida riskide segmenteerimise loogikat, näiteks võib selleks olla kas erinevad protsentiilid või tõenäosuse tasemed, kus kasutaja määrab ise lõikepunktid, mille alusel isikud riskitasemetesse segmenteeritakse. Selle vaate eesmärk on võimaldada kasutajal nihutada segmenteerimise piirmäärasid vastavalt riskide reaalsele jaotusele. Riskiskoor ise on vahemikus 0-1, kuid väga suurel osal elanikest on risk madal, seega on ka segmentide piiri tõmbamine mõistlikum empiiriliselt jaotuse kujust lähtuvalt. Näiteks selliselt, et kõrge riskiga gruppi liigitatakse teatud protsentiil populatsioonist, näiteks 10% kõrgeima riskiga inimesi. Kuna töötus on harv sündmus ehk selle risk on reeglina madal, on mõistlik võimaldada segmenteerimise piire nihutada vastavalt rakendava asutuse analüütikute arusaamale milline harva sündmuse esinemise tõenäosus on piisavalt suur, et oleks vajalik püüda riigi poolt sekkuda antud sündmuse esinemise tõenäosuse alandamiseks.

Töötuse riskiskoorimise masinõppe rakendus on kasutatav kolmel viisil.

Esimene ja otseseim kasutus töötuse ennetamiseks suunatud meetmete parem sihitamine grupile, kelle töötuse risk on suurim. Segmenteerimine võimaldab seda sihtgruppi väga täpselt kirjeldada ning suunata meetmete teavitustööd täpsemini eriti kõrge riskiga gruppidesse. Masinõppe mudeli seletus lubab lisaks identifitseerida, miks teatud profiiliga isikuid just sellisele riskitasemele liigitatakse, milliste tunnuste väärtuste juures on riskitase milline ning läbi selle viidata millised sekkumised just selle grupi juures on parima efektiga.

Teine kasutusviis on individuaalne riskiskoorimine ja selle kaudu teenusele suunamine. Esialgu on seda mõistlik teha pigem nende isikutega, kes pöörduvad Töötukassa poole praegu töötuse ennetamise teenuse saamiseks kuna nad on ilmselgelt motiveeritud midagi ette võtma enda olukorra parandamiseks. Tulevikus võiks kaaluda rakenduse kättesaadavaks tegemist laiemale avalikkusele viisil, mis aitaks skaleerida ennetustegevust ja vähendada tulevikus selle läbi nende osakaalu, kelle puhul risk töötusena realiseerub. Oluline on läbi mõelda mil viisil seda teha nii, et skoorimine ei tekitaks soovitule vastupidist reaktsiooni. Selle juures on oluline tuua juurde ka rakenduse mõjuhinnangu poole kuvamine. Suhtluses karjääriteenuse omanikuga ning Töötukassa ekspertidega ilmnes, et suurim motivaatori ümberõppe koolitustel osalemisel on tulevikus saadav suurem töötasu. Arvestades, et koolitustel osalemine või muu enda oskuste parandamine samal ajal tööl käies nõuab isikult pingutust ja ohvreid vabaaja ning pereelu osas, on see igati arusaadav. Riskiskoorimise mudeli tulemuste kõrval saab seega isikule ka näidata, et juhul kui tal on keskmisest kõrgem töötuse risk, siis valides koolituse A, B või C vahel suureneb tulevikus tõenäoliselt tema sissetulek. Kohati abstraktse riski kõrval on see inimesele reeglina lihtsamini mõistetav ja ta näeb otsest kasu tulevikus, mis motiveerib selle nimel ka rohkem pingutama. Utoopilisemas tulevikus saab selliseid rakendusi osaliselt automatiseerida meetmele kvalifitseerimise hindamise tööriistadena ehk inimene näiteks laseb ennast iseteeninduskeskkonnas riskihinnata, näeb raportit enda kohta, aga seal juures ka teenuste valikut, mis sellise profiili jaoks loodud on, koos nende teenuste tegeliku kvaasieksperimentaalse mõjuhinnanguga just tema profiiliga isikute peal. Ehk ta näeb: a) mil viisil on tal kohati oskustest puudusi, b) kuidas seda parandada ja c) mis on selle parandamise mõju tema sissetulekule tulevikus.

Kolmas kasutusviis on suunatud ennetusmeetme teenuseomanikele ja Töötukassa teenuste kujundajatele. Kuna rakendus kuvab tegelikus populatsioonis toimuvate protsesside alusel arvutatud riskiskoorid ning suudab seda dünaamiliselt kohendada vastavalt alusandmetes toimuvate tööturu protsesside muutustele, on selle abil võimalik valideerida ennetusmeetmetele kvalifitseerumise tingimusi.

Näiteks saab näha kas ja kuivõrd langeb hetkel valitud sihtgrupp(id) kokku empiiriliselt kõige kõrgema riskitasemega gruppidega. See võib viidata vajadusele kvalifitseerumistingimuste kitsendamisele või laiendamisele vastavalt empiirilisele pildile. Teiseks võib tulevikus kaaluda ka riskiskoori enda ühe kvalifitseerimistingimusena arvestamisena või lausa sellisel viisil hinnatud riskiskooride alusel uute tingimuste seadmise dünaamiliselt, vastavalt sellele kuivõrd suurt osakaalu mudel riskantseks hindab ning kuivõrd suures mahus on võimalik ennetusmeetmeid laiendada.

4.2. Tulekahjude prognoosimine

4.2.1. Ülesandepüstitus

MVP 2 (Tuleohutus) keskendus Päästeameti andmestike uurimisele. Eesmärgiks oli: (a) parandada Päästeameti võimekust reageerida tulekahjudele ja võimalusel neid hoopis ennetada ning; (b) paremini suunata ennetustöö käigus tehtavaid koduvisiite ja automaatselt hinnata nende efektiivsust ja mõju. Enamjaolt on tulekahjud seal, kus on inimesed ning tulekahjude põhjused on tavaliselt sotsiaalsed, tehnilised või ilmast tingitud või kombinatsioon nendest faktoritest (vt näiteks Päästeameti aastaraamat 2018, 2019, 2020, Luht et al. 2016, Oidsalu 2018, Chhetri et al. 2018, Troitzsch 2016, Ronan et al. 2016, Yang et al. 2006). Tulekkahjuriskiga seotud tehiolud sisaldavad endas maja vanust, ehitusmaterjali, küttesüsteemi tüüpi ja tehnilist olukorda, elektrisüsteemi olukorda aga ka maja tüüpi (ridaelamu, individuaalelamu, kortermaja) ja lähedal asuvate hoonete tüüpe (KC et al. 2017, Clare et al. 2017, Ceyhan et al 2013). Samuti on tulekahjud tugevalt ajas mittejuhuslikult klasterdunud ehk esinevad rohkem teatud aastaaegadel, kuudel, nädalapäevadel ja kellaaegadel (vt Chhetri et al. 2018, Wang et al. 2014, Taylor 2011, Rohde 2010). Eluhoonete tulekahjude, kui üldiselt väga harvade sündmuste ennustamiseks on rahvusvaheliselt masinõppe ja AI meetodeid varem rakendatud (näiteks Nawaratne 2018, Garg 2018, KC et al. 2017, Balue' Arcoverde 2009, Bahrepuor et al. 2009, Yang 2006).

Lähtudes kirjandusest ja Eestis varem läbiviidud analüüsidest uurisime hoonete seisukorra, ilma ja sotsio-ökonoomsete aspektide mõju tulekahjude tekkele.

Jagasime töö kolme kasutusloo vahel.

1. Tuginedes tulekahjude ja tulekahjuohu, ehitiste ja ilma andmetele, genereerib süsteem automaatse tulekahju tekke riskiskoori. Tulekahjuohu all peame silmas kergeid tuleõnnetusi, mis ei läinud arvesse päris tulekahjuna, näiteks olid päästjate saabumiseks juba kustutatud.
2. Tuginedes tulekahjude ja koduvisiitide andmetele, analüüsib süsteem koduvisiitide mõju ning pakub välja tunnuseid, mis kõige rohkem mõjutavad tulekahju teket hoones.
3. Tuginedes tulekahjude ja inimeste sotsio-demograafilistele andmetele, toob süsteem välja sotsio-demograafilised tunnused, mis enim mõjutavad tulekahju teket.

4.2.2. Andmekogude ülevaade

Päästeameti andmestikest (Päästeinfosüsteemist) vaatlesime järgmiseid andmeid:

1. Tulekahjude ja tulekahjuohu andmed (kasutuslood 1, 2 ja 3) aastatel 2014-2020 (ligikaudu 8000 kirjet), sealhulgas:
   a) hoone asukoht (1 ruutkilomeetrise täpsusega),
   b) andmed hoone kohta (ehitusaasta, kasutusviis, elektri- ja teiste tehniliste süsteemide olemasolu),

   c) andmed tulekahju kohta (põhjus, asukoht, kannatanute arv, hukkunute arv).
2. Koduvisiitide andmed (kasutuslugu 2) aastatel 2018-2020 (ligikaudu 50 000 kirjet), sealhulgas
   a) kodu asukoht (kohaliku omavalitsuse täpsusega),
   b) andmed külastuse kohta (kuupäev, küsimustiku põhjal arvutatud riskiskoor),
   c) andmed kodu kohta (küttesüsteem, ehitusmaterjalid, elektrisüsteemi seisukord, suitsuanduri olemasolu),
   d) andmed elanike kohta (elanike arv ja vanusegrupid, teave alkoholi tarbimise ja suitsetamise kohta).

Ehitisregistrist (ehr.ee, avaandmed) vaatlesime hoonete andmeid (hoone asukoht, hoone tehnilised andmed) (kasutuslood 1, 2 ja 3).

Riigi ilmateenistusest (ilmateenistus.ee, avaandmed) vaatlesime ilmaandmeid (õhutemperatuur, õhuniiskus, tuul, sademed) tunniajase täpsusega tulekahjudele vastavatel perioodidel (kasutuslugu 1).

Statistikaametist vaatlesime inimeste sotsio-demograafilisi andmeid (ligikaudu 1 350 000 kirjet), sealhulgas

1. vanus, rahvus, haridustase,
2. puude olemasolu (jah/ei),
3. töö andmed, auto ja kinnisvara olemasolu.

4.2.3. Kasutuslugu 1: tulekahjude ennetamine tulekahjuandmete põhjal

Päästeameti andmetele juurdepääsu saamiseks konsulteerisime kõigepealt Andmekaitse Inspektsiooniga. Sealt saime kinnitust, et kuna andmestikus ei ole tegemist tundlike isikuandmetega, siis piisab andmetele juurdepääsu saamiseks Päästeameti otsusest. Algselt taotlesime tulekahjude andmed kohaliku omavalitsuse täpsusega. Hiljem taotlesime lisaks tulekahjuohu andmed ning tulekahjuandmed ruutkilomeetrise täpsusega. Selle otsuse tegime kuna kohalik omavalitsus võib olla piisav väikestes valdades, aga suuremates linnades ja valdades teeb see analüüsi raskeks, kui mitte võimatuks. Näiteks Tallinnas on Nõmme ja Pirita omavaheline kaugus suurem kui Pirital ja Viimsil, ning seega võib olukord Pirital ja Viimsis olla sarnasem kui Pirital ja Nõmmel. Ruutkilomeetrine täpsus on koordinaatidest võimalik tuletada ümardades.

Ehitisregistrist laadisime alla ehitiste andmed ning riiklikust ilmateenistusest ilmaandmed. Tulekahjude kohta on meil olemas täpne kuupäev ja kellaaeg. Seetõttu on lihtne neid andmeid ühendada ilmaga, mis oli täpselt tulekahju hetkel ning sellele eelnevatel tundidel. Aga selleks et masinõppe mudelit treenida, on vaja vaadelda ka olukordi, kus tulekahjusid ei toimunud. Kuna meil ei olnud võimalik tulekahju andmeid Ehitisregistri andmetega otseselt ühendada, käsitlesime Ehitisregistris olevaid hooneid sellistena, kus tulekahju ei toimunud.

Kahjuks on keeruline liita nende hoonetega ilmaandmeid, sest ei ole üheselt määratav, mis oli ilm sellel hetk, kui tulekahju ei toimunud. Kuna Ehitisregistri andmed on võrdlemisi staatilised ja ilmaandmed on ajas muutuvad, oli vaja valida meetod, mille abil liita nii-öelda negatiivsed tulekahjujuhtumid ilmaandmetega. Selleks vaatlesime erinevaid variante:

1. valida ilm iga hoone jaoks juhuslikult,
2. valida ilm läheduses tekkinud tulekahju aja järgi,

3. tekitada iga hoone kohta mitu kirjet erinevate ilmadega.

Kuna teine variant mõjutab andmestikku sisuliselt ning kolmas variant teeb negatiivsete kirjete hulga veel palju suuremaks kui ta eelnevalt oli, otsustasime kasutada esimest meetodit ja lisasime igale Ehitisregistri hoonele juhusliku hetke ilma.

Esialgne ülesandepüstitus oli binaarne - ennustada, kas vaadeldaval ruutkilomeetril tekib põleng. Selleks katsetasime binaarse klassifikatsiooni jaoks sobilikke tuntuid meetodeid: logistilist regressiooni, tehisnärvivõrke ning otsustuspuudel põhinevaid meetodeid nagu otsustusmets ja XGBoost. Nendest andis parima tulemuse just XGBoost ning kuna see meetod on ka üsna hästi seletatav (võrreldes näiteks tehisnärvivõrkudega), siis valisime just selle mudelite treenimiseks.

Tulemused

Rakendust on võimalik kasutada Päästeametis, et paremini ressursse suunata, viidates aladele, kus võib tekkida tulekahju. Rakendus ennustab tulekahju tekkimist kasutades Ehitisregistri ja tulekahjude andmeid. Rakenduse põhjal on võimalik ka valida, millistes regioonides ja hoonetüüpides on mõistlik teha koduvisiite ning kus suunatult edendada tuleohutust.

Erinevate meetoditega saadud parimad tulemused on toodud järgmistes tabelis 3. Mudelite testimiseks eraldasime andmed aastast 2020. Kuna andmed olid tasakaalust väljas, siis pidi seda arvestama nii mudeli treenimisel kui tulemuste tõlgendamisel. Nimelt on testandmetes umbes viis korda vähem andmepunkte (ruutkilomeetreid), millel on silt 1, s.t. põleng toimus, kui neid, millel on silt 0, s.t. põlengut ei toimunud. Tõlgendamisel on hea vaadata sildile 1 vastavat saagist, mis näitab, kui palju põlenguid mudel leidis, ning täpsust, mis näitab, kui paljud ennustatud tulekahjud tegelikult toimusid. Kui saagis on väike, siis ei leia mudel põlenguid üles, kui aga täpsus on väike, siis tähendab, et mudel toob välja liiga palju valepositiivseid. F1-skoor on meetrika, mis võtab kokku nii saagise kui täpsuse ning on seega tihti parim ülevaatlik skoor mudeli hindamiseks.

XGBoostiga proovisime ennustada tulekahjusid nii ilmaandmetega kui ilma. Seejuures kasutasime erinevatel juhtudel nii hetkeilma tulekahju ajal kui ka ilmaprognoose ja pikemaid ilmatrende (näiteks Kalmani filtri abil kohandatud nädala ilma). Seejuures osutusid parimaks mudelid, mis kasutasid ilmavaatlusi tulekahju ajal. Olulisemate faktoritena tõi mudel välja õhurõhu ja -temperatuuri, tuulekiiruse ning nähtavuse. Siiski näitas põhjalikum analüüs, et ilmaandmed tekitasid pigem üleliigset müra ning ei parandanud mudeli ennustusvõimekust, mistõttu need andmed hiljem mudeli treenimisel välja jäeti.

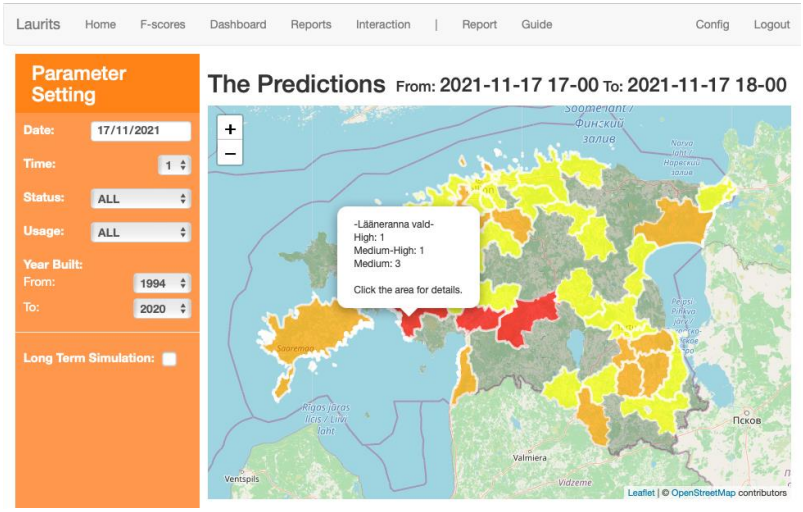
Tabel 3. Tulekahjude mudelite sobivuse näitajad

Mudeli tüüp	Täpsus	Saagis	F1-skoor
Logit regressioon			
Põleng ei (0)	0.96	0.91	0.94
Põleng jah (1)	0.66	0.83	0.73
Tehisnärvivõrgud			
Põleng ei (0)	0.95	0.96	0.96
Põleng jah (1)	0.80	0.75	0.77
Otsustusmets			
Põleng ei (0)	0.95	0.99	0.97

Euroopa Liit
Euroopa
Regionaalarengu Fond
Eesti
tuleviku heaks
RITA

Põleng jah (1)	0.96	0.77	0.85
XGBoost			
Põleng ei (0)	0.98	0.97	0.97
Põleng jah (1)	0.85	0.90	0.88

Lõplik mudel pidas kõige olulisemateks tulekahjuga seotud atribuutideks asukohta, aega, asustustihedust ja hoone tüüpi. Mudeli rakenduse vaade on toodud joonisel 3. Süsteem genereerib tulekahju riskiskoori vahemikus 0..1 ning esitab selle Eesti kaardil kohaliku omavalitsuse täpsusega. Alal klikkides avaneb tabel, mis näitab täpsemalt hooneid, millel on suurem tulekahju tekke oht. Süsteem võimaldab ka mudelit uuendada tegelike intsidentide infoga.



ID	Region	Status	Usage	Year_built	Electricity	Lat	Lng	Building density	Risk
02741020	Lääneranna vald	In use	11101 - Üksikelamu (I)	1996	Working	58.35173823.8565025			0.500792
12849212	Lääneranna vald	In use	11101 - Üksikelamu (I)	2003	Working	58.57279823.98998615			0.419187
22785708	Lääneranna vald	In use	11101 - Üksikelamu (I)	1998	Working	58.686677	23.835232	70	0.862563
32785734	Lääneranna vald	In use	11101 - Üksikelamu (I)	1999	Working	58.68463823.82808678			0.316890
42825264	Lääneranna vald	In use	11101 - Üksikelamu (I)	2003	Working	58.69342223.84257770			0.402002

Joonis 3. Tulekahjuriski rakenduse ekraanivaade (Lääneranna valla riskipildiga).

4.2.4. Kasutuslugu 2: tulekahjude ennetamine koduvisiidiandmete põhjal, koduvisiitide automaatne mõjuanalüüs

Koduvisiitide ja tulekahjude andmestike ühendamiseks väljastas Päästeamet vastavustabeli. See väljastati koos sama väljastusega, millega saime täpsustatud asukohad, ning nende hoonete kohta, kus oli toimunud koduvisiit ja pärast tulekahju, meile täpsustatud asukohti ei antud. Kahjuks ei olnud võimalik andmestike põhjal mudelit treenida, sest koduvisiite oli 50 000 ja ainult ligikaudu 40 juhul toimus pärast koduvisiiti tulekahju ehk 0,08% juhtudest. Neid juhtumeid on liiga vähe, et mingid usaldusväärsed mustrid saaksid avalduda.

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

Ka koduvisiitide automaatset mõjuanalüüsi ei olnud võimalik läbi viia, sest korduskülastusi tehti 70 juhul, millest umbes kolmandikul juhul koduvisiidil arvutatav riskiskoor langes, kolmandikul juhul tõusis ning kolmandikul jäi samaks. Kahel juhul tehti korduskülastust ka kolmandat korda.

Sekkumiseks oleva kodukülastuste arv ise on piisavalt suur võimaldamaks kasutada klassikalisi kvaasieksperimentaalseid meetodeid sekkumise ja sekkumiseta gruppide balansseerimiseks ja seal väljundi - ehk tulekahju – esinemissageduse võrdlemiseks nagu tehti tööturu rakenduse loomise juures. Erinevalt aga tööturu tööpaketist on siin väljundi esinemissagedus ülimalt harv kuna analüüsitavaks ühikuks ei saa olla lihtsalt majapidamine, vaid majapidamine ajas kuna tulekahju mõjutavad faktorid, mis on ajas muutuvad (ilm ja inimkäitumine). On teaduslikke meetodeid, mida on võimalik rakendada ka sellistel erijuhtudel, näiteks meditsiinivaldkonnas välja arendatud väga harva esinevate haiguste tekkemehhanismi ja ravi efektiivsuse hindamise statistilised meetodid (Balzer et al. 2016, Hu & Gu 2021) või tööstuses loodud nn ennustava hoolduse (*predictive maintenance*) lähenemised, mis põhinevad samuti andmekaevel ja masinõppel (Yadav 2017), kuid mis ei ole ilma lisatööta otseselt kohaldatavad antud probleemile. Seega lähtuvalt projekti skoobist hinnati mõjuhindamise arendamine metodoloogilised piirid ja leiti, et sellisel kujul see hetkel ilma olulise lisatööta see tehtav ei ole, mis ei tähenda, et seda võimatu teha on.

Ennetustöödest parema ülevaate saamiseks loodi rakendus, kus kohaliku omavalitsuse tasemel kuvatakse kaardile teatud perioodil toimunud tulekahjud ja koduvisiidid. Lisaks arvutatakse suhtarvud (tulekahjude arv koduvisiitide suhtes, tulekahjude arv elanike arvu suhtes ja koduvisiitide arv elanike arvu suhtes). Võimalik on valida, millist arvu kaardil valdade lõikes värvigradiendi (roheline kuni punane) abil kuvatakse. See annab hea visuaalse ülevaate ennetustööde planeerimiseks.

4.2.5. Kasutuslugu 3: koduvisiitide planeerimine sotsio-demograafiliste andmete põhjal

Kolmanda kasutusloo puhul oli kesksel kohal küsimus, milliseid olemasolevaid sotsio-demograafilisi andmeid on võimalik uuringus kasutada ja mis tingimustel seda võib teha. Uurijatel oli valida, kas taotleda juurdepääsu mitmele erinevale andmete allikale (nt Eesti Hariduse Infosüsteem (EHIS)) või küsida kõiki vajalikke andmeid Statistikaametist, kes kogub ja töötleb vastavaid andmeid riikliku statistika tegemise eesmärgil (edaspidi “konfidentsiaalsed andmed”). Mõlemal juhul on sisuliselt tegemist andmete teisese kasutusega, st andmed on algselt kogutud muul eesmärgil, kuid nüüd soovitakse neid töödelda teadusuuringu eesmärgil.

Statistikaametil on konfidentsiaalsete andmete teaduseesmärgil töötlemiseks välja töötatud eraldi kord, samas kui teised asutused lähtuvad enamasti isikuandmete kaitse seaduses (IKS) ettenähtud üldkorrast isikuandmete töötlemiseks teadus- ja ajaloouuringu ning riikliku statistika vajadusteks (IKS § 6). Lisaks on mõlemal juhul eriliiki isikuandmete (nt puude olemasolu) ja poliitikauuringute tegemiseks vajalik eelnev Andmekaitse Inspektsiooni (AKI) luba. Seega on isikuandmete teaduseesmärgil töötlemiseks vaja läbida eraldi menetlus, mis võib mitme erineva andmeallika korral võtta oluliselt kauem aega kui Statistikaameti puhul, kes koondab erinevaid isikuandmeid eri allikatest ning omab suurt kogemust isikuandmete kaitsel ja turvalisel töötlemisel.

Üldkorra kohaselt võib isikuandmeid töödelda teadus- ja ajaloouuringu ning riikliku statistika vajadusteks ilma andmesubjekti nõusolekuta, kui isikuandmed on pseudonüümitud või samaväärset andmekaitse taset võimaldaval kujul (IKS § 6 lg 1). Selle sätte kohaselt võib isikuandmeid teaduseesmärgil analüüsida mh ka privaatsuskaitse tehnoloogiatega (nt Sharemind), kui seeläbi tagatakse isikuandmete töötlemine

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

pseudonüümitud või samaväärset andmekaitse taset võimaldaval kujul. Samas, Statistikaameti konfidentsiaalsete andmete teisene kasutus on hetkel Eestis kehtiva õiguse järgi piiratud kahel suunal:

1. teaduseesmärgil kasutamise piirangud - Statistikaameti konfidentsiaalsete andmete teaduslikel eesmärkidel edastamise korraga (edaspidi "Kord") detailsem õiguslik analüüs on esitatud aruande lisas X). Korra kohaselt võib kõrge tuvastamise riski ja/või kõrge sensitiivsuse korral konfidentsiaalsusnõukogu loal konfidentsiaalseid andmeid anda kasutada ainult lokaalsel töökohal (Korra p 5.4), sh VPN tunneli kaudu kaugkasutusena (Korra p 6.2), ning selle tulemitele teostab Statistikaamet aimatavuse kontrolli (Korra p 5.5). Samas on taotlejatel võimalus kasutada lokaalsel töökohal oma lähtekoode (Korra juurde lisatud konfidentsiaalsete andmete taotluse näidise punktid 4), samuti on võimalik taotletavate andmetega linkida teistest allikatest saadud andmeid (Korra juurde lisatud konfidentsiaalsete andmete taotluse näidise punkt 7). Seejuures pole täpsustatud, kelle poolt (kas Statistikaameti, taotleja või kolmandate isikute poolt) ja kus (kas lokaalsel töökohal või sellest väljaspool) tuleb linkimist teostada. Järelikult Korraga ei ole täielikult välistatud privaatsuskaitse tehnoloogiate rakendamine konfidentsiaalsete andmete analüüsil teadusuuringu eesmärgil ning tulevikus on võimalik Korda täpsustada nii, et see oleks selgelt ja üheselt lubatud, vajadusel konkreetsete tingimuste alusel.
2. muud piirangud - konfidentsiaalseid andmeid ei ole lubatud kasutada ühelgi muul kui statistilisel ja teaduslikul eesmärgil (riikliku statistika seaduse § 34 lg 1^1). See hõlmab mh halduse, õigusemõistmise, riikliku järelevalve, kriminaaluurimise jms eesmärgil kasutamise välistatust.

Kolmandas kasutusloos uurisime, kas masinõppe abil on võimalik parandada eelmistes kasutuslugudes loodud mudeleid, kui uue mõõtmena lisada ka inimeste demograafilised andmed (sh puude olemasolu). Selleks analüüsisime PÄISi ja Statistikaameti andmeid. Kuigi demograafilised andmed on põhimõtteliselt kättesaadavad ka muudest allikatest (nt EHIS jms), esitasime selle uuringu raames taotluse Statistikaameti konfidentsiaalsete andmete kasutamiseks, sest Statistikaametis on erinevate allikate andmed kokku koondatud ja neile on võimalik kiiremini juurde pääseda.

Statistikaamet palus enne konfidentsiaalsete andmete teaduseesmärgil väljastamist taotleda AKI seisukohta. Seetõttu pöördusime AKI poole palvega selgitada järgmisi küsimusi:

1. kas puude olemasolu andmed on eriliiki isikuandmed (eeldab AKI eelkontrolli IKS § 6 lg 4 alusel) - AKI vastas, et tegemist on eriliigiliste isikuandmetega kaudselt tuvastatava inimese kohta, kuid nende töötlemiseks ei ole vaja AKI luba, sest isikustatult töötleb neid andmeid üksnes Statistikaamet.
2. kas antud uuringu puhul on tegemist täidesaatva riigivõimu analüüsi või uuringuga, mis tehakse poliitika kujundamise eesmärgil (eeldab AKI eelkontrolli IKS § 6 lg 5 alusel) - AKI vastas, et tegemist ei ole poliitikauuringuga, vaid uuringu väljund on praktilist laadi abi riigiasutusele tema igapäevatöös, järelikult AKI luba pole vaja.
3. kas antud uuringu jaoks võib kasutada privaatsuskaitse tehnoloogiaid (sh turvalist ühisarvutust, päringute piiranguid jne), tagamaks et andmeid ei töödelda andmesubjekti tuvastamist võimaldaval kujul (lubatud IKS § 6 lg 3 alusel ainult piiratud juhtudel, mis eeldavad ülekaaluka avaliku huvi olemasolu) - AKI kinnitas, et tema varasemas praktikas on tõepoolest leitud, et Sharemindi privaatsuskaitse tehnoloogiaplatvormi kasutamisel pole AKI luba vajalik, kuna sellisel juhul ei töödelda andmeid isiku tuvastamist võimaldaval kujul. Samas möönis AKI, et selles

seisukohas on teatav vaidlusmoment olemas, sest Statistikaameti enda andmebaasist toimub andmete väljavõtmine isikustatult ning see toiming on hõlmatud Sharemindi abil teadusuuringu eesmärgil isikuandmete töötlemise õigusliku alusega. Niisiis ei saa öelda, et Sharemind'iga andmeid isikustatult ei töödelda.

Lisaks rõhutas AKI, et:

1. Päästeamet ei saa selle uuringu raames väljatöötatud masinõppe mudeleid tulevikus rakendada Statistikaameti andmete peal, sest see väljub teadusuuringu raamidest - tegemist oleks Päästeameti haldustegevusega ning Statistikaameti konfidentsiaalsete andmete kasutamine niisugusel eesmärgil on seaduse järgi keelatud.
2. Päästeamet ei saa selle uuringu raames väljatöötatud masinõppe mudeleid tulevikus rakendada ka nt EHISe andmete peal, sest AKI on selgelt veendunud, et see poleks kooskõlas võimalikult väheste andmete kogumise põhimõttega (isikuandmete kaitse üldmääruse (IKÜM) art 5 lg 1 p (c)) ega andmete algse kogumise eesmärgiga (IKÜM art 6 lg 4). Ka seisaks AKI vastu seadusemuudatusele, mis sellist andmevahetust võimaldama hakkaks.

Niisiis kinnitas AKI, et Statistikaameti konfidentsiaalseid andmeid (sh puude olemasolu) võib antud uuringus koos PÄISi andmetega analüüsida teaduseesmärgil, kuid mitte hiljem Päästeameti haldustegevuse raames (nt ennetustegevuseks). Ka ei näinud AKI põhimõttelisi takistusi privaatsuskaitse tehnoloogiate (nt Sharemind) kasutamisel Statistikaameti andmete analüüsimiseks, kuid tõi välja, et nende andmete esimene töötlemistoiming Statistikaametis toimub paratamatult isikustatult.

AKI seisukoha põhjal rääkisime Statistikaametiga läbi nende demograafiliste andmete kasutamise tingimused. Statistikaamet ei nõustunud andmete töötlemiseks kasutama Sharemind'i, sest see ei ole kooskõlas Statistikaameti andmete teaduslikul eesmärgil töötlemise korraga. Niisiis sõlmisime Päästeametiga konfidentsiaalsete andmete teaduslikel eesmärkidel kasutamise lepingu, mille kohaselt Statistikaamet teeb konfidentsiaalsed andmed selle uuringu teostajatele kättesaadavaks Statistikaameti teadlaste keskkonnas.

Statistikaameti andmed sisaldasid infot inimeste soo, vanuse, elukoha, haridustaseme, tööhõive kohta. Tulekahjudega sai inimesi siduda aadressi identifikaatori kaudu, mis tähendab, et tulekahju seoti üldiselt majaga. Seetõttu seadsime endale kaks peamist masinõppe ülesannet:

1. Ennustada, kas inimene elab majas, mis sattus põlengusse.
2. Ennustada ruutkilomeetri rahvastikuandmeid kasutades, kas sellel alal toimus põleng.

Mõlema ülesande juures rakendasime esmalt XGBoosti algoritmi, kuid selle tõlgendamine osutus keeruliseks. Seda eriti selle tõttu, et kui uurida, milliseid tunnuseid treenitud mudel oluliseks pidas, ei andnud algoritm stabiilset vastust ning sõltus olulisel määral treening- ja testandmete jaotusest. See andis märku, et andmed on tugevalt korreleeritud, mistõttu järgmised proovitud algoritmid põhinesid Lasso reguleeritud meetoditel. Nimelt aitavad sellised meetodid lisaks regressioonikordajate leidmisele kaasa ka kõige olulisemate (kõige rohkem infot pakkuvate) tunnuste valikul. See tähendab, et tugevalt korreleeritud tunnuste puhul valib Lasso tihti välja ainult ühe, kõige olulisema. Ülejäänud kordajad seatakse mudelis võrdseks nulliga.

Kui esimese ülesande puhul andis Lasso regressioon küllalt kesiseid tulemusi (ta eristas ainult seda, et vene rahvusest inimestel on suurem tõenäosus olla põlenud majas ja eesti rahvusest inimestel väiksem),

siis teine ülesanne osutus natuke huvitavamaks. Lasso regressioon oli täpsuselt võrreldav (kohati isegi parem) kui XGBoost, ning osutas küllalt stabiilselt paarile üksikule olulisele tunnusele. Nendest joonistusid selgelt välja töötuse määr, haridustase ja töökoha tüüp.

Kuna absoluutarvude abil on raske öelda, kui palju infot sai mudel vastavatest andmetest rahvaarvu kohta ruutkilomeetril, siis proovisime sama analüüsi läbi teha ka suhtarvudega (kõik sissekanded jagati ruutkilomeetri populatsiooniga). Pärast seda analüüsi joonistus ka XGBoosti algoritmiga treenitud mudeli abil välja olulisema tunnusena just töötuse määr ruutkilomeetril.

4.2.6. Turvaline ühisarvutus

Algselt oli plaanis Statistikaameti andmeid analüüsida turvalise ühisarvutuse platvormi Sharemind[8] abil. Turvaline ühisarvutus on metoodika, mille abil on võimalik arvutusi teha ilma üksikisiku andmeid nägemata. Sharemindi platvormi jaoks on olemas statistikapakett, mis sisaldab paljusid kasutatavamaid meetodeid. Kuna masinõppe meetodeid Sharemindi jaoks veel loodud ei olnud, implementeerisime privaatsust säilitavad (kõrvalkanali rünnete vastu kaitstud) versioonid järgnevatest algoritmidest: juhumets / otsustuspuud, tugivektormasinad, logistiline regressioon, k-keskmiste klasterdamine, lahutus omaväärtusteks ja -vektoriteks, SVD, kernel peakomponentide analüüs, Gaussi protsess, Kalmani ja LiuWesti filter (aegridade jaoks), Lasso logistiline regressioon ja XGBoost.

4.2.7. Kasutusviiside ettepanekud

Projekti käigus loodud prototüüp kasutab csv formaadis ette antud andmestikke. Uuringu ajal majutati rakendus Cybernetica serveris ja tehti Päästeametile kättesaadavaks üle veebi piirates ligipääsu IP põhiselt. Päriskasutuses soovitame programmi integreerida Päästeinfosüsteemi ja selle andmeladudega. Nii on võimalik kiiresti mudeleid uuendada andmelao uuenemisel ning on suurem tõenäosus, et kaardirakendust kasutatakse. Selline uuendamine on väga oluline protsessi osa ennustustäpsuse tõstmisel. Selle käigus märgitakse ära tulekahjujuhtumid, mida mudel õigesti ei ennustanud, ning mis tegelikult juhtus ennustatud intsidentide puhul. See tähendab ka Ehitisregistri teatud andmete lisamist Päästeameti vastavale andmelaole ning nende tabelite hoidmist ajakohasena (piisab perioodilisest uuendusest).

Prototüüp kasutab veebirakenduste raamistikku Python-Flask, masinõppe algoritmi XGBoost, kasutajateabe ja raportite hoidmiseks SQlite'i ning treeningandmestikke csv formaadis. Kõiki neid on võimalik asendada sobivate tehnoloogiatega. Tehnilisi takistusi sellise prototüübi tootestamiseks ja Päästeameti töös kasutuselevõtuks ei ole.

Kuigi meie uuring leidis, et ilmaandmed pigem segavad mudelite loomist kui annavad neile juurde, on võimalik, et teiste meetodite kasutamisel ilma lisamiseks hoonetele, kus tulekahju ei toimunud, võib paremini töötada.

Kokkuvõttes võib öelda, et mudel, mis rakendab ainult Päästeameti enda andmeid ja avaandmeid annab empiiriliselt hea tulemuse, on selgelt teenusena ülesehitatav ning sellisel kujul ilma oluliste takistusteta kasutusele võetav. Samas sai ka selgeks, et potentsiaalselt olulist lisandväärtust andvate sotsio-demograafiliste andmete abil teenuse ehitamine on oluliselt takistatud just andmete ligipääsu ja teisest kasutusviisi kujutava ristkasutuse keelu tõttu praeguste regulatsioonide keskkonnas.

[8] https://sharemind.cyber.ee

4.3. e-Tervis

4.3.1. Probleemipüstitus

Eestis terviseandmekogud jagunevad keskseteks andmekogudeks ja konkreetse tervishoiuteenuse osutaja infosüsteemideks. Eesti on eriti uhke kesksete riiklike haiguslugude ja terviseandmete üle. Käesolev projekt puudutab keskseid riiklikult ülalpeetavaid andmekogusid, täpsemalt uuriti kolme suuremat operatiivtöös kasutatavat andmekogu:

1. retseptikeskust (väljakirjutatud ravimid);
2. Eesti Haigekassa andmekogu (raviarved) ja;
3. Tervise Infosüsteemi (terviselugude kokkuvõtted, laborianalüüside tulemused).

Kuigi kõigis neist kolmes talletatakse Eesti kõikide patsientide ravi kohta olulist infot ja nende n.ö katvus (täielikkus) on üksikpatsiendi ravi vaatest küllaltki hea, pole analüütilises plaanis veel ära kasutatud nende andmete täielikku potentsiaali. Andmekogude andmete omavaheline kokkuviimine (nn linkimine) ja töötlemine, mis patsiendi tervisest tervikpildi saamiseks on hädavajalik, kuid paraku tehniliselt ning õiguslikult keeruline. See teeb raskeks ja kulukaks nii Eesti-siseste terviseuuringute läbiviimise kui ka rahvusvahelistes terviseuuringutes osalemise. Esiteks sisaldavad need andmekogud segu struktureeritud, pool-struktureeritud ja täielikult struktureerimata andmetest, mis ei ole ilma ulatusliku eeltöötluseta kvantitatiivselt analüüsitav ega ka efektiivselt rakendatav näiteks kliiniliste otsustustugedes. Teiseks on eri riikide ja andmestike vaheliste semantiliste erisuste tõttu sageli raske valideerida rahvusvaheliste uuringute tulemusi või soovituste rakendamist Eesti populatsioonis.

Käesoleva alamprojekti eesmärk on linkida ja harmoneerida kolme ülalnimetatud andmekogu andmed ning näidata, kuidas seda rakendada erinevate küsimuste/probleemide lahendamiseks kasutades selleks rahvusvaheliselt parimaid arvutusmudeleid ja teadmust, masinõpet, jne.

4.3.2. Kasutatud andmed ja teisendused

Kasutatud andmestik sisaldas 10% juhuvalimit kõigist Eesti isikukoodiga isikutest ja kõiki nendega seotud terviseandmeid kolmest riiklikust terviseandmekogust (retseptikeskus, Eesti Haigekassa andmekogu, Tervise Infosüsteem). Terviseandmed esinesid perioodist 2012–2019. Kokku sisaldas andmestik terviseandmeid 150 811 patsiendi kohta vanuses 0-106, kokku 20,7 miljonit erinevat tervisedokumenti.

Haigekassa ja TEHIK poolt edastatud pseudonüümitud[9] andmed ühendati üheks andmestikuks ja teisendati ühtsele rahvusvaheliselt tunnustatud OHDSI OMOP CDM andmekujule.[10]

Analüüsitavate terviseandmete teisendamiseks OMOP andmekujule kasutati:

1) ETL (*Extract-Transform-Load*) tööriista Rabbit In a Hat;[11]

[9] Pseudonüümi moodustas TEHIK, kes vahendas linkimiseks vajalikud isikukoodide teisenduse info Haigekassale, kes sama pseudonüümi kasutades edastas oma andmekogu andmed. Teadlasteni ei jõudnud selle protsessiga ühtegi isikukoodi ega nime.

[10] OMOP andmemudel on kirjeldatud siin: https://ohdsi.github.io/CommonDataModel/cdm53.html

[11] http://ohdsi.github.io/WhiteRabbit/RabbitInAHat.html

2) Teisendustööriista Usagi;[12]
3) OMOP sõnastikke.[13]

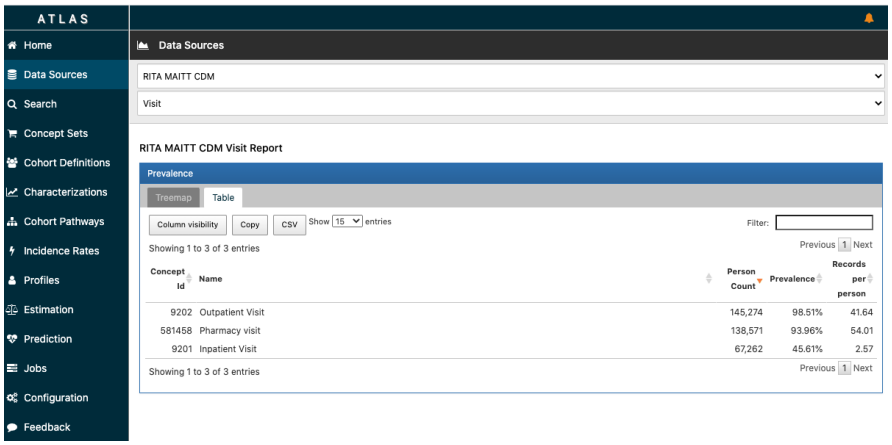
Lisaks olemasolevatele sõnastikele (ICD10 -> SNOMED; ATC -> RxNorm; LOINC), loodi projekti käigus suuremaid ja väiksemaid täiendavaid sõnastikke Eesti andmetele. Suurematest sõnastikest loodi vastavustabelid kirurgiliste protseduuride klassifitseerimisel kasutatavatele NCSP koodidele (NCSP -> SNOMED) ja Haigekassa hinnakirja teenusekoodidele (HK koodid -> SNOMED).

Lisaks sõnastike loomisele eraldati, struktureeriti ja puhastati vabatekstiväljadel esinevat infot (eelkõige Digiloo dokumentidest, aga ka Retseptikeskuse andmetest). Samuti puhastati ja korrastati struktuursel kujul esitatud infot.

Struktureeritud ja puhastatud andmete jaoks loodi automaatsed korduvkasutatavad teisendusskriptid, mis algandmed ühtsele standardsele OMOP andmekujule viiksid. Kogu teisendamisprotsess püüti üles ehitada selliselt, et see oleks võimalikult universaalne ja töötaks edaspidi teistel sarnastel andmestikel. Samuti peeti oluliseks, et loodud teisendusskripte oleks võimalikult lihtne vajadusel täiendada. Näiteks saab nii moodustada võrreldava andmekohordi COVID-19 patsientide andmetest, kui läbida sarnased andmeväljastused ja eeltöötlused.

OMOP andmekujule viidud andmetele rakendati kvaliteedikontrolle, et tuvastada võimalikud ebakõlad andmetes ja/või andmete ülekandmises. Selleks kasutati olemasolevat kvaliteedikontrolli tööriista DataQualityDashboard [14] ja IMI2 EHDEN projekti poolt arendatavat tööriista CdmInspection.[15]

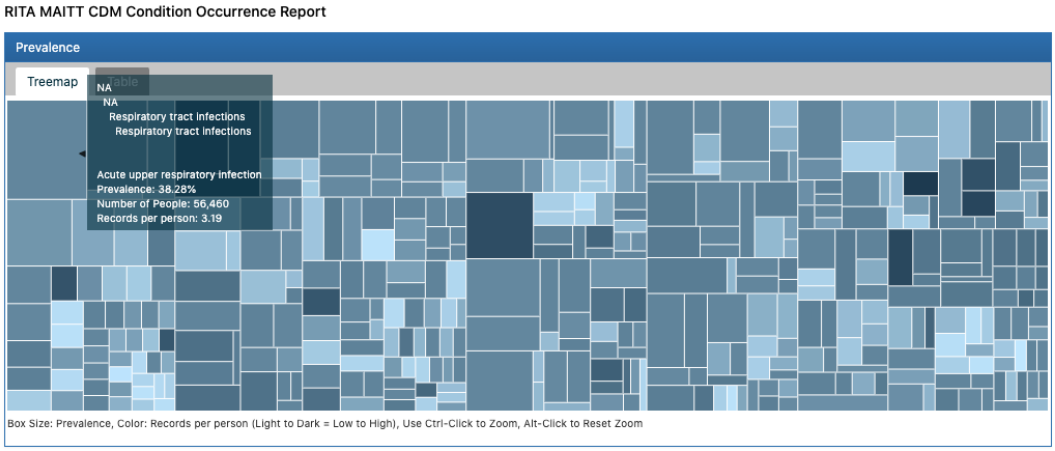
Andmeid hoiti ja analüüsiti Tartu Ülikooli teadusarvutuskeskuses selleks spetsiaalselt loodud turvalises virtuaalmasinas. Analüüside läbiviimiseks ja seal hulgas erinevate päringu-kohortide moodustamiseks kasutati veebipõhist Atlas serveri tarkvara, mis võimaldab saada andmetest standardiseeritud ülevaateid ning viia läbi epidemioloogilisi kohortide võrdlusuuringuid.



[12] https://github.com/OHDSI/Usagi
[13] https://athena.ohdsi.org/search-terms/start
[14] https://ohdsi.github.io/DataQualityDashboard/
[15] https://github.com/EHDEN/CdmInspection

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

Joonis 4. Näide Atlas serveri ekraanivaatest. Pildilt on näha, et OMOP kujul andmestikus oli 67 tuhat statsionaarse ravi patsienti (46% kõigist patsientidest), 145 tuhat ambulatoorse ravi patsienti (99%) ja 138 tuhat retsepti välja ostnud patsienti (94%).



Joonis 5. Näide diagnooside jaotusest uuritavates andmetes. Kõige levinum diagnoos on hingamisteede haigused, mida on esinenud 38% patsientidest. Keskmiselt on iga patsiendi kohta 3 sellist kirjet. Kõige rohkem kirjeid patsiendi kohta (kõige tumedam) on aga glaukoomi patsientidel (37).

4.3.3. Rakenduste loomise protsessi kirjeldus

Terviseandmete uurimiseks on alati vajalik eetikakomitee luba. Selleks tuleb esitada eetikakomiteele uuringu taotlus, mis seejärel kindlate protseduuride alusel läbi vaadatakse ja positiivse otsuse korral ka heakskiidu saab. Tegemist on üldjuhul mitmete kuude pikkuse protsessiga sõltuvalt komiteede koosolekute toimumise intervallist. Eestis on kolm inimuuringutega tegelevat eetikakomiteed, kuid neil on erinevad otsustuspiirid ja käesoleva uuringu jaoks oli vajalik saada heakskiit kahelt erinevalt eetikakomiteelt:

- Tartu Ülikooli inimuuringute eetika komitee - reguleerib andmete kasutust retseptikeskusest ja Eesti Haigekassa andmekogust - taotlus esitati 09.01.2020, kooskõlastus saadi 04.02.2020 (luba 300/T-23), andmed väljastati Haigekassa poolt käesoleva projekti uuringumeeskonnale 13.04.2020.
- Eesti bioeetika ja inimuuringute nõukogu (EBIN) - reguleerib andmete väljastust Tervise Infosüsteemist - taotlus esitati 28.02.2020, kooskõlastus saadi 14.04.2020 (luba 1.1-12/653)

EBIN luba ei olnud andmete uurimismeeskonnale üleandmiseks veel piisav, kuna TEHIK kui Tervise Infosüsteemi volitatud töötleja ütles, et neil on vaja ka Sotsiaalministeeriumi kui vastutava töötleja sellekohast korraldust/luba. Seetõttu pöördusime loa saamiseks Sotsiaalministeeriumi poole 21.04.2020 ja saime loa 29.04.2020 (kiri 1.5-20/1388-2). Sellele järgnes andmete kokkupanek ja uurimismeeskonnale väljastamine, kuid seegi ei laabunud probleemideta, sest esimese korraga said andmed pseudonüümitud vale võtmega. Õige võtmega pseudonüümitud andmed laekusid uurimismeekonnale 29.06.2020.

Projekti jooksul toimus uurimismeeskonnas muudatusi, mis kõik tulid omakorda kooskõlastada läbi eetikakomiteedele esitatud täiendustaotluste ning vastava menetlusprotsessi. Uute tudengite lisandumisel (magistritööd, bakalaureusetööd, doktorandid) on see suhteliselt ajamahukas ja keeruline protsess kõikidele osapooltele, sealhulgas üle koormatud EBIN-ile endale.

Kokkuvõttes saab sellist andmete taotlemise protsessi lugeda võrdlemisi ebamugavaks ja ajakulukaks. Ei saa pidada ratsionaalseks olukorda, kus uuringu läbiviimiseks on vaja mitme erineva eetikakomitee luba (mõlemal erinev taotluse vorm) ning seejärel veel ka ministeeriumi otsust.

Pärast andmete laekumist alustati andmete teisendamist OMOP CDM kujule, mis suures osas sai lõpule viidud aprilliks 2021. a. Väiksemas mahus teisendustööd jätkusid kuni projekti lõpuni. Teisenduste käigus tuli kasutada ja luua ka mitmesuguseid teisendustabeleid. Näiteks tekkis töö käigus vastavustabel NOMESCO kirurgiliste protseduuride teisendamiseks OMOP kujule. Nimetatud vastavustabel vaadati üle ka Tervise ja Heaolu Infosüsteemide Keskuse poolt (TEHIK) ja nad plaanivad selle uue publitseerimiskeskkonna valmides ka avaldada.

Esimeste masinõppemudelite loomist ja katsetamist alustati augustis 2020. a, mis jätkus kuni projekti lõpuni. Projekti käigus katsetati erinevaid masinõppe mudeleid erinevatele tervishoiuküsimustele vastamiseks. Nendest on täpsemalt juttu all. Tulemusi raporteeriti kvartaalselt projekti juhtrühmas, osade mudelite kohta kirjutati ning kaitsti Tartu Ülikooli arvutiteaduse instituudis lõputööd.

Peamise probleemkohana võib välja tuua asjaolu, et võrdlemisi keeruline on leida avalikust sektorist ühte selget osapoolt, kes konkreetset tervishoiualast küsimust püstitaks või selle n.ö kiirest rakendamisest huvitatud oleks. Kõik, mis puudutab konkreetse patsiendi ravi ja selles masinõppe või otsustustugede reaalset rakendamist, on õiguslikult tugevalt reguleeritud valdkond ning avalik sektor võib eelistada teadus-arenduspõhiste katsetuste (innovatiivsete lahenduste piloteerimise) asemel, kus tuleks võtta lahenduse eest ise vastutus, pigem n.ö valmislahendusi koos välise osapoole tootjavastutusega. Kogu tarkvara mis on kasutuses ühe inimese raviotsuste mõjutamisel kuulub uue regulatsiooni järgi meditsiiniseadmete regulatsioonide (MDR; IVDR) alla. Seetõttu on innovatsiooni lihtsam rakendada pigem retrospektiivselt - näiteks rahvatervise valdkonnas, kus tegeldakse haiguste ennetusega populatsiooni, mitte üksikpatsiendi tasemel. Samuti on kõrge huvi andmete uurimiseks teaduslikul tasandil, s.h rahvusvaheliselt. Sellegipoolest on kõigil neil juhtudel - nii konkreetse patsiendi raviks masinõppeliste otsustustugede abil, populatsioonipõhiste uuringute kui rahvusvahelise teadustöö puhul - hädavajalik terviseandmete omavaheline ühendamine erinevatest andmekogudest, tagades ühtlasi nende kõrge kvaliteet ja ühetaolisus. Kuna see vajadus puudutab korraga erinevaid osapooli (erinevad andmekogude haldajad, teadlased teisest asutusest jne), on keeruline leida üht selget osapoolt, kes tajuks probleemi nii teravana, et mitte üksnes andmekvaliteedi olulise tõstmise, vaid ka tervishoiuprotsesside probleemide masinõppelise lahendamise enda vedada võtaks.

4.3.4.Mudeldatud väljundid ja nende põhjendused

Andmete standardiseerimise järel püstitati järgnevad spetsiifilisemad alamküsimused, millele vastamist masinõppe abil testida:

1. Kuivõrd on võimalik automatiseerida ravimite täiendavate riskivähendamise meetmete rakendamise analüüsi

2. Kas ja kuidas on võimalik analüüsida refraktiivsete nägemishäirete ja nendega koosesinevaid oftalmoloogilisi haigusi Eesti Haigekassa raviarvete alusel
3. Kas on võimalik ennustada emakakaelavähi kujunemist ennetuse eesmärgil
4. Kas on võimalik ennustada hospitaliseerimisi krooniliste haigustega patsientide korral (ennetus ja planeerimine)
5. Kas on võimalik ennustada kardiovaskulaarhaiguste kujunemist ennetuse eesmärgil
6. Rahvusvaheline võrdlus - osalesime rahvusvahelises PIONEER eesnäärmevähi ravi uuringus.
7. Katsetasime uut haigustrajektooride uurimiseks mõeldud innovaatilist tarkvarapaketti
8. Hindasime privaatsust säilitava *Sharemind* platvormi sobivust eelnimetatud hospitaliseerimise ennustamise mudeli rakendamise näitel

Kõikide uuritud küsimuste kohta on täpsem kirjeldus toodud järgmises osas.

Ravimite täiendavate riskivähendamise meetmete rakendamise analüüsi automatiseerimine

Mitmete ravimite kasutamisel on arstil kohustus rakendada täiendavaid riskivähendamise meetmeid. Samal ajal on üksikute näidete põhjal teada, et nende meetmete rakendamine on siiski küllaltki harv. Selliste uuringute süstemaatiline läbiviimine keerukas, mille tõttu olukorrast tervikülevaade Eestis puudub. Samas võiks olukord olla märksa parem, kui arstile oleks võimalik anda õigel ajal soovitus konkreetsele patsiendile vastavaid meetmeid rakendada.

Raili Jäe magistritöös (Jäe 2021) kasutati kolme riikliku terviseandmekogu (retseptikeskus, Eesti Haigekassa andmekogu, Tervise Infosüsteem) 2012.-2019. aasta andmete alusel koostatud ja OMOP CDM mudeli kujule viidud ühendandmekogu ja analüüsiti, kuidas rakendatakse tervishoiutöötajate poolt ravimite täiendavaid riskivähendamise meetmeid. Töös täiendati vastavat analüüsi metoodikat ning realiseeriti analüüs agomelatiini näitel taaskäivitatava R Markdown analüüsidokumendina. Loodud analüüsikoodi saab edaspidi rakendada teiste sarnaste täiendavates riskivähendamise meetmetes toodud nõuete rakendamise analüüsimiseks. Tulemused näitasid, et toimeaine agomelatiin täiendavates riskivähendamise meetmes toodud nõudeid täidetakse pigem harva. Kõik täiendavates riskivähendamise meetmetes ettenähtud maksafunktsiooni mõõtmise analüüsid tehti kauem kui 180 päeva kestnud ravi korral vaid 2% kohorti kuulunud patsientidele. Vähemalt üks maksaensüümide mõõtmise analüüs oli enne ravi või ravi ajal tehtud 56%-le kõikidest agomelatiini patsientidest. Töös tuuakse välja viis kuidas niisugust analüüsi saaks teostada automaatselt, nii et arstile kuvatakse õigetel hetkedel vastavaid hoiatusi, kuid selleks on tarvilik tõsta terviseandmete kvaliteeti, m.h muuta andmete kogumise printsiipe. Näiteks tuleks iga laboriuuring raviarvetel kajastada individuaalse koodiga (mitte ühe ja sama koodiga nagu praegu). Samuti on tarvis tõsta Tervise Infosüsteemi esitatavate laborimõõtmiste andmete kvaliteeti (tagada korrektne kuupäev, mõõtetulemus ja ühik).

Refraktiivsete nägemishäirete ja nendega koosesinevate oftalmoloogiliste haiguste analüüs Eesti Haigekassa raviarvete alusel

Eestis ei ole teadaolevalt teostatud laiapõhjalisi uuringuid, mis kaardistaksid refraktiivsete nägemishäirete ja oftalmoloogiliste haiguste esinemisstatistikat ning nende omavahelisi seoseid.

Irina Häelm'i magistritöö (Häelm 2021) eesmärk oli anda ülevaade vaatlusandmete põhjal teostatud epidemioloogilise analüüsi teostamise metoodikast ning kirjeldada selle rakendamist ja täpsust Eesti Haigekassa andmestiku põhjal, uurides oftalmoloogiliste haiguste ja refraktiivsete nägemishäirete esinemisstatistikat ning koosesinemise tõenäosust.

Euroopa Liit
Euroopa
Regionaalarengu Fond
Eesti
tuleviku heaks
RITA

Töö tulemustest selgus, et töös kasutatavad OHDSI (Observational Health Data Sciences and Informatics) Population-Level Estimation analüüsitööriistad võimaldavad leida kirjanduses kirjeldatud seoseid Eesti vaatlusandmetest.

Sarnaselt kirjanduses kajastatule, oli uuritava andmestiku põhjal hüperoopiaga patsientidel tõenäolisem amblüoopia, strabismi ja kinnise nurga glaukoomi diagnooside esinemine. Müoopiaga oli aga tõenäolisem avatud nurga glaukoomi, katarakti ning võrkkesta rebendite ja irrete diagnooside esinemine. Töös teostatud andmeanalüüs ei tuvastanud statistiliselt olulisi seoseid astigmatismi ja oftalmoloogiliste haiguste vahel. Lisaks toodi töös välja uuritavate diagnooside esinemisstatistika Eestis ning võrreldi seda kirjandusega.

Käesoleva uuringu kontekstis on kõige olulisemad järeldused sellest tööst, et Eesti Haigekassa raviarvete andmed on selliseks analüüsiks piisavalt kvaliteetsed ja neid on ka piisaval hulgal, kuid veelgi tähtsam - kui need andmed on viidud OMOP CDM kujule, on sellist analüüsi võrdlemisi lihtne läbi viia, sest saab kasutada olemasolevaid OMOP/OHDSI analüüsitööriistu ehk kogu analüüs on automatiseeritav ja tööriistana kasutusele võetav. Seega on see hea näide kuidas standardiseerimise järel on võimalik mujal välja töötatud algoritmid või seosed Eesti andmetel valideerida ja vajadusel tööriistadeks arendada.

Emakakaelavähi ennustusmudel

Hetkel on emakakaelavähi ennetamise olulisim meetod sõeluuring, kuna HPV-vastase vaktsineerimise mõju ei ole jõudnud veel avalduda. Sõeluuringuid viiakse läbi HPV testina (eelnevatel aastatel oli esmastestiks PAP test). Sellised sõeluuringud on kulukad ja nendes osalemise määr madal. Seetõttu soovisime koostöös epidemioloogide ja naistearstidega luua masinõppe ennustusmudeli, mille eesmärgiks on terviseandmete põhjal tuvastada naised, kellel on suurem risk emakakaelavähki haigestuda. Selline ennustusmudel võimaldaks pöörata kõrgema haigestumisriskiga naistele suuremat tähelepanu ja neid sagedamini uuringutele kutsuda.

Uuringu tegi keeruliseks see, et andmestikus on vähe naisi, kellel on diagnoositud selles ajaperioodis emakakaelavähk. Seega esimese sammuna üritasime uurida neid kohorte, kellel on PAP testi raames leitud histopatoloogilised leiud, mis võivad viia emakakaelavähi tekkeni.

Lisaks pakkus uuritav andmestik võimaluse saada esmane ülevaade inimese papilloomiviiruse (hrHPV) genotüübilisest levimusest Eestis – seda seni tehtud ei ole. PAP testi tulemuste ja HPV tüüpide kõrvutamine võimaldab ennustada HPV vaktsiinide mõju emakakaelavähi ennetamisele. Projekti käigus uurisime, kuivõrd emakakaela intraepiteliaalne neoplaasia ja emakakaelavähk on seotud vaktsiinvälditavate HPV tüüpidega. Vastav analüüs on hetkel veel täiendamisel, tulemuste osas oleme ette valmistamas teadusartiklit.

Seni tehtu kohta on võimalik ülevaade saada järgmistel lehtedel:

- kohortide ülevaade:
  http://omop-apps.cloud.ut.ee/ShinyApps/CCSCharacterization/
- täpsema ülevaade kohortidest koos kitsendustega ja erinevate kohortide võrdlus:
  http://omop-apps.cloud.ut.ee/ShinyApps/CCSStudy/
- rakendus mudelitele:
  http://omop-apps.cloud.ut.ee/ShinyApps/CCSPrediction/

Emakakaelavähi ennustusmudel on hea näide sellest, kuidas OMOP CDM kujule viidud andmeid ja olemasolevaid analüüsivahendeid kasutades on võimalik väga efektiivselt tegeleda ka konkreetse eriala spetsiifilise tervishoiukorralduse küsimuste uurimisega. Efektiivsus tuleneb siin kahest asjaolust - esiteks ei kuluta uurimisgrupp asjatut ressurssi andmete Eesti-spetsiifilisuse tõttu kohortide, visuaalide või tabelite koostamisele (näiteks kiire visuaalne ülevaade tänu OHDSi tööriistadele võimaldab näha, kui palju on igas kohordis patsiente ning milliste tunnuste tõttu ta sellesse kuulub, mida mingi seisundiga patsientidele veel lisaks tehtud on jne); teiseks, selline analüüs ei oleks olnud võimalik ilma andmekogude omavahelise ühendamiseta.

Kõik ülalnimetatud lehtedelt avanevad vaated, samuti järgnevate uurimisküsimuste juurest viidatud vaated, on koostatud automaatselt OHDSI/OMOP tööriistu kasutades.

Hospitaliseerimise ennustamine krooniliste haigustega patsientide korral

Kroonilise haiguste korral võiks tihedam/parem esmatasandil toimuv kontroll vähendada hospitaliseerimise vajadust ning sellest tingitud kulu tervishoiusüsteemile. Probleemiks on nende patsientide väljaselgitamine, kes esmatasandil paremat/tihedamat jälgimist vajaksid ning kel õnnestuks selle läbi ka hospitaliseerimist vältida. Haigekassa on varasemalt läbi viinud pilootprojekti, kus patsientide valikuks kasutati otsustuspuud (World Bank Group 2017). Hiljem on Maailmapanga projekti raames üritanud hospitaliseerimist masinõppe algoritmide abil ka ennustada ning leida faktoreid, mis on hospitaliseerimise ennustamisel määravad. Tegemist oli mahuka projektiga ning lõppes selge tulemita. Samas ei ole probleem kuhugi kadunud ning käesolevas projektis püüti taasluua ennustusmudelit selliste kõrge riskiga patsientide tuvastamiseks.

Uuringu valimisse kuulusid nii 1. tüüpi kui ka 2. tüüpi diabeeti ja hüpertooniatõve põdevad inimesed ehk teatud hulk kroonilisi haigeid, aga välistatud olid patsiendid, kellel vähidiagnoos on juba varasemalt olemas. Kokku oli selliseid inimesi andmestikus umbes 43 000, kes valiti lähtekohordiks. Kõik neid kroonilisi haiguseid põdevad inimesed koondati ühte kohorti. Täiendasime ka hospitaliseerimiste seotud kohorti, täpsustades, kuidas ja millised hospitaliseerimised välja jäetakse.

Analüüsi läbiviimisel kasutati erinevaid mudeleid: Lasso Logistic Regression, Gradient Boosting Machine ja Random Forest. Mudel kasutas oma ennustamisel laia spektrit andmeid: diagnoose, ravimeid, protseduure, mõõtmisi koos mõõtmistulemustega, visiite ja demograafilist infot (sugu ja vanusegrupp), millest koostati tunnused. Tunnusteks oli näiteks diagnoosikood, visiitide arv arstide juurde, välja ostetud retseptiravimite arv jne. Võrreldes eelmise perioodiga lisandusi mõõtmiste tulemused, mis võiksid olla olulised näitajad, aga antud mudelis need oluliseks ei osutunud.

Krooniliste haigete hospitaliseerimise ennustusmudeli kurvi alune pindala (AUC) on 0,73, mis näitab, et arendatud mudelil on juba teatav ennustusjõud olemas. Kõige olulisemateks tunnusteks osutusid isiku kohta esitatud tervisedokumentide arv, mis viitab arstikülastuse sagedusele, ja apteegist väljaostetud retseptide hulk.[16] Samuti südamehaiguste esinemine, mis võib olla seotud sellega, et antud kohordis on kaasatud inimesed, kellel on diagnoositud hüpertooniatõbi.

Hospitaliseerimise mudeli ennustusvõimekus ei ole väga kõrge, samas on käesoleva uuringu vaatest oluliseks järelduseks fakt, et sellistele laialdast huvi pakkuvatele küsimustele vastamine, millele

[16] http://omop-apps.cloud.ut.ee/ShinyApps/WBHospitalizationPrediction/

varasemalt on kulunud märkimisväärselt ressurssi, on OMOP CDM kujule viidud andmete puhul äärmiselt lihtne. Siin toodud küsimuste vastamiseks kasutati täielikult juba olemasolevaid OMOP/OHDSI tööriistu.

Kardiovaskulaarhaiguste ennustusmudel

Käesoleva uurimisküsimuse eesmärgiks oli uurida koostöös rahvatervise arstidega Eesti tervise andmete sobivust kardiovaskulaarhaiguste ennustamiseks. Täpsemalt soovisime välja töötada masinõppel põhineva riskimudeli, kasutades OMOP/OHDSI konsortsiumi poolt loodud vabavara tarkvara tööriista ATLAS. Kardiovaskulaarhaigused on maailmas enim surmasid põhjustavate haiguste grupp ning riskimudel, mis aitaks tuvastada patsiendid, kellel on kõrge risk infarkti või insuldi saamiseks, aitaks säästa paljusid elusid kui ka aega, mida arstid kulutavad hetkel kõikide, ka olematu riskiga inimeste uurimiseks sõeluuringusse (skriiningule) kutsumisega. Meie andmestikus on olemas haiguste diagnoosid, ravimid ja mõõtmised, mida saab kasutada mudeli parameetritena. Varasemalt välja töötatud riskimudelites (QRISK3, SCORE) on samuti kasutatud mudeli sisenditena kliinilisi näitajaid lisaks tavapärastele parameetritele nagu sugu, vanus ja kehamassiindeks.

Sihtrühmaks valiti inimesed (target cohort), kellel pole varasemalt diagnoositud ühtegi RHK-10 diagnoosikoodide G45, I20-I25, I63-I64 alla kuuluvaid haigusi ning samuti ei tohtinud neile olla välja kirjutatud statiine (ravimid, mida kasutatakse liigkolesterooliveresuse raviks). Uurisime, kuidas vanuseline koosseis mõjutab mudeli tulemusi ning seetõttu on ühes mudelis kaasatud inimesed vanuses 25 kuni 84 aastat (sellist vanuselist koosseisu kasutati ka QRISK3 riskimudeli loomisel) ning 40 kuni 74 aastat (selline vanuseline koosseis määrati kuna enamasti on kardiovaskulaarhaigustest mõjutatud pigem vanem populatsioon).

Tulemuskohort koosnes inimestest, kellel jälgimisaja järgselt diagnoositakse üks või mitu kardiovaskulaarhaiguste diagnoosidest (G45, I20-I25, I63-I64).

ATLAS tööriist võimaldab täpsustada huvipakkuvaid parameetrid, mis kardiovaskulaarhaiguste puhul võivad suurema tõenäosusega mõjutada mudeli headust (nö “kitsas” mudel). Sellised parameetrid võivad olla kaasuvad haigused nagu esimest või teist tüüpi diabeet, kõrgvererõhutõbi, reumatoidartriit, südame arütmia jne, ravimite nagu kortikosteroidide tarvitamine ning vererõhu või kolesterooli mõõtmised. Samuti on huvitav uurida mudelit, mis kasutab kõiki andmebaasis olevaid andmeid (nö “lai” mudel) kuna selliselt võib leida mudelile sisendeid, mida algselt poleks kaalunud ning kasutatav mudel (Lasso Logistic Regression) valib andmetest ise olulisimad parameetrid, mis mudeli headust enim mõjutavad.

Tulemuste tõlgendamiseks on erinevaid meetrikaid, siinkohal kasutasime AUC (Area Under the Curve) ja AUPRC (Area Under the Precision Recall Curve). AUC (vahemikus 0 kuni 1, baseline 0.5) näitab, kui hästi suudab mudel eristada huvipakkuva diagnoosiga inimesi tervetest inimestest. AUPRC soovitatakse kasutada juhtudel, kui huvipakkuva diagnoosiga inimesi on võrreldes tervete inimestega tunduvalt vähem, meie mudeli puhul on tulemuskohordi suurus vahemikus 4% kuni 5% (ühtlasi tähendab see seda, et AUPRC baseline tulemus mudeli puhul on 0.04 kuni 0.05).

Parima tulemuse (AUC 0.83 ja AUPRC 0.17) andis “lai” mudel, mille sihtrühma kuulusid inimesed vanuses 25 kuni 84 aastat.[17] Samas sihtrühmas andis “kitsas” mudel tulemuseks AUC 0.82 ja AUPRC 0.15.[18] See mudel töötas kõige paremini kuna hõlmas kõige suuremat sihtrühma (üle 83 000 inimese). Vanuselises

[17] http://omop-apps.cloud.ut.ee/ShinyApps/CVDRiskModelLai/
[18] http://omop-apps.cloud.ut.ee/ShinyApps/CVDRiskModelKitsas/

mõttes konservatiivsemas sihtrühmas oli inimesi 50,403 (“lai” mudel AUC 0.74 ja AUPRC 0.14, “kitsas” mudel AUC 0.72 ja AUPRC 0.11).

Ennustusmudeli tulemused olid lootustandvad, kuid probleemiks oli see, et sellise uuringu põhjal järelduste tegemiseks oleks vaja pikemast ajavahemikust andmeid. Praegusel juhul on kasutatud kolmeaastast perioodi, mille jooksul vaadatakse kardiovaskulaarhaiguste teket, aga tavaliselt kasutatakse vähemalt kümneaastast perioodi. Kolmeaastane piirang tuleneb sellest, et andmed tuleb jagada mitmeks osaks, et tekiks: 1) vaatlusperiood, mille jooksul kogutakse andmeid, mille põhjal ennustatakse kardiovaskulaarhaiguste teket ning 2) periood, mille jooksul vaadatakse kardiovaskulaarhaiguste teket.

Samuti ei olnud mudelisse kaasatud mitmeid andmeid, mis on eelnevalt osutunud kardiovaskulaarhaiguste ennustusmudelite korral oluliseks (näiteks sissetulek, haridus jms), aga mida terviseandmetes ei kajastata. Lisaks on ka suitsetamine ja alkoholi tarbimine olulised riskitegurid, mida ei ole hästi tervisedokumentides dokumenteeritud.

Arvestades asjaolu, et Eesti terviseandmekogudesse lisandub andmeid pidevalt juurde, muutub ka andmestik niisuguste uuringute läbiviimiseks aasta-aastalt järjest paremaks. Kaaluda tuleks haiguste oluliste riskitegurite kvaliteetsemat kogumist. Teatud haiguste puhul võib paremate ennustuste tegemiseks vaja olla ka teisi andmeid peale terviseandmete.

PIONEER Study-a-thon

Eesnäärme vähk on enam levinumaid vähke meeste hulgas. Eesnäärme vähi ravimisel kasutatakse erinevaid raviviise alates aktiivsest jälgimisest (oota ja vaata) kuni aktiivse ravimiseni. Aktiivse jälgimise käigus jälgitakse haiguse progressi, aga aktiivset ravi ei toimu. Samas on selle ravimeetodi kohta vähe informatsiooni, kuidas haigus progresseerub ja kuidas komorbiidsused mõjutavad oodatavat eluiga.

IMI2 projekt PIONEER (*prostate cancer diagnosis and treatment enhancement through the bid data in Europe*[19]) on koos IMI2 EHDEN projekti (European Health Data & Evidence Network[20]) ja OHDSI kogukonnaga[21] praegu läbi viimas suuremahulist uuringut eesnäärme vähi aktiivse jälgimise ravimeetodi uurimiseks. Selle läbiviimiseks korraldati esmalt 5 päevane virtuaalne study-a-thon (8.-12. märts 2021), mille käigus defineeriti välja uuringu kohordid ja tarkvara esmane seadistus tulemuste analüüsiks. Hetkel käib aktiivne tulemuste analüüs ning ka Eestis on kaasatud MAITT projekti käigus loodud andmebaas.

Uuringusse on kaasatud mitmeid andmebaase üle maailma, näiteks Hollandi vähikeskuse andmebaas, USA suured raviarvete ja haiglate andmebaasid. Eesti andmebaasi unikaalsus seisnes selles, et meie andmebaasis olid ühendatud Haigekassa raviarved, retseptide andmed ja samuti epikriiside info, mille tulemusena omasime patsiente praktiliselt kõikide uuritavate kohortide kohta. Samas võis miinuseks pidada asjaolu, et kuna andmebaasis oli umbes 150 000 inimest, siis sobilikke eesnäärme vähi juhtumeid oli kokkuvõttes vähe, eriti võrreldes suurte USA andmebaasidega.

[19] https://www.imi.europa.eu/projects-results/project-factsheets/pioneer
[20] https://ehden.eu/
[21] https://www.ehden.eu/uncovering-the-natural-history-of-prostate-cancer-in-data-from-millions-of-patient-lives-across-the-globe/

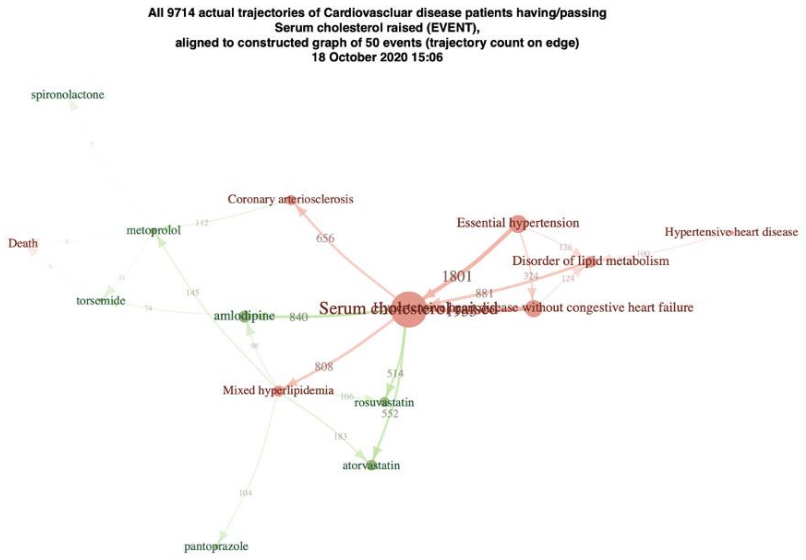
Esialgsed tulemused näitasid,[22] et aktiivse jälgimise ja kohese ravimise vahel ei olnud suuri erinevusi. Samas 6 kuud hiljem on haiglasse sattumine tõenäosus kõrgem neil meestel kelle rakendati aktiivset jälgimist. Esimesed tulemused esitatakse EAU (European Association of Urology) kongressil märts 2022 uuringu juhtide poolt ning plaanis on selle uuringu eri tahkudest mitu rahvusvahelist teadusartiklit.

Saadud kogemus näitab ilmekalt, et Eesti patsientide hulk on liialt väike iseseisva taolise uuringu läbiviimiseks. Seega saame nendes küsimustes, näiteks eesnäärme vähi ennustusmudeli loomises, toetuda eelkõige rahvusvahelisele kogemusele ning suurematele andmekogudele. Samas on Eesti andmed täiesti piisavad selliste rahvusvaheliste andmete peal treenitud mudelite valideerimiseks ja tänu andmete ühetaolisusele (OMOP mudelile) on mudelite rakendatavust võimalik Eesti andmetel väga kergesti kontrollida. See annab hea võimaluse kiirelt veenduda, kas välismaal leitud seosed kehtivad ka Eestis ja kas väljapakutud lahendusi saaks ravijuhendites rakendada. Lisaks sellele andis OMOP CDM kujule viidud andmestik suurepärase võimaluse Eesti teadlastel selles suures rahvusvahelises uuringus osaleda. Kuna taolisi uuringuid on lisandumas teisigi, on see erakordne võimalus sellises suures teaduses kaasa rääkida.

4.3.6. Haigustrajektooride uurimise tarkvarapaketi katsetamine

Viimasel aastakümnel on järjest rohkem huvi hakanud pakkuma terviseandmetest haigustrajektooride automaatne tuvastamine. See võimaldab mitte üksnes tuvastada haiguste vahel uusi seoseid, vaid ka otsida ravitrajektooride kitsaskohti, et neid siis tervishoiukorralduslike meetmetega parandada. EHDEN projekti raames on Tartu Ülikooli ja Hollandi teadlaste koostöös välja arendatud trajektooride tuvastamise tarkvarapakett, mida käesoleva uuringu andmestikul oli võimalus katsetada. Tegemist on unikaalse selletaolise tarkvaraga, mis võimaldab otsida trajektoore mistahes OMOP CDM andmestikul mitmesuguste sündmuste - haiguste, protseduuride, ravimite jne - vahel.

[22] https://data.ohdsi.org/PioneerWatchfulWaiting/ ja
https://data.ohdsi.org/PioneerWatchfulWaitingDiag/



Joonis 6. Näide tarkvara poolt automaatselt tuvastatud haigustrajektooridest.

Katsetus kulges väga edukalt ja andis kindluse, et tarkvara toimib probleemideta ning lähiajal on plaanis avaldada ka sellekohane rahvusvaheline teadusartikkel. Joonis 6. näitab üht sellist tuvastatud haigustrajektoori. Jällegi ei oleks sellist katsetust ei oleks saanud läbi viia ilma Eesti OMOP CDM kujul andmeteta. Tarkvara ennast kirjeldav artikkel ilmus 16.03.2022 ajakirjas JAMIA Open.

4.3.7. Privaatsust säilitavate tehnoloogiate rakendamine masinõppemudelite loomisel

Turvaline ühisarvutus on metoodika, mille abil on võimalik arvutusi teha ilma üksikisiku andmeid nägemata. Käesolev projekt demonstreeris selle metoodika võimekust masinõppemudelite treenimisel terviseandmete peal. Selleks kasutati turvalise ühisarvutuse platvormi Sharemind,[23] mis võimaldab vajalike arvutuste läbiviimist ilma, et analüütikud üksikuid kirjeid näeksid. Platvormi jaoks on loodud ka SecreC programmeerimiskeel, mis lihtsustab algoritmide implementeerimist.

Sharemindi rakendamiseks on vaja kolme serverit. Nende serverite vahel jagatakse andmed ja tehakse analüüse privaatsust säilitavalt viisil. Selles projektis olid andmete paiknemisel ja käsitlemisel ranged piirangud, seetõttu seati vastavad serverid üles pilve, kus terviseandmed asusid. Selleks paigaldati pilve kolm uut virtuaalmasinat, mis täitsid arvutusosapoolte rolle.

Andmete viidi Sharemindi tabelite kaupa. Sealjuures sorteeriti välja tulbad, mis olid tühjad või ei andnud lisainformatsiooni. Selline filtreerimisprotsess ei mõjuta jõudlustulemusi ning sai tehtud puhtalt mugavuse ja kettaruumi kaalutlusel.

Enne masinõppe meetodite rakendamist tuli andmed viia vastavaks analüüsiks vajalikule kujule. Seda võimaldavad erinevad juba varasemalt ETL (extract-transform-load) protsessiks implementeeritud meetodid. Kuna ülesande peamine eesmärk oli demonstreerida ja uurida just masinõppemudelite

[23] https://sharemind.cyber.ee

treenimist, siis rakendati ajapiirangute tõttu ETL osas ka algoritme, mis võivad teatud juhtudel andmete kohta vähesel määral infot lekitada (lekib see, kui palju on erinevatel doonoritel erinevaid haiguseid, aga sealjuures ei leki, kes need doonorid on või mis haigused neil olid).

Pärast andmete sobilikule kujule viimist, rakendati tuntud masinõppe meetodeid. Peamiseks tööriistaks oli projekti käigus SecreC keeles implementeeritud LASSO logistiline regressioon, mis näitas nende andmete peal head täpsust. Lisaks on Sharemind platvormil selle meetodi abil mudelite treenimine oluliselt kiirem kui näiteks otsustuspuudel põhinevaid meetodeid kasutades.

4.3.8. Kasutusviiside ettepanekud

Tööpaketi eesmärkide täitmiseks oli esimene ülesanne koondada erinevad terviseandmed ja viia need OMOP CMD kujule ning teiseks siht katsetada vähemalt kolme erinevat terviseväljundi prognoosimise masinõppe mudelit tekkinud andmestiku peal. Selle katsetuse eesmärk oli näha, kas ja kuivõrd toimib standardiseeritud andmestik platvormina, millel on juba väiksema täiendava lisatööga võimalik kiirelt valideerida erinevaid algoritme terviseväljundite prognoosimiseks ja teatud haiguste kujunemise seletamiseks ning kujunemisteede identifitseerimiseks. Tulemused näitasid, et see just nii toimibki:

1. osade väljundite puhul ilmnes, et rahvusvaheliselt välja töötatud algoritmid ja tööriistad annavad valiidseid tulemusi, mis on otseselt kasutatavad nii rahvatervise kui individuaalses ennetustöös kui ka indiviidide ravi toetamises;
2. osade väljundite puhul ilmnes, et Eesti andmestikud ei ole piisavad, et saaks empiirilise täpsuse, mis lubaks ennustusmudeleid reaalselt rakendusse võtta;
3. standardiseeritud andmestikel saab juba suhteliselt kergemini kasutada juba välja töötatud tööriistu analüütilise ja kasutatava teadmise loomiseks.

Andmestike väljavõtmine, linkimine, ühtlustamine ning standardkujule viimine ja kogu selle protsessi korrataval kujul elluviimine annab tulevikus võimaluse hoida kokku ca 60-70% ajast, mis iga uurimisküsimuse juures ettevalmistava tööna on alati vaja ära teha veel enne kui mudeleid saab treenida, sobitada ja hinnata, kas need empiiriliselt üldse töötavad.

See aja kokkuhoid vähendab olulisel määral ka masinõppe/AI projektidesse investeerimise riski. Kogu ettevalmistav töö tuleb nagunii alati ära teha enne, kui saab selgeks, kas loodud masinõppe mudel üldse töötab või valideerida, et see Eestis siiski ei tööta. MAITT projekti abiga panustasime masinõppe/AI rakenduste testimisse e-tervise valdkonnas läbi OMOP CMD kujul andmetele rakendamise.

Projekti tulemused on kasutatavad vähemalt neljal olulisel viisil:

1. OMOP CMD andmete standardiseerimise protokoll, skriptid ja loodud OMOP andmestik kujutavad endast testplatvormi, millel saavad ülikoolid, Haigekassa, TEHIK, Sotsiaalministeerium, Tervise Arengu Instituut jne kiirelt valideerida rahvusvaheliste uuringute tulemusi ja erinevate tervisenäitajate masinõppe ennustusmudelite täpsust Eesti populatsiooni andmestikul;
2. loodud ennustusmudelid on otseselt kasutatavad rahvatervise ennetustöö täpsemaks sihitamiseks gruppidesse, kus mudelid on tuvastanud suurimad haiguse kujunemise riskid, kaugema eesmärgiga riskide realiseerumist ära hoida;
3. loodud ennustusmudelid on kasutatavad ka indiviidi tasandil ennetustegevuses läbi isikule vastavate riskiskooride arvutamise ja nende alusel soovituste andmise, kuid selleks on vaja esmalt

lahendada ülal mainitud meditsiiniseadme tootmise/tootjavastutuse küsimus mis pidurdab hetkel osapoolte valmisolekut selliseid otsustugesid kasutusele võtta;

4. ravimite kasutuse täiendavate riskivähendamise meetmete analüüs näitas, kuidas on andmekvaliteedi parandamise järel võimalik andmepõhiseid otsustustugesid kasutada ka juba otseselt individuaalse ravikvaliteedi hindamiseks ja soovituste andmiseks.

Hetkel on Euroopas käimas IMI2 EHDEN projekt, mille üheks partneriks on Tartu Ülikool, ning mille eesmärk toetada tegevusi, et teisendada OMOP CDM kujule üle 200 miljoni Euroopa patsiendi andmed. Projekti raames toetatakse nii teisendamist ise kui ka väike-ettevõtete koolitamist ja sertifitseerimist selleks tööks. Eestis on tänaseks sertifitseeritud kaks ettevõtet - OÜ STACC ja OÜ Quretec. Käesolevas projektis tegelesid teisendamisega nii STACC kui Tartu Ülikooli töötajad. Eesti andmekogudel on jätkuvalt võimalus saada rahalist toetust andmekogude viimiseks OMOP kujule, sest Ida-Euroopa andmekogusid on teisendatud vähe.[24]

4.4. Küberjulgeolek

4.4.1. Probleemipüstitus

Lühikokkuvõttena on selle tööpaketi ülesandeks hõlbustada küberkaitse spetsialistide tööd, kes tegelevad võrguliikluse monitoorimisega. Hõlbustamine tähendab siin nii nende töömahu vähenemist kui töö kvaliteedi parandamist. Tööpaketis katsetatav lahendus on ohtude äratundmise automaatõpe: süsteem õpib, millised on potentsiaalsed ohud ja millised mitte, ning suudab seeläbi eelfiltreerida/eelmärgendada leitud hoiatusi. Anname järgnevas veidi detailsema ülevaate probleemist.

Küberkaitse on IT süsteemide kvaliteetseks funktsioneerimiseks hädavajalik töölõik igas IT süsteemidest sõltuvas organisatsioonis. Küberkaitse peab tagama nii seda, et süsteemide tööd ei õnnestuks teadlikult häirida (kättesaadavus), et andmeid ja dokumente ei saaks pahatahtlikult muuta või kustutada (terviklikkus) kui seda, et süsteemid ei lekiks välistele osapooltele andmeid ja dokumente (konfidentsiaalsus).

Küberkaitse spetsialistide üheks oluliseks töölõiguks on süsteemide töö monitoorimine, leidmaks võimalikke ründeid ja muud pahatahtlikku või hooletut käitumist, mis liigitub küberkaitselise probleemi alla. Süsteemide monitoorimise peamine vahend on nende genereeritud logifailide uurimine.

Üks spetsiifiline monitoorimisviis on organisatsiooni ja välismaailma vahelise interneti-liikluse detailne jälgimine, eesmärgiga leida ründeid ja muid küberkaitselisi probleeme. Tüüpiline viis selle jaoks on kasutada võrguliiklusest taoliste probleemide automaatse otsimise tarkvarasüsteeme. Tuntuim ja enimkasutatud taoline süsteem maailmas on vabaraline Suricata.[25]

Suricata kasutab probleemsete sündmuste leidmiseks suurt hulka spetsiaalseid reegleid, mis on koostatud küberkaitse spetsialistide poolt. Taolisi reeglikomplekte pidevalt uuendatakse ning levitatakse, nii tasuta kui tasulistena.

Suures organisatsioonis – näiteks Tallinna Tehnikaülikool – genereerib Suricata pidevalt tohutu hulga hoiatusi võimalike ohtlike sündmuste kohta. Nende hulk ühes päevas on TTÜs suurusjärk kuni miljon. Rõhuv enamus hoiatusi ei viita mitte tegelikele probleemidele, millega tuleks tegeleda, vaid

[24] https://www.ehden.eu/data-partner-calls/
[25] https://suricata.io/

potentsiaalsetele ohukohtadele. Küberkaitse spetsialistidel on teatud kogemuse omandamise järel – eeskätt organisatsiooni süsteemide arhitektuuriga kursisolemine – reaalne võime hinnata, millised hoiatused nõuavad edasist kontrolli ja tegelemist, ja milliseid hoiatusi võib ignoreerida.

Arusaadavalt ei jõua ükski spetsialist igapäevaselt vaadata läbi miljonit hoiatust. Praktikas kasutatakse meetodit, kus hoiatused nende sageduse järgi sorteeritakse, ning uuritakse uusi/vähemsagedasi hoiatusi. Ka taoliste hoiatuste hulk võib olla väga kõrge, mis teeb spetsialisti töö nende läbivaatamisel väga mahukaks ja raskeks.

Antud tööpaketi tulemuseks on sobivate tehisintellekti-meetodite valik ja süsteemi väljatöötamine, mis võimaldab hoiatuste olulisust automaatselt õppida: nii spetsialisti kaasamata kui spetsialisti tegevust automaatselt õppides, tulemusena võimaldades oluliselt paremini ja automaatselt filtreerida hoiatusi tähtsateks ja vähemtähtsateks, vähendades seega spetsialisti töömahtu ja tõstes tema töö kvaliteeti.

4.4.2. Ülevaade kasutatud andmetest

Meie töölõigu peamisteks andmeteks on Tallinna tehnikaülikooli välisvõrku monitooriva Suricata süsteemi genereeritud hoiatused ehk logikirjed. Nagu varasemas öeldud, on nende suurusjärk päevas kuni miljon.

Täiendavalt on AS Cybernetica poolt kasutusel nende ettevõtte välisvõrku monitooriva Suricata süsteemi genereeritud hoiatused, mille maht on mitu suurusjärku väiksem.

Lisaks toor-hoiatustele kasutatakse täiendavat infot, näiteks IP aadresside ja geograafiliste asukohtade (a la linnad/riigid) seostamise andmebaasi: Suricata logikirjeid rikastatakse väliste osapoolte asukohtadega.

Kolmandana on kasutusel ka võrguliikluse sisu väljavõtted: neid kasutavad küberkaitse spetsialistid uurimaks olulisemate hoiatuste reaalseid põhjusi ja tegelikku ohtlikkust. Kuna need andmed on väga tundlikud - krüpteerimata ehk ilma näiteks https-i kasutamata võivad nad loetavalt sisaldadada isikuandmeid, paroole jne jne - siis neid andmeid loodavale süsteemile ega meetodite uurijatele tegelikult kättesaadavad ei ole. Neid kasutavad ainult hoiatusi uurivad spetsialistid, kellele on organisatsiooni poolt antud vastav õigus ja sõlmitud vastavad NDA-d.

Konkreetsemalt on töö käigus kasutatatud erinevatel perioodidel TTÜs 2021 aastal genereeritud logiaandmeid: tervet aastat andmete suure mahu tõttu kasutatud ei ole. Uuritud andmete perioodid/mahud on meie ülesande jaoks piisavad: juhendatud masinõppe jaoks kasutati spetsialisti poolt märgendatud ca 16.000 hoiatust, juhendamata masinõppe jaoks uuriti mitu suurusjärku suuremat kogust hoiatusi. Lisaks on AS Cybernetica poolt meetodite efektiivsuse võrdluseks kasutatud AS Cybernetica 2020 aastal kogutud analoogilisi andmeid väiksemas mahus.

4.4.3. Rakenduse loomise protsess

Rakenduse loomise protsessi **esimene etapp** oli andmetele ligipääsu taotlemine. Suricata logikirjed ja nendega seotud võrguliikluse väljavõtted olid ja on TTÜs reaalselt olemas, kuid nendele ligipääs on nende tundlikkuse tõttu väga piiratud. Ligipääsu võimaldamiseks analüüsis juhtiv TTÜ küberkaitse spetsialist (töögrupi liige), kel on neile andmetele ligipääs, nende väljade tundlikkuse astet ja kõrvaldas liig-tundlikud andmeväljad. Kokkulepitud kogus andmevälju ja ligipääsu taotlevad teised töögrupi liikmed ja vajadused vaadati üle TTÜ turvalisuse eest vastutavate isikute poolt, kes sõlmis töögrupi liikmetega ka vajalikud NDA-d.

Rakenduse loomise protsessi **teine etapp** oli masinõppeks sobiva arvuti hankimine ja riistvara- ning tarkvarasüsteemi ülesseadmine selliselt, et süsteem saaks ligipääsu lubatud andmetele turvalises keskkonnas, ning väljapoolt seda keskkonda oleks ligipääs piiratud ainult töögrupi liikmetele.

Rakenduse loomise protsessi **kolmas etapp** – mis toimus juba esimese ja teise etapiga paralllelselt – oli ülesande ja vajaduste detailsem analüüs koos analoogiliste senitehtud uuringutega tutvumisega ja võimalike lahendusdetailide arutamisega.

Neljas etapp oli andmete eelanalüüs masinõppe jaoks koos andmete sildistamist mittevajavate (juhendamata) masinõppe meetodite kasutamisega. Selle käigus loodi ka uus juhendamata masinõppe meetod SCAS, mis on kirjeldatud uues avaldatud artiklis ja mida lühidalt kirjeldame järgnevates peatükkides.

Viies etapp oli andmete sildistamine/märgendamine: töögruppi kuuluv küberkaitse spetsialist uuris läbi mitme eri perioodi hoiatused ja märgistas ohtlikud ja kriitilised hoiatused, et süsteem saaks seejärel neid märgistusi õppida automaatselt hoiatustele külge panema.

Kuues etapp oli masinõppe süsteemi loomine, mis võimaldaks katsetada erinevaid masinõppe meetodeid selliselt märgistatud andmete pealt õppimiseks.

Seitsmes etapp oli masinõppe reaalne tegemine, katsetades seejuures hulga erinevate meetoditega, nii juhendamata kui juhendatud meetodite hulgast, eesmärgiga parimad välja valida ning nende efektiivsust ja rakendatavust hinnata.

Kaheksas etapp oli edukamate meetodite täiendav katsetamine AS Cybernetica andmetel: väiksem kogus ja küllalt teistsugused Suricata logikirjed.

Üheksas etapp oli tulemuste analüüs: mis on paremad meetodid ja miks. Kaheksas etapp toimus iteratiivselt ja paralleelselt seitsmenda etapiga: analüüsi tulemusena muudeti parameetreid ja katsetati teiste meetoditega. Etapi lõpus asuti kirjutama artiklit juhendatud masinõppe uuringu tulemustest.

4.4.4. Testitud meetodid ja nende valiku põhjendused,

Projekti raames uuriti nii juhendamata (ei vaja spetsialisti poolset regulaarset märgendamist) kui juhendatud (süsteem õpib spetsialisti käitumist „järele tegema“) masinõpet IDS-i (intrusion detection system, konkreetselt suricata) Suricata küberkaitse-alarmide eelfiltreerimiseks.

Masinõppe tehnikate rakendamine on järginud standardset töövoogu, mis koosneb andmete eeltöötlusest, tunnuste valikust ning klassifikaatorite treenimisest ja valideerimisest.

Juhendatud masinõppe meetodid

Eeltöötluse etapis eemaldati valesti formateeritud vaatluspunktid ja andmekogum tasakaalustati. Selle tulemusel loodi ligikaudu 16 000 märgistatud kirjest koosnev andmekogum. Seejärel jagati andmed treenimise ja valideerimise alamhulkadeks, samas kui tunnuste valiku jaoks viidi läbi treenimisseti teine jaotamine. Andmekogum sisaldas kahte klassi – ohutuid hoiatusi ja hoiatusi, mida inimekspert peaks analüüsima.

Kirjandusest saadavaid tulemusi järgides analüüsiti kategooriliste- ja numbriliste tunnuste eristamisvõimet. Vaid viie tunnuse valimine entroopia väärtuste põhjal võimaldab treenida mudeleid, mille headuse parameetrid varieeruvad 0,96 ja 0,98 vahel. Sellele vaatamata on sellel klassikalisel lähenemisel üks oluline puudus. Nimelt nõuavad standardsed masinõppetehnikad kategooriliste tunnuste kodeerimist, mistõttu on sama mudeli kasutamine erinevate ajaperioodide jaoks keeruline. Seejärel pöörati esmane tähelepanu arvväärtustele. Hinnati standardsete masinõppe klassifikaatorite (nagu *k*-lähim naaber, otsustuspuu, juhumetsad, AdaBoost, tugivektormasinad ja logistiline regressioon) headust. Otsustuspuu ja AdaBoosti klassifikaatorid on näidanud parimat jõudlust. Seejärel treeniti ja valideeriti valitud mudeleid, kasutades andmestikke, mis koguti peaaegu kahekuulise intervalliga (treenimisandmed märtsis ja valideerimis andmed mais). Mudelite klassifitseerimisvõime ja sobivuse näitajad on toodud tabelis 4. Õigsus näitab õigesti prognoositud hoiatuste esinemise ja mitteesinemise osakaalu kogu hoiatustest. Täpsus näitab tegelikult esinenud hoiatuste osakaalu kõikidest prognoositud hoiatustest, saagis näitab õigesti prognoositud hoiatuste esinemise osakaalu kõikidest esinenud hoiatustest.

Tabel 4. Mudelite klassifitseerimisvõime ja sobivuse näitajad

Mudel	Õigsus	Täpsus	Saagis	F1 - skoor
Otsustuspuu	0.969	0.949	0.990	0.969
AdaBoost	0.966	0.942	0.992	0.966

Nimetatud juhendatud masinõppe algoritme koos valitud spetsiifiliste parameetrite ja põhimõtetega katsetati seejärel AS Cybernetica andmestikul. Viidi läbi kaks katsegruppi:

1. TTÜs treenitud mudeli katsetamine AS Cybernetica hoiatustel. Ootuspäraselt olid tulemused halvad, st mudeli ennustusvõime oli liiga madal selleks, et olla kasulik. See kinnitab, et õpitud klassifitseerimine sõltub väga tugevalt konkreetse organisatsiooni IT infrastruktuurist: hoiatus, mis ühe organisatsiooni jaoks on tõsine, ei pruugi seda olla teise organisatsiooni jaoks. Ehk teisisõnu, süsteem õpib muuhulgas organisatsiooni IT-struktuuri.
2. TTÜs väljatöötatud algoritmide/parameetrite katsetamine AS Cybernetica spetsialisti märgendatud hoiatustel, tehes nende peal läbi kogu masinõppe, ehk, ehitades mudeli AS Cybernetica jaoks. Tulemused olid väga head, kuigi natuke nõrgemad kui TTÜ andemetel treenituna: põhjus tõenäoliselt selles, et AS Cybernetica andmekogus oli oluliselt väiksem.

Juhendamata masinõppe meetodid

Lisaks juhendatud masinõppe (*supervised machine learning*) meetoditele uuriti projekti raames ka juhendamata masinõppe (*unsupervised machine learning*) meetodite sobivust IDSi alarmide analüüsiks. Juhendamata masinõppe meetoditel on juhendatud meetodite ees üks suur eelis -- nad ei vaja märgendatud treeningandmeid. Antud asjaolu on IDSi alarmide ja võrgumonitooringu valdkonnas väga oluline, kuna selles valdkonnas koguneb lühikese aja jooksul väga suur hulk andmepunkte ja piisavalt realistlike treeningkorpuste loomine on seetõttu äärmiselt ajamahukas ülesanne. Kuna aja möödudes

kerkivad esile uued ohud, muutuvad ka treeningkorpused mõne aja jooksul vananenuks, mistõttu nende loomise kallist ja töömahukat protseduuri tuleb küllaltki tihti korrata.

Projekti raames loodi juhendamata andmevoogude klasterdamise algoritm SCAS (*Stream Clustering Algorithm for Suricata*), mis kasutab tööks alarme tekitanud pakettide sisu põhjal loodud atribuute (näiteks HTTP paketi "user-agent" ja "url" väli). Suricata ja teised moodsad IDS platvormid toetavad selliste atribuutide tuvastamist ning loodud algoritm kasutab kokku 25 sellist tekstilise iseloomuga atribuuti. Algoritm tuvastab IDSi alarmivooge reaalajas analüüsides sageli esinevad alarmikombinatsioonid, mis vastavad tihti esinevatele rutiinsetele ja madala prioriteediga rünnetele, luues iga sellise alarmikombinatsiooni põhjal klastritsentroidi. Järjekordse alarmikombinatsiooni reaalajas ilmumisel mõõdetakse selle sarnasust kombinatsioonile vastava tsentroidiga või tsentroidi puudumisel loetakse alarmikombinatsioon anomaalseks.

Antud lähenemine võimaldab SOCi (küberkaitse keskuse) operaatoril fokuseerida oma tähelepanu anomaalsetele alarmikombinatsioonidele (nende puhul tsentroid puudub) või siis nendele kombinatsioonidele, mis ei ole neile vastavatele tsentroididele piisavalt sarnased. Kuna kõik IDSi alarme kirjeldavad atribuudid on tekstilised, tugineb tsentroidi ja alarmikombinatsiooni sarnasuse mõõtmise funktsioon atribuutide väärtuste sagedustabelitele. SCAS algoritmist on olemas vabavaraline realisatsioon,[26] mida on TTÜ SOCis kasutatud enam kui aasta vältel, ning algoritmi täpsem kirjeldus koos 3 kuu jooksul kogutud detailse jõudlusstatistikaga on avaldatud konverentsiartklis (Vaarandi 2021).

Pikema aja jooksul kogutud jõudlusstatistika näitab, et algoritm suudab eristada väheolulisi rutiinseid ründeid rohkem tähelepanu vajavatest alarmidest ning suudab alarme reaalajas analüüsida vähese protsessori- ning mäluressursi kuluga.

4.4.5. Tulemuste kirjeldus ja rakenduse lühiülevaade

Projekti ühe tulemusena leidis kinnitust hüpotees, et nii juhendamata kui juhendatud masinõpe annavad küberkaitse spetsialistidele olulist abi organisatsiooni välisvõrgu liikluse monitoorimisel: aitavad hoida kokku tööaega (tänu eelfiltri kasutamisele) ja tõsta töö kvaliteeti (tehakse vähem vigu). Sarnaseid uuringuid on tehtud mitme uurimisgrupi poolt maailmas ka varem, kuid me ei tea publikatsioone, kus neid uuringuid oleks tehtud suure organisatsiooni tegeliku monitoorimise käigus, ehk praktilises situatsioonis, ning võrreldud tulemusi teise organisatsiooni andmetel.

Samuti andis projekt selge kinnituse, millised juhendatud masinõppe meetodid on tegelikus praktikas sobivamad: juhendatud meetodite osas otsustuspuu ja Adaboost.

Projekti ühe tulemusena töötati välja ja realiseeriti uus juhendamata masinõppe algoritm antud ülesande jaoks: SCAS ehk *Stream Clustering Algorithm for Suricata*. Analoogiliselt eelmisele punktile leiti, et sel viisil tehtud juhendamata masinõpe annab olulise tööaja kokkuhoiu.

Projekti üks tarkvaraline rakendus – juhendamata masinõppeks - on kättesaadav eelviidatud SCAS repositooriumist. Juhendatud masinõppe tarkvara koosneb mitmest komponendist, mis teevad eeltöötlust, koguvad märgendatud andmeid, teevad reaalset masinõpet ning kasutavad loodud mudelit

[26] https://github.com/ristov/scas

hoiatuste filtreerimiseks. Nimetatud tarkvarakomplekt ei ole hetkel avalikult publitseeritud ning selle praktiline kasutamine teistes organisatsioonides nõuab lisatööd, mida kirjeldame all.

Projekti teadusliku tulemusena on publitseeritud eelmainitud artikkel juhendamata masinõppe meetodist SCAS ja hetkel kirjutab töögrupp artiklit juhendatud masinõppe tulemustest.

4.4.6. Kasutusviiside ettepanekud

IDSi (näiteks Suricata) kasutamine praktikas nõuab küberkaitse spetsialistidelt suurt hulka regulaarset tööd, mis tihtipeale muudab IDS-i kasutamise ebarealistlikuks. Meie projekt näitab, et kombinatsioonis juhendamata ja juhendatud masinõppe meetoditega see töömaht väheneb oluliselt, mis võiks julgustada organisatsioone Suricatat tegelikult kasutusele võtma.

Suricata tegelikul kasutamisel soovitame rakendada mõlemat masinõppe meetodit. Juhendamata masinõppe jaoks loodud algoritm on kättesaadav repositooriumist[27] koos kirjeldusega artiklis Vaarandi (2021). Juhendatud meetodi puhul on vaja sisse seada protseduur, kus hoiatusi läbi vaatav spetsialist olulisteks peetud hoiatused märgendab ning need märgendid süsteemile andmetena sisse annab, samuti tuleb süsteemi regulaarselt üle treenida. Samuti on vaja lahendada küsimus, kuidas neid filtreeritud hoiatusi spetsialistile paremini kuvada.

Taoline andmevoogude juhtimine ja kasutaja jaoks mugava liidese ehitamine ei ole väga keeruline, kuid mitte ka triviaalne ülesanne, ning sõltub kuigivõrd küberkaitse infrastruktuurist organisatsioonis. Meie soovitus on käivitada projekt, mis analüüsiks selle ülesande jaoks riigiasutuste küberkaitse infrastruktuuri ning looks selle põhjal võimalikult laialt kasutatava ettevalmistatud andmevoogude juhtimise, masinõppe käivitamise ja spetsialisti kasutajaliidese. Taolise süsteemi integreerimine organisatsioonide küberkaitse infrastruktuuri oleks oluliselt kergem ja kokkuhoidlikum, kui selle realiseerimine mitmes organisatsioonis eraldi ja sõltumatult.

5. Mõju avalikule sektorile

RITA projekti üheks eesmärgiks oli tuvastada masinõppepõhiste lahenduste väljatöötamise ja kasutuselevõtuga seotud mõjud avalikus sektoris. Selleks loodi esmalt kirjanduse põhjal ülevaade AI-põhiste lahenduste võimalikest mõjudest, millest allpool tuuakse ära lühikokkuvõte. Seejärel kaardistati projekti lahenduste väljatöötamisega seotud osapoolte kogemused võimalike mõjude ilmnemise osas. Kuivõrd projekti üksi lahendus ei ole tänaseks jõudnud rakendamise faasi, ei ole hetkel võimalik anda ka põhjalikku mõjuhinnangut. Seetõttu on käesoleva raporti empiiriline mõjuhinnang väga esialgne, pigem tulevikku vaatav ning tuues välja mõned peamised tendentsid.

5.1. Kuidas mõista võimalikke mõjusid?

AI-põhised lahendused võivad avalduda nii töötaja, organisatsiooni kui ka laiemalt ühiskonna tasandil (vt Tabel 1 Lisas 3). Esiteks, indiviidi tasandil võivad AI-põhised lahendused töötajaid võimestada läbi parema protsesside juhtimise. (Veale & Brass 2019). Läbi kasutajagruppide detailsema segmenteerimise ning sellest tuleneva personaliseerituma ja sihipärasema sekkumisega, on töötajatel võimalik oma aega ja ressursse paremini jaotada (Veale & Brass 2019). Lisanduvalt on võimalik AI-põhistele lahendustele

[27] https://github.com/ristov/scas

delegeerida rutiinsemad tegevused, jättes loomingulisemad ja keerulisemad juhtumid inimeste lahendada (Mehr et al. 2017). See võimaldab protsesse teostada väiksema ajakuluga ning suunata inimressurssi suuremat väärtust loovatesse tegevustesse (Sun & Medaglia 2019). AI-põhiste lahenduste kasutusega on võimalik parandada töötajate analüütilist võimekust ja oskuseid, mille läbi suurendada tegevuste mõjusust ja lõppkasutajate rahulolu.

Teiseks, organisatsiooni tasandil võivad AI-põhised lahendused pakkuda kiiremat ja täpsemat alternatiivi ühiskondlike probleemide tuvastamiseks (Pencheva et al. 2018), võimaldades ennetada erinevate probleemide süvenemist (Höchtl et al. 2016). Erinevate andmestike kooskasutus AI-põhistel lahendustel võib aidata kaasa uute teadmiste sünteesile (Kankanhalli et al. 2019). Lisanduvalt on võimalik seeläbi luua proaktiivseid automatiseeritud protsesse, suurendades organisatsioonide tõhusust, hoides kokku kulusid ning vähendades ühtlasi ajakulu nii tarbijale kui ka teenuseosutajale (Kuziemski & Misuraca 2020). AI-põhinevate tööriistade kasutus võib pakkuda rohkem võimalusi *ex-ante* poliitikaanalüüside teostamiseks, võimaldades etteulatuvalt paremini mõista poliitikaplaanide mõjusid (Pencheva et al. 2018) või kodanike ootuseid (Desouza & Jacob 2017). AI-põhinevate lahenduste läbi on võimalik tuvastada uusi seoseid läbi erinevate suurandmete andmekaeve tehnikate, mis võib viia poliitikaprobleemi ümbermõtestamiseni kui ka uute lahenduste sõnastamiseni (Mehr 2017). AI-põhiseid lahendusi võidakse kasutada ka organisatsioonis paremaks ressursside jaotamiseks (Pencheva et al. 2018). Siinkohal on võimalik hinnata nii tööprotsesside teostamist, ressursside kasutamist, rutiinsete töölõikude ja teiste ebaefektiivsete protsesside eemaldamist. Ka võimalike pettuste tuvastamine kas lõpptarbijate (Eggers et al. 2017) või teenuste osutajate poolt (Lima & Delen 2020). Lisanduv kasu võib samuti ilmneda näiteks AI-põhiste lahenduste rakendamisel riigihangete teostamisel. Näiteks lepingute koostamine, lisanduva informatsiooni pakkumine, hangetega seonduva otsusetegemine parima väärtuse tuvastamine, pakkujatega seonduvate riskide tuvastamine ja protsesside automatiseerimine (Van der Peijl et al. 2020). See võimaldab täpsemaid pakkumisi, võib vähendada eelarvest üle minekut, parandada hangetega seonduvat tõhusust ja säästa kulusid organisatsioonile läbi administratiivse koormuse vähendamise.

Kolmandaks, AI-põhiste lahenduste kasutus pakub mitmeid potentsiaalseid kasutegureid süsteemi- ja ühiskonna tasandil. Nende kasutus poliitikakujundamises aitab algatada uusi initsiatiive ja luua ühtset arusaama erinevate osapoolte vahel läbi tunnetuslikult objektiivse andmekasutuse (Kolkman 2020). Masinõppe lahenduste kasutus teenuse osutamisel võimaldab seeläbi suurendada avaliku sektori otsuste legitiimsust. See on eriti relevantne kontekstides, kus usaldus nii otsusetegemise protsesside ja riigiteenistujate puhul on minimaalne ning masinotsustus loob tunnetuslikult erapooletu otsusetegemise konteksti (Miller & Keiser 2020). Automatiseeritud otsustusmehanismidega on võimalik tagada marginaliseeritud või ka väljajäetud gruppidele võrdne kohtlemine. Näiteks automaatne dokumentide tõlkimine võib pakkuda vähemusgruppidele võimalusi enda eelistuste väljendamiseks ning AI-põhised lahendused võivad pakkuda kodanikele rohkem informatsiooni poliitikakujundamise protsessides osalemiseks (Savaget et al. 2019). Poliitikakujundamise protsessis võib AI-põhiste lahenduste kasutamine viia otsusetegemise võimekuse kasvuni läbi parema andmetöötlemise võimekuse (Mcneely & Hahm 2014). Teenuseosutamise puhul võib kasu ilmneda läbi informatsiooni tõhusama kogumise ja andmeanalüüsi, mis võib luua ressursside kokkuhoidu nii organisatsioonidele kui ka kodanikule (Pencheva et al. 2018).

Samas tuleb arvesse võtta, et täna on kirjanduses käsitletud positiivsed mõjud valdavalt eelduslikud ning positiivsete mõjude empiiriline valideerimine on olnud piiratud (Kuziemski & Misuraca 2020).

Nagu iga tehnoloogia puhul, võib ka AI-põhiste rakenduste kasutamine viia terve rea erinevate negatiivsete tagajärgedeni (O'Neil 2016; Green 2019; Kitchin 2017). Põhjused võivad olla nii tehnilised (nt alusandmete ebatäpsus) kui ka institutsionaalsed (nt automatiseerimisega seotud kitsaskohtade teadlik eiramine). Ka negatiivseid tagajärgi saab eristada nii indiviidi, organisatsiooni kui ka ühiskonna tasandil. Need võivad omalt poolt ilmneda esmasel rakendamisel kui ka võivad olla tegemist tagajärgedega, mille mõjud avalduvad aastate jooksul (Bailey & Barley 2020).

Esiteks, indiviidi tasandil võivad masinõppel põhinevad lahendused vähendada ligipääsu kriitilisele informatsioonile ning sellest tulenevalt võimekust erinevaid olukord tõlgendada. Seeläbi võib see tekitada lisakoormust töötajatele, kes peavad otsima alternatiivseid viise vajaliku kontekstipõhise informatsiooni saamiseks (Thunman et al. 2020). Samuti võib see viia tunnetatud otsustusõiguse vähenemiseni ja kommunikatsioonikanalite ahenemiseni, mida seostatakse töövõimaluste kadumisega. AI-põhiste lahenduste kasutamine muudab organisatsioonis erinevate indiviidide rolle ja võimujaotust (Zouridis et al. 2018), mõjutades seda, kes ja kuidas organisatsioonis panustavad. Samuti võib masinõppel põhinevate lahenduste kasutus viia liiase tsentraliseerituseni.

Teiseks, organisatsioonide tasandil võib AI-põhiste lahenduste kasutamisel muutuda organisatsioonide eesmärkide seadmise protsess, lähtudes mitte enam probleemide olemusest, vaid tehnoloogilistest võimalustest (Pencheva et al. 2018). Kuigi masinõppel põhinevad ja AI-põhised lahendused pakuvad abi keeruliste ühiskondlike lahendustega tegelemisel, siis üksikud lahendused ei ole piisavad kompleksete probleemide lahendamisel (Green 2019). Sotsiaalsete probleemide keerukust on keeruline tõlkida masinanalüüsiks sobivasse keelde ning seeläbi on risk liiga arbitraarsete või lihtsustatud näitajate kasutamisel (Morozov 2013). Teisalt andmekvaliteedi madal tase võib viia tõlgendamise probleemideni ja vigadeni otsusetegemisel, mis omakorda võib viia ressursikulude suurenemiseni (Kettl 2016) ning andmeõigluse vähenemiseni (Wirtz et al. 2019).

Kolmandaks, ühiskonna/süsteemi tasandil võib AI-põhise protsesside automatiseerimine viia töökohtade kaotuseni, samuti isikuõiguste riive ning andmeõigluse vähenemiseni tulenevalt andmete ebatäpsusest, nende kujunemise ajaloost ning algoritmide aluseks olevatest vääreeldustest (Kitchin 2020; O'Neil 2016). Automatiseeritud otsustusprotsesside läbipaistmatus, mis võib tuleneda nt ärisaladuse kaitsest või masinõppeprotsesside keerukusest, võib viia vastutuse hajumiseni ning seeläbi avaliku sektori legitiimsuse ja usaldusväärsuse vähenemiseni. Olukorras, kus avalikul sektoril endal puudub sisemine võimekus AI-põhiseid lahendusi luua või hinnata, võib nende lahenduste sisseostmine erasektorist viia avaliku halduse otsustusprotsesside varjatud „erastamiseni". See võib näiteks kaasa tuua olukorra, kus avalik sektor ei saa mõjutada masinõppelahenduste disainivalikuid ega ole hiljem võimeline tarnija poolt tehtud valikutest aru saama (O'Neil 2016). Eelneva tulemusena on risk erinevate ühiskonna gruppide diskrimineerimiseks.

5.2. Elluviidud projektide mõju

Mõjuhinnangu läbiviimise eesmärk oli kaardistada projektis väljatöötatud lahenduste mõjud lähtudes kirjanduse põhjal loodud analüütilisest raamistikust. Selleks toetuti nii semi-struktureeritud intervjuudele arendajate ning tellijatega kui ka projektiga seotud dokumentide analüüsile. Lumepallimeetodil loodud intervjueeritavate valimi moodustamisel lähtuti nende: 1) olulisusest ja 2) rollist vastavas projektis, saamaks projektide mõjudest võimalikult mitmekülgne vaade. Kokku on seni läbi viidud 6 intervjuud 10 projektidega seotud osapoolega (sh 4 tellijate esindajat ja 6 arendajate esindajat). Tänu kõikide projektide varajasele faasile, oli intervjueeritavate hulk piiratud. Intervjuude läbiviimiseks loodud küsimustikku kohandati vastavalt intervjueeritava rollile projektis: arendajate puhul kasutati intervjuude protokolli

disainiprotsessi sektsiooni, tellijate puhul käsitleti lisaks disainiprotsessile ka organisatsiooni struktuuri, ideoloogiat jm puudutavat sektsiooni.

5.2.1 Kesksed leiud intervjuudest

Nagu sissejuhatuses viidatud, siis projektid on rakenduslikus mõttes täna alles väga varases staadiumis, mistõttu oluliste mõjude ilmnemisest on täna veel vara rääkida. Küll aga on võimalik välja tuua mõned tendentsid, mis pikemas perspektiivis võivad ära määrata, millist mõju projekti käigus väljatöötatud lahendused avaldama hakkavad:

1. Masinõppel põhinevate lahenduste puhul on piisavate ja usaldusväärsete andmete olemasolu võtmetähtsusega. Intervjuudest nähtub, et siinjuures esineb täna Eesti avalikus sektoris mitmeid süsteemseid ja tehnilisi probleeme, mis raskendab lahenduste väljatöötamist ning pikemas perspektiivis nende mõju. Lähemalt kirjeldavad neid probleeme konkreetsete projektide kirjeldused (vt osa 3.6. ja 4).
2. Lisanduv väljakutse ilmnes andmete konfidentsiaalsusega, mis võib pikemas perspektiivis mõjutada AI-põhiste mudelite ja otsustusprotsesside loomise võimalikkust. Esiteks, tänu masinõppe põhinevate lahenduste uudsusele eksisteeris projektides suurem vajadus tuvastada nii andmete ristkasutuse võimalusi kui ka ulatust. See hõlmab lisanduvat ajakulu, osapoolte kaasamist ning teadmiste arendamist rakendajaorganisatsioonides. Antud väljakutse pole iseloomult vaid juriidiline, sest osapooltel võib esineda isiklik konflikt teatud tüüpi andmete kasutusega analüüsi teostamiseks (nt tundlikud isikuandmed riskiskoori määramisel). Detailsem informatsioon on kättesaadav eespool konkreetsete rakenduste alt (osa 4.).
3. Intervjuudest nähtus ning seda kinnitavad ka eelpool olevad projektide kirjeldused, et mudelite võimalik omaksvõtt ning kasutus tellijaorganisatsioonide ning nende töötajate poolt on täna veel väga lahtine. Mudelite loomist domineerisid arendajad, kel oli reeglina selgeim idee ja visioon lõpplahenduse võimalustest. Neljast projektist ühe puhul on näha tellijapoolset võimekust arendajate poolt loodud lahendust oma organisatsiooni kontekstis lahti mõtestada, mis loob paremad eeldused ka võimalike positiivsete mõjude ilmnemisele tulevikus. Ühes projektis oli tellijatel olemas arusaam protsessist ja lõpplahendusest, kuid see arusaam ulatus vähesel määral kaugemale üksikutest projekti kaastatud isikutest. Kahes projektis selget tellijat ei olnud. Seega, nende kolme projekti puhul on väga raske spekuleerida nende mudelite võimalike positiivsete mõjude üle olukorras, kus nende võimalik kasutus on väga lahtine.
4. Kõigi projektide puhul on ka keeruline hoomata üksikspetsialistide/lõppkasutajate tegelikku käitumist, kuivõrd nende kaasatus on seni olnud minimaalne. Potentsiaalne väljakutse võib olla seniste praktikate kokkusobimatuse (nt hoiatuste sildistamisega küberturvalisuse MVP puhul) kui ka mudelite omaksvõtuga (nt mittenõustumine muutujatega töötukassa MVP riskihindamise osas).
5. Rakendaja organisatsiooni puhul tekib vajadus lahendus rakendajaorganisatsiooni sees „maha müüa“ lõppkasutajatele. Esialgsed lahendused töötati reeglina välja piiratud hulga inimestega, mis võib tekitada lisanduvaid väljakutseid rakendamise erinevates faasides. Seoses lõppkasutajate-töötajate rollist tingitud vastutusest, on nende tundlikkus riskide ja ebakindluse suhtes kõrgem. See mõjutab nende otsuseid olukordades, kus eksisteerib ebamäärasus ja teadmatus masinõppepõhiste mudelite toimimisest. Kokkuvõttes võib see kaasa tuua kas tulevase vähese kasutuse või kasutuspraktikad, mis hälbivad oluliselt algsest disainiidees.
6. Olemasolevate uuringute pinnalt võib eeldada, et masinõppelahenduste potentsiaalsed kasutajaorganisatsioonid Eesti avalikus sektoris on väga erineva majasisese tehnoloogiliste

võimekustega sõnastamaks ning tellimaks enda vajadustele vastavaid masinõppelahendusi. Väliste arenduskompetentside kasutamisel võib olla suuremaid väljakutseid erinevate arenduste omavahelise sõltuvusega, kus muutused ühe arenduse raames toovad probleeme teiste arendustega seoses. Siinkohal võivad need probleemid olla nt metoodilised, kus andmete klassifitseerimine kui ka andmeväljade muutmine ühes andmekogus võib tekitada probleeme teises arenduses.

7. Intervjueeritud arendajate ning rakendajate hinnangud on suuresti ettepoole vaatavad ning spekulatiivsed tänu piiratud kasutuskogemusele. Arendajate puhul on selge omapoolne rõhuasetus olnud läbipaistvuse tagamisel ja selgituse pakkumisel, vähendamaks mudeli „musta kasti“ iseloomu. See on ka olnud üheks kriteeriumiks mudelite valikul, et oleks võimalik seletada mudelite toimivust ka lõppkasutajale. Seega on näha praktikaid, kus läbi mudelites olevate muutujate omavahelise seose selgitamise võimaldab potentsiaalsetel lõppkasutajatel paremini hoomata lõppotsust mõjutavatel protsessidel.

Kokkuvõtvalt on selge, et masinõppe/AI toega teenustel on võimalik ette näha nii ulatuslikke positiivseid mõjusid kui ohte mõtlematu rakenduse puhul (vt tabel 5). Eeldades, et teenuste loomise ja haldamise tehnilised raksused on ületatavad vastava kompetentsi kasvatamisega teenuse omanike seas, on ilmselt suurim väljakutse nn operatiivse tasandi kasutaja ja masinõppe/AI võimekusega infosüsteemi efektiivse koos töötamise tagamine. See tähendab masinõppe lugemisoskuse tekitamist ja teenuse osutamise äriprotsessi muutmist. Vaid sel juhul on võimalik saavutada ka need kitsad organisatsioonilised eesmärgid, teenuse osutamise suurem efektiivsus, täpsus, kulutõhusus ja kvaliteet, mida teenuste omanikud või rakendavate asutuste juhid eelkõige näha soovivad.

6. Soovitused

Antud peatükis on lühidalt ära toodud raporti soovitused, need on jagatud üldisteks, lahenduste kasutusviise ja disaini puudutavateks ja spetsiifilisi domeene puudutavateks.

6.1. Üldsoovitused

Soovitus 1. *Prioriseerida ennetustegevuseks mõeldud teenustes masinõppe/AI toe rakendamist.*

Põhjendus. Ennetustegevust viiakse läbi suurema ebakindluse olukorras ehk sellega soovitakse vähendada veel mittetoimunud negatiivsete sündmuste riski. Selleks on vajalikud täpsed prognoosid tulevikusündmuste osas. Sobivate alusandmete olemasolul on masinõpe/AI selleks erakordselt võimekas tehnoloogia. Andmetega kaetus riigi erinevate andmekogude näol on väga hea, andmed on lingitavad ja teisese kasutuse raskustest hoolimata on tehniliselt võimalik luua täpseid mudeleid. Ennetustegevus lubab suurendada ühiskonna turvalisuse taset ja läheb hästi kokku riigi teenuste pro-aktiivseks muutmise initsiatiiviga, lisaks vähendab see vajadust tegeleda hiljem kallite negatiivsete tagajärgedega (vt osa 3.1. ja 3.2.).

Soovitus 2. *Luua ESA juurde masinõppe toega teenuste arendamise sandbox.*

Põhjendus. Reeglina on teenuse toeks oleva mudeli arendamiseks vajalik kasutada kõikseid andmeid ehk näha nii neid juhte, kus mingi organisatsiooni jaoks relevantne sündmust esineb (töötus, tulekahju, haigus) kui ka ei esine. Viimasel juhul aga ei ole tihti õiguslikku alust neid andmeid haldustegevuse läbiviimiseks näha ega kasutada. ESA sees on sellised analüüsid aga võimalikud ning andmete kaitstud käitlemine juba

olemasolevate protseduuridega tagatud. Tasuks kaaluda viise kuidas ESA saaks hakata ka selliste teenuste toimimist tulevikus tehniliselt pakkuma (vt ka osa 3.6.1., 4.1. ja 4.2.).

Soovitus 3. *Tuua sündmusteenuste arendamisse masinõppe/AI komponent.*

Põhjendus. Sündmusteenused on hetkel teadaolevate plaanide järgi eelkõige ex-post ja deterministlikud ehk juba toimunud sündmuse järel nende teenuste pakkumine, mille puhul on tekkinud riigil kohustus neid pakkuda või kodanikul sündmuse tulemusena õigus soovi korral teenust tarbida. Sündmusteenus toimib selle juures üle eri asutuste halduspiiride. Andmete ristkasutusse tehtavate investeeringute ja arenduste raames oleks loogiline kasutada võimalust luua masinõppe tuge just neile sündmusele reageerimise osadele, kus riigi teenusega sekkumise või abi pakkumise tulemus on lahtine ehk mitte determinineeritud tulemiga (vt osa 3.3).

Soovitus 4. *Tugevdada andmeteaduse kompetentsi organisatsioonide juures.*

Põhjendus. Masinõppe/AI toega teenuse oluline osa on andmetele ehitatud masinõppe mudeli ja selle korrektse toimimise tagamine läbi mudeli monitoorimise ja regulaarse ümbertreenimise. Juhul kui masinõpet kasutatakse suure mõjuga individualiseeritud teenuse toe juures on omakorda oluline mudeli empiirilise täpsuse valideerimine. Need ülesanded lähevad kaugemal baas IKT toe pakkumisest ja nõuavad ideaalses olukorras andmeinseneri, rakendusstatistiku/matemaatiku ja limiteeritud kasutajaliidese arendaja oskust katva ametikoha ehk andmeteadlase tuge. Selline võimekus võiks olla kas organisatsiooni enda majas või seda võiks pakkuda ESA või RIA tsentraalselt erinevatele asutustele, sest organisatsioon ise ei pruugi täiskohaga andmeteadlast ära koormata (vt ka osa 3.6.3).

Soovitus 5. *Parandada andmekogude metaandmete kvaliteeti ja täiendada andmedoonoreid ja tarbijaid ühendav muudatuste automaatse raporteerimise reeglistik.*

Põhjendus. Arvestades projektis katsetatud rakendusvaldkondi on just mudelite empiiriline täpsus suure kaaluga. Nende alusel kas sekkutakse otse inimeste ellu või antakse soovitusi, millel on reaalsed elulised tagajärjed inimesele või organisatsioonile ehk need on isikule või ühiskonnale potentsiaalselt suurte otseste mõjudega AI/ML rakendused. Täpsus saavutatakse eri andmete ristkasutusega, kus võimenduvad tehnilised vead, mis võivad muuta mudeleid vigadeks kui neid ei suudeta eelnevate kontrollidega tuvastada (vt Lisa 1.). Näiteks võiks kaaluda RIHAs dokumenteeritud infosüsteemide ja andmekogudele ka andmete väärtuste kirjelduse lisamist ning andmete genereerumise protsessi lühidast kirjeldust (vt ka osa 3.6.4.). Oluline on ka andmekogude kuni väljade tasemeni minevate muudatuste raporteerimise süsteem andmedoonorite ja võimalike ristkasutajate vahel.

Soovitus 6. *Täiendada MKM juhiseid tehisintellekti süsteemide loomiseks, lisades sinna teavet eetiliste aspektide ja õiguslike nõuete kohta.*

Põhjendus. Nii projekti läbi viidud õiguslik hinnang, kui ka andmete taotlemise protsess ning hilisem mudelite spetsifitseerimise protsess, tõid välja rida kohti, kus kas teatud kasutusviis oleks eetiliselt ja õiguslikult problemaatiline või oleks vaja õiguslike probleemide vältimiseks vajalik kasutada teistsugust tehnilist rakenduse loomise viisi (näiteks mudelobjekti vs päris andmete migreerimine). Teenusele masinõppe tuge tellival asutusel on tihtilugu keeruline kõiki eetilisi ja õiguslikke aspekte eelnevalt hinnata, sest masinõppe/AI mudeli toimimine on empiiriline küsimus ja olenevalt lõplikult kasutavavast andmekoosseisust võivad osad küsimused mudeli ehitamise käigus laheneda. Oluline on, et neid küsimusi ei langetaks mudelit ehitav tehniline töötaja, vaid et tellija oleks suuteline neid eelnevalt enda jaoks

teadvustada ja modelleerijatele kommunikeerida. Täiendavad juhised lubaks selliseid kitsaskohti kiiremini märgata.

Soovitus 7. *Kaaluda algoritmilise mõjuhinnangu või eelhinnangu nõude sisseseadmist.*

Põhjendus. Projektis piloteeriti eelkõige negatiivseid riske ennetavate teenuste prototüüpe. Negatiivsed riskid tabavad suurema tõenäosusega ühiskonna kõige nõrgemaid ja haavatavamaid isikuid. Samas oleks nende aitamine kõige suurema mõjuga üldise ühiskonna turvalisuse taseme tõstmisel. Algoritmiline mõjuhinnang lubaks tuvastada ja leevendada võimalikke ühiskondlikke riske ja ära hoida kahju enne tehisintellekti süsteemi kasutusele võtmist. Samuti vähendaks see riske teenuse arenduse tellimisel. Tehisintellekti süsteemide keelustamise asemel seadusandlikul tasandil võiks Eesti loobuda selliste tehisintellekti süsteemide kasutamisest, mille mõjuhinnang näitab, et süsteem tõenäoliselt rikub põhiõigusi (vt Lisa 4).

Soovitus 8. *Täiendada olemasolevat seadusandlust või kehtestada kodanikel konkreetne õigus saada selgitust tehisintellekti/masinõppe kasutamise kohta avalikus sektoris.*

Põhjendus. Eesti õigusaktid ei sisalda konkreetseid sätteid kodaniku õiguse kohta nõuda selgitusi tehisintellektipõhiste otsuste kohta. Seega võiks Eesti seadusandja läbipaistvuse tagamiseks kaaluda võimalust konkretiseerida olemasolevaid sätteid ja/või kehtestada eraldi selgesõnaline õigus saada *tagantjärele* selgitusi temale teenuse osutamisel kasutatud tehisintellekti süsteemide kohta koos selleks tema kohta kasutatud alusandmete koosseisu kirjelduse ja mudeli väljundväärtusega. See annaks ka teenuste arendajale selge motivatsiooni võtta selgitatavuse nõue üheks keskseks teenuse masinõppe/AI statistilise mootori kriteeriumiks ning looma selgitatavust toetavat analüütilist tuge ning sisuliselt teadmist teenuse lõplike omanike juures (vt Lisa 4).

Soovitus 9. *Kaaluda kõrge riskiga tehisintellekti süsteemide siseriikliku registri loomist suurema läbipaistvuse tagamiseks.*

AIA ettepanekus on ette nähtud üleeuroopalise registri loomine kõrge riskiga tehisintellekti süsteemide jaoks ilma, et liikmesriikidel oleks keelatud selliseid algatusi ka riiklikul tasandil kasutusele võtta. Sellise registri loomine võiks aidata parandada läbipaistvust ja algoritmilise vastutuse standardeid. Näiteid avaliku halduse tehisintellekti süsteemide registritest leiab mh Prantsusmaal, Soomes ja Kanadas,[28] samuti kogub Eesti Majandus- ja Kommunikatsiooniministeerium teavet avalikus sektoris kasutatavate tehisintellektide kohta ja on teinud avalikult kättesaadavaks olulisemate tehisintellekti süsteemide nimekirja veebilehel https://www.kratid.ee/kasutuslood . Kehtivad Eesti seadused ei näe ette ei kõrge riskiga süsteemide ega muude tehisintellekti süsteemide õiguslikku registreerimise kohustust, kuid seda võiks kaaluda arvestades mõne sellise toega teenuse suurt mõju (vt Lisa 4).

Soovitus 10. *Kaaluda seadusemuudatusi kaitsmaks kodanikke masinõppel põhinevate diskrimineerivate otsuste eest.*

Eesti peaks kaaluma täiendavad seadusmuudatused kaitsmaks kodanikke masinõppel põhinevate diskrimineerivate otsuste eest. Kuna ELi diskrimineerimisvastane õigus ei takista liikmesriikidel sätestada rangemaid eeskirju inimeste kaitseks diskrimineerimise eest, on Eesti seadusandja vaba kaaluma

[28] "Algorithmic Accountability for the Public Sector." (2021) Ada Lovelace Institute, AI Now Institute and Open Government Partnership (n 45) 19 <https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/>.

erinevaid õiguslikke võimalusi, et ära hoida diskrimineerimist tehisintellekti kasutamisel avalikus sektoris. Probleemid tekivad muuhulgas seoses diskrimineerimisvastaste õigusaktide (piiratud) kohaldamisalaga, "kaudse diskrimineerimise" ebaselge mõistega tehisintellekti süsteemide poolt juhitud suurandmete analüüsi kontekstis ning seoses raskustega diskrimineerimise tuvastamisel algoritmiliste otsuste tegemisel, mis on tingitud algoritmide tööprotsesside läbipaistmatusest.

Näiteks võiks Eesti (i) töötada välja "andmete kvaliteedi" kriteeriumid, et vältida andmekogumite kallutatust algusest peale; (ii) võtta kasutusele meetodid algoritmide läbipaistvuse ja selgitatavuse suurendamiseks, et kodanikud oleksid võimelised ära tundma võimalikku diskrimineerimist, ja/või (iii) otsustada keelata teatavate tehisintellekti süsteemide kasutamine haldusotsuste tegemisel, kui esineb tõsine diskrimineerimise oht ja/või otsustada neid süsteeme mitte kasutada (vt Lisa 4).

6.2. Kasutusviiside ja disaini soovitused

Soovitus 11. *Prioriseerida vertikaalsete kasutusviisidega masinõppe/AI toega teenuste arendust.*

Põhjendus. Masinõppe/AI mudel toob kasu ainult siis kui tema väljundi alusel asutuse äriprotsessi muudetakse. Efekt tuleb teistmoodi tegutsemisest, mitte mudelist endast. Seega on oluline veenda ka asutuse nn operatiivne ametnike tasand lahenduste kasulikkusest, sest nende tegevusest summeerub tegelik ühiskondlik kasu. Kuigi teenustele masinõppe/AI toe loomine on ilmselgelt võimalik vaid juhatuse ja kesktaseme juhtide ja teenuseomanike soovi korral asutuse protsesse efektiivsemaks ja kulutõhusamaks ning teenuseid kvaliteetsemaks muuta ehk soov tuleb ülalt alla, võiks teoorias masinõppe/AI rakenduste suurim kasu tekkida kui operatiivsel tasandil tekkiv väärtus alt üles agregeeritakse ja igal astmel nii tekkivat analüütikat protsesside ja inimeste juhtimisel kasutatakse (vt ka osa 3.4.).

Soovitus 12. *Koolitada lõppkasutajaid masinõppe/AI sisus.*

Põhjendus. Olenevalt teenuse täpsest kasutusviisist tuleks planeerida lõppkasutajatele üldiseid masinõppe/AI koolitusi. Näiteks AI abil otsustavale ametnikule/teenistujale masinõppe ja AI mudelite lugemise oskust domeenides, kus mudeli väljundi alusel viikase läbi mingi sekkumine või antakse soovitus isikule (näiteks määratakse toetus, suunatakse töötuse riskiga inimene ümberõppele, külastatakse tuleohu riskipiirkonna majapidamisi, tehakse raviotsus jne) (vt osa 3.6.5.).

Soovitus 13. *Lisada masinõppe/AI toega teenustele arendamisel mudelite lugemist toetava kasutajaliidese ehitamise nõue.*

Põhjendus. Suure mõjuga ML/AI toega teenuste loomisel on vajalik nii teenuse omanike kui lõppkasutajatele põhjaliku arusaama tekitamine, mida teenuse aluseks olev mudel teeb ehk luua ML/AI kriitiline lugemisoskus. Selliste teenuste arenduse hankimisel peaks seega eraldi rõhutama ka vajadust mitteekspertide poolset lugemisvõimet toetava kasutajaliidese disaini tellimist ja kasutuskoolituse läbiviimist. Lisaks on oluline demonstreerida näidisjuhtudega kuidas masinõppe mudel indiviidi andmeid kohtleb, näiteks kui masinas on juhumets peaks kasutaja üldistatud tasemel mõistma, et juhumetsas leitakse andmetest otsustuspuud, kuidas need tegelike kodanike andmete peal luuakse, kuidas nende pealt kombineeritakse kokku metsa ehk mis on üldised rakendatud reeglid antud lahenduses, ja kuidas sealt valitakse prognoos, mida tema näiteks kliendi kohta näeb. Lisaks mõistma, mis andmeid masin näeb, saamaks aru, mida see masin ilmselgelt arvesse ei oska võtta (vt osa 3.6.5.).

Soovitus 14. *Lisada masinõppe/AI toega teenuste arendamisel kasutaja tagasiside korjamise nõue.*

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

Põhjendus. Struktureeritud ja lihtne tagasisidestamise süsteem, mis oleks näiteks ennetuseks mõeldud riskiskoorimise mudelitel põhineva teenuse puhul kasutaja hinnang riskiskoorile ja selle seletuse adekvaatsusele, arendab esiteks kriitilist masinõppega koostöö oskust ning võimaldab teiseks mudelit tulevikus paremaks teha, sest kasutaja suudab kohati erijuhte identifitseerida ja sel kujul paremini sildistada. Eriti sotsiaalteenuste osutamisel on see oluline säilitamaks teenistuja diskretsiooni õigust otsuste juures (Thunman et al. 2020) ning vähendada ohtu nn masinaga mittekonformse käitumisviisi vältimiseks kuna tööprotsess, selle monitoorimine juhtide poolt või asutuse muud protsessid soodustavad just konformset/protseduurist lähtuvat otsustamist (Meijer et al. 2021) (vt osa 3.6.5.).

Soovitus 15. *Lisada masinõppe/AI toega teenuste arendamisele teenuse automatiseeritud mõju hindamise sisseviimise soovitus.*

Põhjendus. Masinõpe mudelite loomise raames selgus, et teenuste mõju hindamine on kohati metodoloogiliselt keeruline ehk organisatsioonidel endal puudub teaduslikult vettpidav viis kuidas hinnata, kas osutatav teenus annab soovitud efekti. Kuna teenusele masinõppe tugede loomiseks on vaja ette võtta ulatuslik andmete eeltöötlus ja liidestamine ning ka väljundi otsa analüütika loomine, oleks mõistlik selle sama töö raames luua vähemalt ka tehniline võimekus teenuse efekti automatiseeritud hindamiseks (vt osa 3.5.).

6.3. Domeenispetsiifilised soovitused

Soovitus 16. *Sotsiaalministeerium võiks võtta kandva rolli Eesti andmekogude teisendamisel OMOP CDM kujule, ning kaaluda õigusruumi piiride raames võimalusi innovaatiliste lahenduste piloteerimiseks ilma meditsiiniseadme tootmise/tootjavastutuse vajaduseta.*

Põhjendus. Erinevalt teistest antud uuringus analüüsitud domeenidest ei ole tervisevaldkonnas sama asutus vastutav kogu protsessi eest ehk populatsioonitaseme ennetuse ja patsienditasandi ennetuse ning patsiendi tasandi ravi ja ravi efektiivsuse hindamise eest vastutavad erinevad osapooled. Kombinatsioonis mudeliga, kus raviasutuste rahastus on seotud ravi, mitte ennetusega, ei ole ka kedagi kelle huvi oleks just sellisel kujul kogu populatsiooni hõlmavate indiviiditasandi mikroandmetel loodud, aga vertikaalselt kasutatavat (populatsioonitasandi ennetus, patsienditasandi ennetus, patsiendi ravi, raviefektiivsuse hindamine), masinõppe põhist abimeest luua ja nii horisontaalselt kui vertikaalselt juurutada. Antud projekt demonstreeris, et andmete standardiseerimine piisava kvaliteeditasemega on võimalik, mis annab olenevalt olukorrast võimaluse lõigata ära see ligi 70-80% ajakulust, mille võtab masinõppe mudeli rakendamisel andmete liitmine, puhastamine ja ümberkodeerimine ja mida teevad kõik uuringumeeskonnad üha uuesti ja uuesti. Selle tulemusena saaks oluliselt kiiremini valideerida Eesti andmete peal mujal väljatöötatud algoritme terviseriskide paremaks haldamiseks (vt osa 4.3.3).

Soovitus 17. *Ühtlustada terviseandmete kasutamise lubade taotlemine Haigekassa, TEHIKU ja Sotsiaalministeeriumi kui kesksete terviseandmete kogude volitatud ja vastutavate töötlejate vahel ning võimaldada terviseandmete uurimine vaid ühe eetikakomitee loa alusel vältimaks dubleerivate lubade küsimise vajadust.*

Põhjendus. Antud uuringu puhul tuli EBINi kui „Sotsiaalministeeriumi juures tegutsev ning valdkonna eest vastutava ministri loodud sõltumatu komitee“ loa järel uuesti küsida luba Sotsiaalministeeriumilt, kelle loa järel TEHIK sai vastavad andmed väljastada. See kujutab ennast olulist lisatakistust andmete analüütilisel väärindamisel ja ei ole selge, miks ministeeriumi poolt uuringulubade väljastamiseks volitatud komitee otsuse järel sama volitaja käest uuesti luba küsima peab. Olukorras, kus uuringueesmärgi

saavutamiseks on vajalik erinevad andmekogud kokku panna, on seega paratamatult vaja küsida kahe eetikakomitee luba, mille taotlusvorm ja protsess küll erineb, kuid mille sisuks olev uuringu põhjendatuse, vajalikkuse ja uuritavate kaitseks loodud protseduuride kirjeldus on sisult samaväärne (vt osa 4.3.3).

Soovitus 18. *Analüüsida riigiasutuste küberkaitse infrastruktuuri masinõppepõhise küberohtude leidmise rakendatavuse vaatepunktist.*

Põhjendus. Antud projekti raames demonstreeriti masinõppe rakendatavust massiivse koguse küberohtude automaatse markeerimise ja prioriseerimise jaoks. Samas selgus ka, et rakenduse kasulikkus sõltub olulisel määral asutuse küberkaitse organisatsioonist ja nende käsutuses olevatest andmevoogudest. Seega printsiibis ei ole sellisel kujul organisatsiooni küberkaitsevõime tugevdamisel hetkel ilmselt takistusi, see on tehniliselt tehtav nagu projektis demonstreeriti, kuid vajab olulist organisatsioonipõhist kohandamist loomaks võimalikult laialt kasutatav ettevalmistatud andmevoogude juhtimise ning masinõppe käivitamise kasutajaliides (vt osa 4.4.).

7. Kokkuvõte

Projekti eesmärk oli piloteerida masinõppe/AI rakendatavust neljas erinevas valdkonnas saamaks aru:

1. millised avaliku sektorid teenused oleksid kõige sobivamad masinõppe ja AI otsustustoe rakendamiseks;
2. kas riigi IT infrastruktuur on sobiv masinõppe ja AI laiemaks kasutamiseks;
3. mil viisil mõjutaks masinõppe/AI otsustustugede laiem kasutuselevõtt avaliku sektori toimimist;
4. kas praegune õiguskord toetab selliste tööriistade kasutuselevõttu;
5. mil viisil oleks masinõppe/AI toega teenuseid võimalik kasutusele võtta suurima kasu saamiseks, samas tekkivaid riske minimeerides;
6. millised õppetunnid masinõppe/AI sisuga rakenduste katsetamisest on olulised tulevikus analoogsete projektide läbiviimisel avalikus sektoris.

Rakendusvaldkondadeks olid tööturg, sisejulgeolek, e-tervis ja küberjulgeolek. Rakenduste sisuks oli vastavalt: töötuse riskiskooride masinõppe ennustusmudel; eluhoonete tulekahjude riskide masinõppe ennustusmudel; andmete standardiseerimise lahendus ja terviseväljundite prognoosimise masinõppe mudelid; küberohtude tuvastamise masinõppe mudel.

Reaalselt väljatöötatud masinõppe lahenduste tulemusena võib uurimisküsimuste osas kokkuvõtvalt järeldada järgmist.

Esiteks, kõiki valdkondi ühendav probleem abstraktsel tasemele on vajadus otsustada suure ebakindluse olukorras, mis omakorda tingib vajaduse saada otsuse aluseks võimalikult täpseid prognoose riskide osas, mille suhtes otsus langetatakse. Just sellises olukorras annab masinõppe/AI tugi suurima lisandväärtuse, sest antud tehnoloogia on võimeline andma täpseid prognoose suuremahulise ja kohati vähestruktureeritud andmestiku pealt. Kõik rakendused on suunatud nn halbade riskide realiseerumise ärahoidmiseks ennetava tegevuse kaudu. Kuivõrd suure ühiskondliku mõjuga selliste riskide ärahoidmine on sõltub kahest asjaolust:

1. kas need riskid puudutavad suurt osa elanikkonnas/organisatsioonidest, kelle pealt summeerub kokku oluline mõju;

2. või nende riskide realiseerumise tagajärjed on isikutele/organisatsioonile väga tugeva otsese tagajärjega.

Esimest tüüpi mõjuga on selgelt tööturu, e-tervise ja küberjulgeoleku rakenduskohad, teist tüüpi ühiskondliku mõjuga on tulekahjude ennetamise ning osaliselt ka e-tervise rakenduskoht. Lisaks nendele näidetele hinnati ka masinõppe/AI kasutusvõimalusi teenustel, mida riik pakub pärast mingi sündmuse toimumist, aga mida iseloomustab samuti vajadus saada teenuse pakkumiseks prognoose potentsiaalselt ebakindla tuleviku suhtes. Sellised kasutuskohad on kindlasti kõrge riskiga, kuid eduka rakendamise korral on need ka kõige suurema ühiskondliku kasuga. Kokkuvõtvalt on soovitus arendada masinõppe toega teenuseid rohkem just ennetusse suunatud teenuste toeks ning nn sündmusteenuste kontekstis, kus sündmuse järel on vajadus kodanikule abi/soovitusi anda, teadmata deterministlikult, mis selle tagajärg olla võib (näiteks tööturuteenuse soovitus töötule).

Teiseks, masinõpe on valitud valdkondades Eestis empiiriliselt rakendatav, loodud prognoose pakkuvate mudelite puhul on võimalik saavutada kõrge või väga kõrge täpsuse tase. Masinõppe/AI kasutamisele IT tehnilisi takistusi ei ole, takistused on pigem andmete ristkasutuse korralduses ja õiguslikus ebakindluses teatud kasutuslugude puhul. Empiiriliselt täpsete mudelite loomine on võimalik tänu väga heale andmetega kaetusele, kuid on võimalik vaid andmete ristkasutusega ning relevantse valdkonna nn kogupopulatsiooni vaatlemisega (näiteks töötuse sündmuse esinemine ja mitteesinemine; haiguse esinemine ja mitteesinemine, tuleõnnetuse esinemine ja mitteesinemine), mis võib olenevalt kasutuskohast tähendada vajadust kaasata andmeid, mida teenust pakkuv amet/organisatsioon hetkel oma haldusülesannete täitmisel ei kasuta.

Kõikide rakenduste, välja arvatud küberkaitse juures kasutati andmeid erinevatest allikatest, mis genereeruvad omaette äriprotsesside tulemusena, mille omanikud ja haldajad on erinevate valdkondade organisatsioonid. Kasutati küll kahte nn agregaatorit – TEHIK ja ESA – kes hoiavad väga suures mahus erinevaid andmeid, kuid sellest hoolimata toob selline andmete lahus genereerumise protsess endaga kaasa oluliselt suurema vajaduse andmete mikrotasandi koordineerimise järele tagamaks omakorda kolmanda organisatsiooni teenuse toimimine. Seega on masinõppe/AI teenustesse integreerimise sujuvamaks toimimiseks vaja tagada andmekogude muutuste raporteerimine ja andmete genereerumise protsessi detailne kirjeldus. See vähendaks oluliselt halduskoormust andmete omanikele tulevikus ning teeks teenuste omanikele lihtsamaks oma masinõppepõhise teenuse toimepidevuse tagamise.

Masinõppe/AI suuremahulisem kasutuselevõtt avalikus sektoris tekitab lisaks baas IKT toele aga ka vajaduse masinõppe mudelitele andmeteadlase toe pakkumise järele. Eriti olukorras, kus teenus on potentsiaalselt suure mõjuga, on oluline tagada mudeli korrektne ümbertreenimine, täpsuse diagnoosimine ja tekkida võivate vigade parandamine. Mõistlik on selline võimekus osaliselt tsentraliseerida ja pakkuda asutustele/organisatsioonidele domeenipõhist tuge.

Kolmandaks, mõju avaliku sektori toimimisele on hetkel hinnatav pigem oodatud efektide, mitte vaadeldud mõjude osas kuna masinõppe/AI rakendusi on hetkel reaalses kasutuses vähe. Esile kerkis tugev vajadus võimalikult täpselt defineerida andmete ristkasutuse ulatus ja asutuse äriprotsesside tulevane sõltuvus sellest. Antud väljakutse pole iseloomult vaid juriidiline, sest osapooltel võib esineda isiklik konflikt teatud tüüpi andmete kasutusega analüüsi teostamiseks (nt tundlikud isikuandmed riskiskoori määramisel). See tähendab oluliselt lisanduvat ajakulu, täiendava õigusanalüüsi vajadust ka otseselt asutuse sees, suuremat osapoolte kaasamist teenuse arendamisel organisatsiooni sees väljaspool kitsast antud teenuse omanike ringi ning laiemat teadmiste arendamist rakendajaorganisatsioonides.

Masinõppe/AI rakenduste omaksvõtu tagamine nn operatiivse tasandi kasutajate poolt on vaja rakenduste tellimise ja arendamise osas pidevalt arvesse võtta. Oodatud asutusesisese efektiivsuse saavutamiseks tähendab see reeglina vajadust koos masinõppe elemendi sissetoomisega ka sisulise äriprotsessi muutmist ehk kasutuslugude muutmist, vastasel juhul ei võeta rakendusi tööprotsessis omaks või kasutatakse neid viisil, mis võivad minna vastuollu eetiliste ja õiguslike põhimõtetega.

Tugevalt masinõppe mudeli põhise teenuse ehitamine suurendab sõltuvust arendajatest suuremal määral kui tavaline arendustöö. Tellival asutusel peab olema suutlikkus kriitiliselt hinnata mudeli ehitamisel tehtud andmeteaduslikke valikuid ja mudelite toimimise loogikat. Seda eelkõige olukordades, kus kitsad andmeteadlase metodoloogilised valikud omavad elulisi tagajärgi hilisemal teenuse osutamisel. Näiteks kodanike segmenteerimine riskigruppidesse kasutades masinõpet ja grupikuuluvuse alusel diferentseeritult teenuse pakkumine / ennetustegevuse läbiviimine, tähendab, et segmenteerimismudelis olevate klassifitseerimisreeglite sisu hakkab mõjutama kodanikele osutavate teenuste sisu.

Olemasoleva info põhjal võib eeldada, et masinõppelahenduste potentsiaalsed kasutajaorganisatsioonid Eesti avalikus sektoris on väga erineva majasisese tehnoloogiliste võimekustega sõnastamaks ning tellimaks enda vajadustele vastavaid masinõppelahendusi. Tagamaks masinõppe toega teenuste kvaliteeti ja normidele vastavaid kasutusviise oleks vajalikud nii põhjalikumad juhendmaterjalid masinõppe/AI rakendamise printsiipidest avalikus sektoris, kasutusviiside tutvustusi näidete abil kui ka otsest andmeteaduse tuge arenduse faasis ja teenuse elutsükli jooksul.

Neljandaks, praegune õiguskord ei takista selliste tööriistade kasutuselevõttu, kuid vajalik on läbivalt täpsustada suuniseid vältimaks masinõppe/AI laiema kasutuselevõtuga kaasnevaid võimalikke ohtusid. Lisaks peaks Eesti valitsus kaaluma nende suuniste õiguslikku staatust, eriti seda, kas suunised - kuigi nad ei ole kodanike poolt jõustatavad - peaksid olema vähemalt haldusasutuste jaoks siduvad.

Tehisintellekti süsteemide keelustamise asemel seadusandlikul tasandil võiks Eesti loobuda selliste tehisintellekti süsteemide kasutamisest, mille mõjuhinnang näitab, et süsteem tõenäoliselt rikub põhiõigusi. Võimalikud ohud, mis võivad tuleneda tehisintellekti kasutamisest avaliku halduse poolt, võivad olla väga erineva intensiivsusega. Seetõttu tuleks valitsuse tasandil algatada arutelu võimalike piirangute üle, et tagada ühtsed hindamiskriteeriumid ja nende ühtlane kohaldamine.

Eesti jaoks tähendab ELi poolt planeeritav tehisintellekti regulatsiooni (Artifical Inteligence Act - AIA) ennetav mõju, et esialgu on mõistlik hoiduda kõrge riskiga tehisintellekti süsteeme käsitlevate asjakohaste õigusaktide vastuvõtmisest riigisiseselt. Selle asemel on soovitatav, et Eesti tagaks oma seisukohtade kuuldavaks tegemise AIA õigusloomeprotsessis ELi tasandil ja võimalusel ka juba Euroopa standardiorganisatsioonide poolt tehtava standardimistöö käigus.

AIA ettepanekus on ette nähtud üleeuroopalise registri loomine kõrge riskiga tehisintellekti süsteemide jaoks ilma, et liikmesriikidel oleks keelatud selliseid algatusi ka riiklikul tasandil kasutusele võtta. Sellise registri loomine võiks aidata parandada läbipaistvust ja algoritmilise vastutuse standardeid. Kehtivad Eesti seadused ei näe ette ei kõrge riskiga süsteemide ega muude tehisintellekti süsteemide õiguslikku registreerimise kohustust. Luues registri vaid kõrgema riski kategooria tehisintellekti süsteemidele tuleb samas näha ette kontroll haldusorgani otsuse üle, mis klassifitseerib süsteemi “kõrgema riski kategooriaks”.

Viimaks tekib küsimus, kas Eesti peaks vastu võtma täiendavad seadusmuudatused kaitsmaks kodanikke masinõppel põhinevate diskrimineerivate otsuste eest. Kuna ELi diskrimineerimisvastane õigus ei takista

liikmesriikidel sätestada rangemaid eeskirju inimeste kaitseks diskrimineerimise eest, on Eesti seadusandja vaba kaaluma erinevaid õiguslikke võimalusi, et ära hoida diskrimineerimist tehisintellekti kasutamisel avalikus sektoris.

Näiteks võiks Eesti (i) töötada välja "andmete kvaliteedi" kriteeriumid, et vältida andmekogumite kallutatust algusest peale; (ii) võtta kasutusele meetodid algoritmide läbipaistvuse ja selgitatavuse suurendamiseks, et kodanikud oleksid võimelised ära tundma võimalikku diskrimineerimist, ja/või (iii) otsustada keelata teatavate tehisintellekti süsteemide kasutamine haldusotsuste tegemisel, kui esineb tõsine diskrimineerimise oht ja/või otsustada neid süsteeme mitte kasutada.

Viiendaks, masinõppe/AI toega teenuste kasutuselevõtu juures riskide minimeerimine eeldab läbivat arusaama tekitamist, mida antud meetodid teevad ja lõppkasutaja valmisolekut ja/või suutlikkust masinõppe mudeleid tööprotsessis kriitiliselt lugeda ja kasutada. On oluline anda teenust osutavale teenistujale või masinõpet oma töös kasutavale ametnikule võimalikult hea arusaam, mida masinõppe/AI mudel tehniliselt teha suudab ja konkreetses kasutuskohas täpsemalt teeb. Ehk teisisõnu anda inimesele võimekus masinõpe/AI väljundit kriitiliselt hinnata, vajadusel masina soovitusega mitte nõustuda ja anda talle selleks ka tehnilised abivahendeid hästi disainitud kasutajaliidese näol. Lisaks tekitab see vajaduse lõppkasutajate koolitamiseks masinõppe sisust nende tööprotsessis tagamaks korrektset kasutusviisi.

Tuleks kaaluda ka masinõppe põhiste teenuste juures lõppkasutajale mingil viisil tagasisidestamise andmise kohustust kuna eesmärk on tagada inimese ja masinõppe või tehisintellekti süsteemi koostöö, mitte ühe domineerimine teise üle. Tagasiside lubab tulevikus ka vastavaid mudeleid praktiku kogemuse abil parandada, näiteks uute tunnuste lisamise kaudu mudelisse.

Viimaks oleks vajalik juba teenuse arendusprotsessi etapis läbi mõelda kuidas toimub sellise toega teenuse mõju hindamine (mõju teenuse otsesele väljundile, aga ka teenuse osutamisele asutuse poolt). Teatud juhtudel on suhteliselt vähese lisatööga selline mõju hindamine automatiseeritav. Läbiv ja võimalikult automaatne mõju hindamine lubaks kiirelt tuvastada, kas tugeva tehnoloogilise komponendiga teenus annab soovitud efekti väljundile (näiteks vähendab riske), aga ka kas seda kasutavad teenistujad kasutavad masinõppe liidest viisil, mis on kooskõlas algselt plaanitud disainiga ja eriti, kas esineb ka nn kriitilist mudeli lugemist ja mittenõustumist.

Kuuendaks, juba lisaks ülaltoodud punktidele ilmnesid ka spetsiifilise õppetunnid masinõppe/AI sisuga rakenduste katsetamisest, mis võiks olla olulised tulevikus analoogsete projektide läbiviimisel avalikus sektoris. Täpsemad kirjeldused on leitavad osas 4, lühikokkuvõte mõnest kesksest probleemist on järgmine:

1. Adekvaatse masinõppe või AI mudeli ehitamine on reeglina võimalik kui rakendusasutus saab kasutada andmeid kogu populatsiooni kohta, mille seest tekib tema haldusülesannete objekt. Teisisõnu peaks asutus nägema nii sündmusi kui mittesündmusi - näiteks töötuse sündmuse esinemist ja mitteesinemist; haiguse esinemist ja mitteesinemist, tuleõnnetuse esinemist ja mitteesinemist. See probleem on lahendatav vaid väliste andmete ristkasutusega;
2. Andmete ristkasutus tähendab reeglina aga teisest kasutamist, mille osas on AKI seisukohal, et seda võib teha ameti tegevust praktiliselt toetava teadusuuringu raames, kuid neid andmeid ei tohi kasutada haldustegevuses. Projektis katsetati viise, mis pakuks tehnilisi lahendusi teisese kasutuse probleemile;
3. Teisese kasutuse keelu mõju ulatuseks testiti tulekahju riskiskooride tööpaketis mudeleid koos ja ilma sensitiivsete isikuandmetega ja leiti, et olulist võitu sensitiivne andmestik ei toonud. See on

aga empiiriline asjaolu, mis ei pruugi kehtida laiemalt teiste modelleeritavate väljundite puhul ehk igas olukorras ei ole ilma sensitiivsete andmeteta võimalik empiiriliselt adekvaatset mudelit luua;

4. Võimaliku tehnilise lahendusena eelnevale probleemile testiti tööturu tööpaketis võimalust luua masinõppe mudel ESA keskkonnas, kus on teadusuuringu raames lubatud sensitiivseid andmeid liita ja mudeldada ning selle järel tõsta välja mudelobjekt ilma andmeteta. Sellisel kujul on teenuseid võimalik üles ehitada ESA nn *sandbox'is* mudel luues ja treenides ning teenuse osutamisel küsida isikult tema nõusolekul tema profiili andmed ja arvutada talle teenuse osutamiseks vajalik väljund alles siis. Sisuliselt oleks tegemist teadusuuringu perioodilise uuendamisega, kus ESA täidaks nn testserveri rolli ja teenuse omaniku juures oleks live-server mudelobjektiga;
5. eTervise tööpaketis testiti terviseandmete OMOP CMD kujule standardiseerimise võimalust ja tekkiva andmestiku kasutusvõimalust testplatvormina masinõppe/AI terviseriskide prognoosimise algoritmidele. Tulemused on väga paljulubavad, sest kinnitust sai just soovitud tulemus, kus juba mujal väljatöötatud algoritme saab kiirelt olemasolevate tööriistadega Eesti andmete peal valideerida ja näha milliste riskide hindamisel need täpseid tulemusi annavad (kandidaadi ennetustööks kasutuselevõtuks) ning millistel juhtudel ei ole Eestis võimalik riske täpselt leida ja sellel kujul ennetada. Sellisel kujul on standarditud andmekogu näol loodud tulevikuks algoritmide kiire testimise platvorm, mida saaks tuleviku tervisevaldkonna ennetusteenuste ehitamisel juba kasutada;
6. Masinõppe laiemat kasutuselevõttu populatsioonitaseme ja indiviiditaseme ennetustöös aga takistab kohati selge tellija puudus. Ravil aitav masinõpe on hetkel väga tugevalt reguleeritud, mille tõttu tunduvad Eesti vastavad asutused eelistavat valmislahendusi koos välise osapoole tootjavastutusega, mitte aga kodumaise teadus-arendustegevuse kasutamisega loodud rakendustega kaasnevat vastutust.

8. Kasutatud allikad

Alexopoulos, C., Lachana, Z., Androutsopoulou, A., Diamantopoulou, V., Charalabidis, Y., & Loutsaris, M. A. (2019). How machine learning is changing egovernment. *In Paper presented at the 12th international conference on theory and practice of electronic governance, Melbourne, Australia.* https://dl.acm.org/doi/abs/10.1145/3326365.3326412

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, *36*(2), 358–367. https://doi.org/10.1016/j.giq.2018.10.001

Bailey, D. E.; Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, *30*(2), 100286. https://doi.org/10.1016/j.infoandorg.2019.100286

Bahrepour, M.; Meratnia, N. ; Havinga, P. (2009). Use of AI Techniques for Residential Fire Detection in Wireless Sensor Networks. *Proceedings of the Workshops of the 5th IFIP Conference on Artificial Intelligence Applications and Innovations*. Lk. 311-321.

Balue´ Arcoverde, G. ; Hansen, M. ; Maeda, E. ; Formaggio, A. ; Shimabukuro, Y. (2009). Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation*, 11: 265–272.

Balzer, L. ; Ahern, J. ; Galea, S. ; van der Laan, M. (2016). Estimating Effects with Rare Outcomes and High Dimensional Covariates: Knowledge is Power" Epidemiologic Methods, vol. 5, no. 1, 2016, pp. 1-18. https://doi.org/10.1515/em-2014-0020

Bovens, M.,Zouridis, S. (2002). From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control. *Public Administration Review,* Vol. 62, No. 2, 174-184.

Bratteteig, T., Wagner, I. (2016). Unpacking the Notion of Participation in Participatory Design. *Computer Supported Cooperative Work, 25,* 425–475.

Brhel, M., Meth, H., Maedche, A., Werder, K. (2015). Exploring principles of user-centered agile software development: A literature review. *Information and Software Technology, 61,* 163–181.

Buffat, A. (2015) Street-Level Bureaucracy and E-Government. *Public Management Review, 17(1),* 149-161.

Bullock, J. (2019). Artificial Intelligence, Discretion, and Bureaucracy. *American Review of Public Administration, Vol. 49(7),* 751–761.

Bullock, J., Young, M., Wang, Y.-F. (2020). Artificial intelligence, bureaucratic form, and discretion in public service. *Information Polity, 25,* 491–506.

Ceyhan, E. ; Ertugay, K. ; Duzgun, S. (2013). Exploratory and inferential methods for spatio-temporal analysis of residential fire clustering in urban areas. *Fire Safety Journal*, 58: 226-239.

Chen, Y.-C., Lee, J. (2018) Collaborative data networks for public service: governance, management, and performance. Public Management Review, 20(5), 672-690. https://doi.org/10.1080/14719037.2017.1305691

Chen, T., Ran, L., Gao, X. (2019). AI innovation for advancing public service: The case of China's first Administrative Approval Bureau. In Proceedings of dg.o 2019: 20th Annual International Conference on Digital Government Research (dg.o 2019), June 18 20, 2019, Dubai, United Arab Emirates. ACM, New York, NY, USA. https://doi.org/10.1145/3325112.3325243

Chhetri, P. ; Corcoran, J. ; Ahmad, S. ; KC, K. (2018). Examining spatio-temporal patterns, drivers and trends of residential fires in South East Queensland, Australia. *Disaster Prevention and Management*, https://doi.org/10.1108/DPM-09-2017-0213

Clare, J. ; Townsley, M. ; Birks, D.J. ; Garis, L. (2017). Patterns of police, fire, and ambulance calls-forservice: scanning the spatio-temporal intersection of emergency service problems. *Policing: A Journal of Policy and Practice*, available at: https://doi.org/10.1093/police/pax038

Digiühiskonna arengukava 2030. (2021). https://mkm.ee/sites/default/files/mkm_arengukava_digiuhiskond_26-10-2021.pdf

Degeling, M., Berendt, B. What is wrong about Robocops as consultants? A technology-centric critique of predictive policing. *AI & Society* **33,** 347–356 (2018). https://doi.org/10.1007/s00146-017-0730-7

Egbert, S.; Krasmann, S. (2020) Predictive policing: not yet, but soon preemptive?, Policing and Society, 30:8, 905-919, DOI: 10.1080/10439463.2019.1611821

Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*.

Fountain, J. (2001). *Building the Virtual State: Information Technology and Institutional Change.* Brookings Institution Press.

Garg, S. ; Aryal, J. ; Wanga, H. ; Shahe, T. ; Kecskemeti, G. ; Ranjan, R. (2018). Cloud computing based bushfire prediction for cyber–physical emergency applications. *Future Generation Computer Systems* 79: 354–363.

Gilboa, I. (2009). Theory of Decision Under Uncertainity. Cambridge University Press.

Green, B. (2019). The Smart Enough City. In *The Smart Enough City*. https://doi.org/10.7551/mitpress/11555.001.0001

Hand, L., Catlaw, T. (2019). Accomplishing the Public Encounter: A Case for Ethnomethodology in Public Administration Research. *Perspectives on Public Management and Governance,* 125–137.

Høybye-Mortensen, M. (2015). Decision-Making Tools and Their Influence on Caseworkers' Room for Discretion. *The British Journal of Social Work, 45(2),* 600-615.

Hu, L. ; Gu, C.(2021). Estimation of causal effects of multiple treatments in healthcare database studies with rare outcomes. *Health Serv Outcomes Res Method* 21, 287–308 (2021). https://doi.org/10.1007/s10742-020-00234-4

Häelm, I. (2021). Refraktiivsete nägemishäirete ja nendega koosesinevate oftalmoloogiliste haiguste analüüs Eesti Haigekassa raviarvete alusel. TÜ magistritöö 2021. https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=72289&year=2021

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 147–169. https://doi.org/10.1080/10919392.2015.1125187

James, G.; Witten, D.; Hastie, T.; Tibshirani, R. (2013) An Introduction to Statistical Learning. Springer Nature.

Jaton, F. (2021). The Constitution of Algorithms. Ground-Truthing, Programming, Formulating. *MIT Press.* https://doi.org/10.7551/mitpress/12517.001.0001

Juhised määruse „Teenuste korraldamise ja teabehalduse alused“ rakendajatele. (2019). https://www.mkm.ee/sites/default/files/content-editors/lyhijuhised_tkta_rakendajatele_vers_1_1.pdf

Jorna, F., Wagenaar, P. (2007). The 'Iron Cage ' strengthened? Discretion and digital discipline. *Public Administration, 85(1),* 189–214.

Jäe, R. (2021). Ravimite täiendavate riskivähendamise meetmete rakendamise analüüsi automatiseerimine. TÜ magistritöö 2021. https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=72375&year=2021

Kankanhalli, A., Charalabidis, Y., & Mellouli, S. (2019). IoT and AI for Smart Government: A Research Agenda. *Government Information Quarterly*, *36*(2), 304–309. https://doi.org/10.1016/j.giq.2019.02.003

KC, K.; Corcoran, J. (2017). Modelling residential fire incident response times: a spatial analytic approach. *Applied Geography*, 84: 64-74.

Kettl, D. (2016). Making Data Speak: Lessons for Using Numbers for Solving Public Policy Puzzles. *Governance: An International Journal of Policy, Administration, and Institutions, Vol. 29, No. 4,* 573–579. https://doi.org/10.1111/gove.12211

Kitchin, R. (2020). Civil liberties or public health, or civil liberties and public health? Using surveillance technologies to tackle the spread of COVID-19. *Space and Polity, 24(3),* 362-381. DOI: 10.1080/13562576.2020.1770587

Kolkman, D. (2020). The usefulness of algorithmic models in policy making. *Government Information Quarterly*, *37*(3), 101488. https://doi.org/10.1016/j.giq.2020.101488

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 101976. https://doi.org/10.1016/j.telpol.2020.101976

Lima, M. S. M., Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly, 37(1).* https://doi.org/10.1016/j.giq.2019.101407

Loi, M.; Rodrigues, M. (2012). A note on the impact evaluation of public policies: the counterfactual analysis. MPRA Paper No. 4244. https://mpra.ub.uni-muenchen.de/42444/

Luht, K. ; Käerdi, H. ; Tammepuu, A. ; Valge, A. ; Kull, T. ; Mumma, A. (2016). Eluruumide tuleohutuse riskihindamise metoodika ja kodukülastuse ankeedi väljatöötamine. Eesti Sisekaitse Akadeemia.

Masso, J; Eamets, R; Philips, K. (2005). Job creation and job destruction in Estonia: labour reallocation and structural changes, IZA Discussion Papers, No. 1707, Institute for the Study of Labor (IZA), Bonn

Masso, J.; Eamets, R.; Philips, K. (2007). Job Flows and Worker Flows in the Baltic States: Labour Reallocation and Structural Changes. In: Paas, T.; Eamets, R: (Ed.). Labour Market Flexibility, Flexicurity and Employment: Lessons of the Baltic States (61−100). . Nova Science Publishers.

Meijer, A. ; Lorenz, L. ; Wessels, M. (2021). Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems. *Public Administration Review.* https://doi.org/10.1111/puar.13391

Mcneely, C. L., & Hahm, J. on. (2014). The big (data) bang: Policy, prospects, and challenges. *Review of Policy Research*, *31*(4), 304–310. https://doi.org/10.1111/ropr.12082

Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government. *Harvard Ash Center Technology & Democracy*, *August*, 19. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

Mehr, H., Ash, H., & Fellow, D. (2017). Artificial intelligence for citizen services and government. In *Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch., no. August*. Ash Center, Harvard Kennedy School.

Meijer, A., Lorenz, L., Wessels, M. (2021). Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems. *Public Administration Review.* https://doi.org/10.1111/puar.13391

Meskó, B. , Gergely Hetényi, G., Győrffy, Z. (2018). Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Services Research, 18(545).* https://doi.org/10.1186/s12913-018-3359-4

Meriküll, J (2011) Labour market mobility during a recession: the case of Estonia. Bank of Estonia, Working Paper Series 1/2011

Mikhaylov, S. J., Esteve, M., Campion, A. (2018). Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. *Phil. Trans. R. Soc. A 376: 20170357.* http://dx.doi.org/10.1098/rsta.2017.0357

Miller, S. M.; Keiser, L. R. (2020). Representative Bureaucracy and Attitudes Toward Automated Decision Making. *Journal Of Public Administration Research And Theory*, 1–16. https://doi.org/10.1093/jopart/muaa019

Miller, T. (2000). Journal of Transport Economics and Policy, Vol 34, No. 2, pp. 169-188.

Morozov, E. (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism* (1st ed.). Public Affairs.

Nawaratne, R. ; Alahakoon, D. ; De Silva, D. ; Chhetri, P. ; Chilamkurti, N. (2018). Self-evolving intelligent algorithms for facilitating data interoperability in IoT environments. *Future Generation Computer System*, 86: 421-432.

Oidersalu, E.; Laube, T. (2018). Eluruumides toimunud tuleohutusalaste nõustamise kokkuvõte 2017. Päästeamet.

Oliver, N., Calvard, T., Potocnik K. (2017). Cognition, Technology, and Organizational Limits: Lessons from the Air France 447 Disaster. *Organization Science, 28(4),* 729–743.

Orlikowski, W. (2000). Using Technology and Constituting Structures; A Practice Lens for Studying Technology in Organizations. *Organization Science, 11(4),* 404-428.

Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity, 25,* 507-522.

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2018). Big Data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration*. https://doi.org/10.1177/0952076718780537

Põldver. M 2021 „Statistilise elu väärtuse kasutamine covid-19 vastu võitlemise meetmete tasuvusanalüüside kontekstis", TÜ magistritöö 2021.

Päästeameti aastaraamat. (2018). https://www.rescue.ee/files/2019-04/1555570808_pa-aastaraamat-2018-final-est-17042019.pdf

Päästeameti aastaraamat. (2019). https://www.rescue.ee/files/2020-04/pa-aastaraamat-2019-est-veeb-rgb.pdf

Päästeameti aastaraamat. (2020). https://www.rescue.ee/files/2021-04/aastaraamat-est-final.pdf

Rohde, D., Corcoran, J., Chhetri, P. (2010). Spatial forecasting of residential urban fires: a Bayesian approach. *Computer, Environment and Urban Systems*, 34: 58-69.

Ronan, T., Teeuw, R. (2016). London's burning: integrating water flow rates and building types into fire risk maps. *International Journal of Emergency Services*, 5: 34-51.

Savaget, P., Chiarini, T., & Evans, S. (2019). Empowering political participation through artificial intelligence. *Science and Public Policy*, *46*(3), 369–380. https://doi.org/10.1093/scipol/scy064

Sun, T. Q., Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly, 36,* 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Sündmusteenuste analüüs. (2020). Lõpparuanne. AS PWC ja Trinidad Wiseman. https://pilv.mkm.ee/s/mZQOshroHT19lBr/download?path=%2F&files=S%C3%BCndmusteenuste%20anal%C3%BC%C3%BCsi%20l%C3%B5pparuanne.pdf

Taylor, M. ; Higgins, E. ; Francis, M.; Lisboa, P. (2011). Managing unintentional dwelling fire risk. *Journal of Risk Research*, 14: 1207-1218.

Thunman, E. ; Ekström, M. ; Bruhn, A. (2020). Dealing With Questions of Responsiveness in a Low-Discretion Context: Offers of Assistance in Standardized Public Service Encounters. *Administration & Society, Vol. 52(9),* 1333–1361.

Troitzsch, J. (2016). Fires, statistics, ignition sources, and passive fire protection measures. *Journal of Fire Sciences*, 34: 171–198.

Vaarandi, R. (2021). A Stream Clustering Algorithm for Classifying Network IDS Alerts. 2021 IEEE International Conference on Cyber Security and Resilience (CSR).

https://ieeexplore.ieee.org/document/9527926

Veale, M., & Brass, I. (2019). Administration by Algorithm? Public Management meets Public Sector Machine Learning. *Algorithmic Regulation*, 1–30. https://doi.org/10.31235/OSF.IO/MWHNB

Võrk, A.; Leetmaa, R.; Kupts, M.; Kirss, L. (2015). Counterfactual Impact Evaluation (CIE) of Estonian Adult Vocational Training Activity.

Wang, J., Li, S. (2014). Time-clustering Behaviors of Urban Fires. *Procedia Engineering*, 71: 214-219.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector-Applications and Challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

World Bank Group. (2017). Enhanced CareManagement:Improving Health forHigh Need, High RiskPatientsin Estonia. https://www.haigekassa.ee/sites/default/files/enhanced-care-management.pdf

Yadav, P.; Kurup, U.; Shah, M. (2017). Structured Causal Inference for Rare Events: An Industrial Application to Analyze Heating-Cooling Device Failure. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017, pp. 1055-1060, doi: 10.1109/ICMLA.2017.00-10.

Yang, L. ; Brown, M. ; Dawson, C. ; Gell, M. (2006). Neural network and GA approaches for dwelling fire occurrence prediction. *Knowledge-Based Systems,* 19: 213–219.

Young, M., Bullock, J., Lecy, J. (2019). Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration. *Perspectives on Public Management and Governance,* 1–13.

Zouridis, S., van Eck, M., Bovens, M. (2018). Automated Discretion. In: Evans, T., Hupe, P. (eds.) *Discretion and the Quest for Controlled Freedom. Palgrave Macmillan.* 313-329.

Zuiderwijk, A., Chen, Y., Salem, F. (2021). “Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda.” *Government Information Quarterly.* https://doi.org/10.1016/j.giq.2021.101577.

Lisa 1. Andmete koordineerimise vajaduse suurenemine

Jaotusliku erinevuste veaks muutumine

Näiteks - kui andmete omanik on mingil põhjusel hakanud tunnuse haridus kategooriate A ja B, C, D, E, F korral mingil hetkel liitma kategooriad A ja B ei muutu esmapilgul midagi, kuid tegelikkuses tekkib andmestikku sellest ajahetkest teistsugune jaotus, millel võivad olla tunnuse olulisusest tulenevalt mudeli empiiriline väljund suuresti valeks muutuda. See on esmapilgul pisike detail, kuid piltlikustab hästi olukorda, kus tugevalt andmepõhiste tõenäosusmudelite alusel toimivate teenuste puhul tekib juurde üks ühise andmehalduse lisakihi vajadus, mis läheb kaugemale tavalistest automaatkontrollidest ja üksteise teavitamisest baasi muudatuste osas. Ehk teisisõnu ei pruugi sellisel juhul masinõppe mudel isegi mitte katki minna, sest vajalik tunnus esineb edasi, aga kuna ta jaotus on tegelikult muutunud, kandub see automaatselt ka live'i protsessi edasi, tehes mudeli empiiriliselt valeks ilma, et keegi seda märkaks. Sellega on seotud mudeli diagnostika monitoorimine teenuseomaniku poolt, kes peab suutma hinnata, kas muudatused tulenevad alloleva protsessi muudatustest, näiteks muutub inimkäitumine ja masin kohendub sellele dünaamiliselt ja „õpib" uued mustri teatud viiteajaga selgeks või on viga põhjustatud vea tekkest alusandmetesse. Masinõppe teenuste arendamise tellimisel tuleks rõhutatult sisse kirjutada mudelite diagnostika ja automaatkontrollide olulisus, see kergendab teenuse elutsükli haldust oluliselt kui oluline on ka selliste muudatuste raporteerimise või monitoorimise abivahendid. Näiteks võiks kaaluda RIHAs dokumenteeritud infosüsteemide ja andmekogudele ka põhjalikuma andmeväljade väärtuste kirjelduse lisamist.

Andmete genereerumise protsessi erinevuste tagajärjed

Andmeid genereeriv protsess on inimese, ettevõtete ja institutsioonide käitumine, osaliselt selguvad seda käitumist iseloomustavad parameetrid riigi registritesse ja andmestikesse *ex post* teatud viiteajaga. Vahepealne lõtk, kus isiku parameetrid on muutunud hetkel *t*, aga masinõppepõhine teenus langetab otsuse või annab soovituse *t-1* andmekooseisu pealt prognoosides riski ajahetke *t+1,* võib reaalajaliselt toimivate mudelite puhul tekitada olukorra, kus tagantjärgi uuenenud andmete alusel oleks isiku suhtes ajahetkel *t* tehtud teistsugune otsus. See probleem on Eestis erinevate avalike teenuste osutamise puhul tavaline ja on kohati lahendatud viisil, kus inimesele endale on pandud kohustus nendest asjaoludest riiki teavitada vältimaks tema suhtes vale otsuse või võimalike nõuete teket – näiteks erinevad sissetuleku saamise või selle määraga seotud toetuste määramise otsused antud kuul, mis sõltuvad osaliselt Maksu- ja Tolliametisse järgneval kuul laekuva TSD deklaratsioonilt nähtavast tegeliku sissetuleku infost. Reaalajas toimivate masinõppe rakenduste juures võimendub see probleem aga oluliselt, sest andmete ristkasutuse puhul on masinõppe teenuse omaniku jaoks tegemist väljaspool tema haldusala isegenereeruva andmestikuga. Sellised viiteajad ja lüngad andmetes on küll erijuhud, kuid kui nende arv on suur on andmevigade tagajärjeks valed või mittesobivad otsused (vt ka Kettl 2016). Isegenereeruvate andmete tekke ja muutumise protsessi saab teenuse arenduse tellija või omanik arvesse võtta vaid juhul kui tal on detailne info selle kohta. Kui andmestikke on palju, tähendab see vajadust erinevate andmedoonorite äriprotsessidega detailset tutvumist, mis raskendab omakorda selliste teenuste arendamist ja väljatöötamist. Vältimaks olukordi, kus sellise probleemi tulemusena avaldub isikule negatiivne mõju, oleks äärmisel juhul kohati ka vajadus koordineerida ülalmainitud lõtkust tulenevalt isikule pandud teavituse kohustust ka nende asutustega kelle masinõppe/AI toega teenus sõltub reaalajas teise asutuse haldusalas kodanike käitumisest genereeruvatest andmetest. See on aga omakorda lisakoormus kõikidele osapooltele üksikute erijuhtude vältimiseks.

Andmedoonorite parem andmestike metainformatsiooni dokumenteerimine, kus lisaks andmevälja sisu kirjeldusele ja koodide/väärtuste tähendusele on võimalik välja lugeda ka andmete genereerumise protsessi detailid suurendaks oluliselt andmete kvaliteeti teistele võimalikele kasutajatele. See oleks teatud mõttes üldine hüve ja aitaks vältida masinõppe rakenduste kasutamisega järgneda võivad negatiivseid efekte. Loomulikult on parema metainformatsiooni lisamine seotud lisakuludega, mida andmete omanik/haldaja peaks kandma, kuid see vähendab tulevikus koormust ka talle ärajäävate lisapäringute näol.

Lisa 2. Tööturu MVPs rakenduse loomise kirjeldus

Lisa 2.1. Ennetavate teenuste mõju hindamine

Töötust ennetavate meetmete mõjuhindamiseks on kasutatud klassikalist tõenäosuse alusel sobitamist (*propensity score matching*), mis on kombineeritud täpse sobitamisega (*exact matching*). Selleks leitakse igale teenusel osalenud isikule *N* lähimat naabrit - isikut, kes ei ole hinnatavat teenust saanud, kuid on töötanud ajal, mil osalusvaatlus on teenusel osalemist alustanud, ja on muude jälgitavate tunnuste alusel teenusel osalenud isikule võimalikult sarnased. Vaikimise on lähimate naabrite arvuks *N=5*, kuid osalusvaatluste lõikes võib võrdlusvaatluste arv mõnevõrra erineda - see võib olla väiksem, kui rangetest reeglitest tulenevalt ei pruugi alusandmestikus soovitud arvu võrdlusvaatlusi esineda, ja suurem, kui vaadeldavate tunnuste lõikes peaks esinema identseid võrdlusvaatlusi ehk isikuid. Nimekiri sobitamiseks kasutatavatest tunnustest on esitatud raporti lisas.

Mõjuhinnangu **perioodiks on 1 kuni 24 kuud**. Seejuures on hindamise algoritmis võimalik määrata, kas teenuse mõju soovitakse hinnata meetme alguses või lõpust. Vaikimisi teostatakse mõjude hindamine antud meetme algusest.

Teenuse mõju hinnati kahele väljundile: 1) TSD alusel leitud töötamise fakt (0/1) ja 2) keskmine palk eurodes. Töötavale inimesele suunatud teenustest valiti mõju hindamiseks järgmised teenused:

- Tööturukoolitus koolituskaardiga töötavatele
- Koolitustoetus tööandjale muutuste olukorras
- Koolitustoetus tööandjale töötaja eesti keele oskuse arendamiseks
- Koolitustoetus tööandjale töötajate värbamiseks
- Tasemeõppes osalemise toetus
- Kvalifikatsiooni saamise toetamine töötavatele

Mõjuhindamise väljundiks on mikroandmestik, kus igale osalusvaatlusele on lisatud *N=5* võrdlusvaatlust. Väljundandmestik sisaldab nii hinnatavaid väljundeid (töötasu saamist faktina ja töötasu suurust) ning sobitamiseks kasutatud tunnuste väärtuseid. Teenusel osalemise tõenäosusmudeli (antud juhul logit mudel) hindamiseks valitakse arvutusliku efektiivsuse huvides alamhulk võrdlusvaatlustest. Hetkel on ca 16 miljonist osalusvaatlusest (üle kõigi hindamisperioodi kuude) valitud 1 miljon vaatlust. Osalusvaatlustest kaasatakse mudelisse kõik kirjed. Osalemise tõenäosusskoor arvutatakse aga sellegipoolest kõigile osalus- ja võrdlusvaatlustele, st sobitamise aluseks olevat võrdlusvaatluste hulka seeläbi ei vähendada.

Sobitamise järel näidatakse osalus- ja võrdlusvaatluste keskmisi erinevusi vastavalt töötamise tõenäosuses ja keskmises palgas, antud erinevus näitab valitud teenuse põhjuslikku mõju töötamisele TSD alusel ning palgale.

Lisa 2.2. Rakenduse loomine

Järgnevalt antakse ülevaade rakenduse loomise protsessist ning sellise lähenemise nõrkadest ja tugevatest külgedest võimalike tulevikurakenduste ehitamisel.

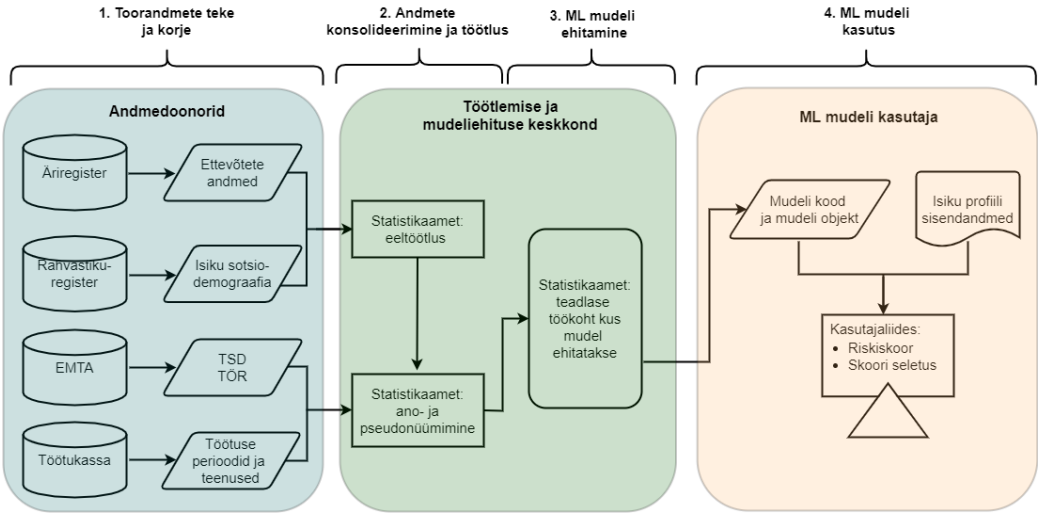
Tehniline protsess

Rakenduse ehitamiseks valiti viis, kus kasutatakse vaid Statistikaametis olemasolevaid andmeid. Selle lähenemise eesmärk oli testida kaht olulist aspekti sääraste rakenduste tulevikus ehitamiseks:

1) kas sel viisil on võimalik kiiremini ehitada rakendust, mille alusandmed sõltuvad erinevate asutuste ning registrite poolt korjatud ja salvestatud andmetest ilma, et peaks eraldi antud asutustest andmeid liidestamise eesmärgil küsima;
2) kas sel viisil on võimalik ehitada rakendusi mille väljundi korrektseks modelleerimiseks on vaja näha kogupopulatsiooni mingitest sündmustest – näiteks nii registreeritud töötuid kui ka töötavaid inimesi – olukorras kus rakendust kasutav asutus kas ei tohi või ei saa sellist täielikku andmestikku omada.

Mõlemad probleemid on olemuses andmete juurdepääsupiirangute ja teisese kasutuse keelust tulenevad probleemid. Töötuse riski rakenduse puhul testiti viisi kuidas nendest piirangutest lähtuvalt on siiski võimalik luua rakendust, mis kasutab andmeid mida rakendav asutus ei oma kas üldse või ei oma kogu populatsiooni kohta ning mida korjatakse kodanike kohta muul eesmärgil kui nende tööturu riskide analüüsiks.

Tööturu masinõppe rakenduse loomise protsessi detailid on toodud Joonisel 2. Rakenduses kasutatavad andmed on kahetised, nii otsesed registriandmed, mis tulevad registripidajatelt kui ka juba erinevate registriandmete alusel tekkinud töödeldud andmed, näiteks Statistikaameti ettevõtete statistilised profiilid. Samuti kasutatakse Töötukassa enda andmeid väljundtunnuse esinemise kohta (töötus), aga seda tehakse antud viisil kuna nii saab need liidestada populatsiooniandmetega, mis moodustab väljundtunnuse teise poole (töötamine). Andmed läbivad Statistikaameti endapoolse eeltöötluse ning ja anonüümimise ja pseudonüümimise protsessi, mille järel tehakse tabelid kättesaadavaks Statistikaameti kontrolli all olevas teadlase töökohas.



Joonis 2. Töötuse riski prognoosiva rakenduse toimimise protsessijoonis.

Teadlaste töökohas viiakse läbi järgmised mudeli ehituseks vajalikud tegevused:

1. andmed liidestatakse ja puhastatakse ning kontrollitakse andmete kvaliteeti;
2. viiakse läbi deskriptiivne analüüs;
3. luuakse sünteetiliste tunnuste genereerimise skript;
4. testitakse erinevaid mudeleid, valitakse välja ja treenitakse sobivaim masinõppe mudel antud väljundi jaoks;
5. luuakse kogu protsessi automatiseeriv skript lähtudes lõpuks valitud mudelist ja selle kasutuses olevatest andmekogudest ja tunnustest;
6. salvestatakse mudelobjekt koos vajaliku lisainfoga.

Teadlase töökohal luuakse seega kogu rakenduse masinõppe mudeli sisu ning salvestatakse need osad, mis on vajalikud mudeli rakendamiseks väljaspool Statistikaameti turvalist töökohta.

Rakendamise faasis luuakse eraldi kasutajaliides, kas eraldiseisva rakendusena või kliendi infosüsteemi integreeritavate vaadetena. Antud rakendusse tõstetakse Statistikaametist mudelobjekt ilma alusandmeteta või siis nende andmetega, mida mõne mudeli metoodika nõuab, näiteks võib kohati vaja olla võtta kaasa ka osa treeningandmestikust, kuid see sõltub valitud mudelitest ja nende jooksutamiseks loodud teekidest.

Valitud protsessi eelised

Kirjeldatud viisil rakenduse ehitamisel on selgeid eeliseid aga ka miinuseid, esmalt miinustest.

Esiteks on sellisel viisil võimalik luua kogu populatsiooni andmetel põhinevat riskiskoorimise mudelit ilma, et lõppkasutajal oleks juurdepääs populatsiooni andmetele, vajalik on vaid treenitud masinõppe mudeli mudelobjekti integreerimine kasutajaliidesega ja rakendus on tehniliselt valmis profiilidele nende riske arvutama. Olukorras kus modelleeritav väljund ei muutu reaalajas - töötuse risk on isiku profiilis, töökohast ja tööturu olukorrast mõjutatav latentne risk mis ei muutu kiirelt - on vajalik masinõppe mudel

Euroopa Liit
Euroopa Regionaalarengu Fond
Eesti tuleviku heaks
RITA

uuenenud alusandmete peal ümberhinnata näiteks mitte tihedamalt kui korra kvartalis ja siis uuendatud mudelobjekt uuesti rakendusse tõsta. Tegemist on sisuliselt Statistikaameti teadlase töökoha kasutamine eksperimentaalse *sandbox'ina,* kus saab anonümiseeritud või pseudonümiseeritud populatsioonitaseme registriandmetega testida erinevate väljundite prognoosimisega, mida piiravad vaid teadusliku eetika piirid.

Teiseks on sellisel viisil võimalik vältida andmete teisese kasutamise keeldu, mis on antud projektis ilmnenud ühe tõsise takistusena paremate AI/masinõppemudelite ehitamisel. Teiseseid andmeid kasutatakse tehniliselt vaid mudeli ehitamise ja treenimise faasis Statistikaameti teadlase töökohal teadusuuringu faasis. Kuigi selle raames leitakse kogupopulatsioonile vastavad riskid kasutades kõiki andmeid, tehakse seda pseudonümiseeritud andmete peal ning individuaalne riskiskoor ei liigu sellest keskkonnast välja, välja liiguvad vaid populatsiooni kirjeldavad jaotused ning mudelobjekti sisu. Lihtsustatult väljendades kasutatakse spetsiifiliste isikute sissetuleku infot leidmaks sissetuleku ja töötuks jäämise vahelist seost kirjeldav koefitsient või puu struktuur või , mis seejärel see ilma andmeteta Statistikaametist välja tõstetakse.

Kolmandaks on juhul kui kogu vajalik andmekoosseis on Statistikaametis olemas võimalik vältida pikaajalist ja ebakindla tulemusega andmetele juurdepääsu taotlemise protsessi.

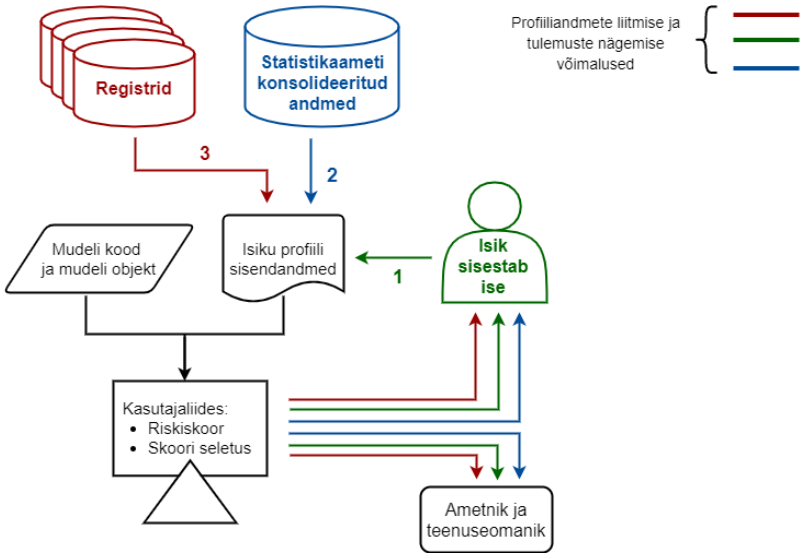
Valitud protsessi miinused

Samas on sellisel rakenduse loomise ja kasutusse võtmise viisil ka selgeid miinuseid.

Esiteks on spetsiifilise inimese riskiskoori leidmiseks vajalik tema profiili pilt rakendusse viia. Viise kuidas seda teha piltlikustab joonis 3. Võimalikke viise on kolm:

1) esiteks (joonisel rohelisena toodud viis nr 1) võib isik vastava info rakendusse ise sisestada ja talle arvutatakse skoor, sellisel kujul võib rakenduse teha vabalt kättesaadavaks veebis ja võimaldada inimestel ennast ise skoorida, pidades muidugi silmas millised positiivseid või negatiivseid efekte see endaga kaasa võib tuua (kõrge riskiga isikute erinevate hirmude tekitamine, mittesoovitud käitumisele suunamine, jne). Isiku enda skoorimise tulemused saab samamoodi kättesaadavaks teha Töötukassa teenuseomanikule või konsultandile kelle ülesanne on isikut vastavalt siis töötuse ennetamiseks mõeldud teenuste juurde suunata. Teatud juhul võib see toimuda juba osaliselt automaatselt iseteeninduskeskkonnas, kuna töötust ennetavatele teenustele kvalifitseerimise tingimused on Töötukassa kodulehel niikuinii vabalt leitavad.
2) teiseks (joonisel sinisena toodud viis nr 2) võib kaaluda isiku nõusolekul tema suhtes vastava päringu tegemisele juba konsolideeritud andmestikust, millel juures peaks olema andmestiku tagasiisikustamise võti. Sel juhul ei peaks isiku uuesti oma andmeid ise esitama ning oleks võimalik pärida ka need sünteetilised tunnused mida inimene ei pruugi ise enda tööajaloo kohta kiirelt mäletada. Hetkel selliseid isikupõhiseid päringuid Statistimaametist teha ei ole võimalik ja see vajaks vastava teenuse loomist, mille võimalikkust/võimatust ei ole antud projektis hinnatud. Oluline on märkida, et andmestikust ei küsita isiku töötuse riski, vaid tema profiili, sest riski arvutamine saab isikustatud kujul toimuda vaid juba tema nõusoleku alusel mudelobjekti info ja profiili abil.

3) kolmandaks (joonisel punasena toodud viis nr 3) on võimalik teatud profiili infot küsida töötava isiku kohta tema nõusolekul X-tee päringuga otse vastavatest registritest, näiteks riigi uues nõusolekuteenuse abil. Selle lähenemise üheks miinuseks on vajadus teha päringuid mitmesse erinevasse registrisse, samas mitme registri kohta on need juba olemas ja Töötukassa neid oma klientide andmete pärimiseks ka kasutab. Küll peab päringutega saadud andmetele rakendama veel ühte, eeldatavasti küll lihtsat, sünteetiliste tunnuste arvutamise skripti, et oleks võimalik tekitada võimalikult sama tunnuste koosseis, mida mudelis kasutati.



Joonis 3. Töötuse riskiskoori rakenduse isiku profiili lisamine skoorimiseks.

Teiseks on vajalik tagada mudeli jaoks vajalike andmete regulaarne laekumine Statistikaametisse ning ka ameti enda majasiseste protsesside automatiseeritus tasemeni kus mudeli ümberhindamise töö oleks tehtav andmeteadlase poolt regulaarselt ilma vajaduseta ameti sees andmekooseise uuesti küsida ja nende pseudo- või anonümiseerimist paluda. Kuna puudub otsene kontroll andmedoonorite üle tekib ka probleem tunnuste tehnilise muutumise osas üle aja kujul mis tekitab vajaduse kas andmete liidestamise ja/või mudeli hindamise skripte muuta. See probleem ei ole väga suur kui mudelit hinnatakse regulaarselt andmeteadlase poolt üle keskkonnas, kus ta saab vastavad muudatused kiirelt ellu viia ja mudelobjekti uuesti luua. Lõpprakendus ise saab töötada ka eelmise mudelobjekti abil ilma, et midagi katki läheks. Juhul kui aga ka isiku profiili küsimine automatiseerida tekib andmedoonorite poolt läbiviidud tunnuste muudatuste monitoorimise ja õigeaegse raporteerimise probleem. Samas on selge, et igasugune andmepõhine AI/masinõppe mudel, mille üheks oluliseks kasutuse legitiimsuse kriteeriumiks on seletatavus, vajab rohkem regulaarset hooldust ja kogu mudeli loomise ahela pidevat kontrolli.

Kolmandaks on sellisel kujul rakenduste loomine hetkel limiteeritud Statistikaameti teadlase töökohtade tehnilise piiratuse tõttu. Töökohad on väikese kõvaketta ja operatiivmälu mahuga ja piirid suurte andmete peal arvutusmahukate masinõppemudelite jooksutamisel tulevad kiirelt ette. Näiteks antud MVP loomisel testitud lasso ridge regressiooni üks mudel jooksis keskeltläbi 1 kuni 2h. Selliseid mudeleid oli iga kuu kohta üks ehk 12 tükki ja nii mõlema mudeldatud väljundi kohta ehk kokku tuli hinnata 24

mudelit. Ka limiteeritud kujul arvutuste paralleelseks viimine tähendas ikkagi ligi 48 tunnist arvutust olukorras kus sama piiratud mälumahtu ja operatiivmälu tahavad kasutada ka teised teadlaste töökohtade kasutajad. Paradoksaalselt ei ole siin tegelikult tegemist sugugi nii suurte andmetega, antud mudeleid hinnati tabelitel milles oli suurusjärgus 25 miljonit rida, mis ei kvalifitseeru kuidagi suureks andmestikuks. Kuid ka selliste andmestike peal andmetöötlusskriptide jooksutamine ja arvustumahukamate mudelite hindamine nõuab kohati oluliselt rohkem mälu, mida ka arvutuste paralleelseks viimine ei lahenda, sest vaja on paremat riistvara. Lisaks on probleemiks ka töökorraldus. Erinevate mudelite ehitamiseks vajalike teekide laadimine toimub RMITi poolt, kellele vahendab seda andmeteadlase palvet edasi Statistikaamet. Kuna sensitiivsete andmetega töötamise puhul puudub võimalus üle VPNi kontrollida kas vastav pakett või teek on installeeritud (mida ta mõnikord ka pärast kinnitust tegelikult ei ole) aeglustub modelleerimise protsess tunduvalt kuna põhiline töövahend selleks on kaasaegne masinõppe ja AI mudelite sobitamiseks vajalik tarkvarapakett. Puudub ka mudelite meeskonnaga arendamiseks väga vajalik koodihalduse süsteem Gitlab või Githubi kujul, mis lubaks analoogselt tarkvaraarendusega mitmel andmeteadlasel korraga mudelite ehitamise kalla tööd teha ning mudeli keerulisi ja üksteisega seotud skriptide efektiivselt hallata ja uuendada.

Lisa 3. AI lahenduste mõju avaldumise kohad

Tabel 1. AI põhiste lahenduste mõju tasandid ja võimalikud positiivsed ja negatiivsed tagajärjed

Mõju tasand	Oodatav kasu	Oodatav kahju
Indiviidi tasand	Ebavajalike rutiinide ja isiklike kokkupuudete vähendamine Uute oskuste ja võimekuste loomine Tähendusrikkamad ülesanded töötajale Personali võimestamine (informeeritum otsusetegemine)	Vähem võimalusi kontekstipõhise informatsiooni kogumiseks Piiratud võimekus reageerida *ad hoc* äärmuslikele juhtumitele Masinotsustusest sõltumine otsusetegemisel Vähenenud otsustusõigus töötajal (töömõtte kadumine) Piiratud ligipääs kriitiliste juhtumite menetlemiseks
Organisatsiooniline tasand	Parem informatsiooni kvaliteet ja ligipääs, suurem/sihipärasem avalike teenuste tarbimine Suurem legitiimsus läbi tunnetatud objektiivsuse suurenemise Lisanduvate kommunikatsioonikanalite võimaldamine Parem rahulolu jälgimine Kiirem ja täpsem sekkumine teenuseosutamisel ja poliitikakujundamisel Täpsem ülevaade ressursside kasutusest nende tõhusamaks jaotuseks Ressursside kokkuhoid	Uued pudelikaelad tööprotsessides ja teenuseosutamises Probleemid vastutuse jaotamise, võtmise ja usalduse kaotusega Probleemid tööriistade väljundite tõlgendustega Võimu liigne tsentraliseerimine Liiane AI-põhiste lahendustele toetumine probleemide ning lahenduste sõnastamisel, organisatsioonilise mälu kadu Lahenduste loomise ja hooldamisega seonduvad kulud ületavad kasu

Ühiskondlik tasand	Varasem ja täpsem ühiskondlike probleemide tuvastus	Töökohtade kaotus
	Võrdne kohtlemine avalike teenuste osutamisel	Andmeõigluse vähenemine ja isikuõiguste riive
	Riik kodanikule lähedamal	Ühiskondlike protsesside ennustamise näilisus
	Personaliseeritud ja proaktiivse teenuse osutamine	Teatud ühiskonnagruppide diskrimineerimine
	Teenusekasutajate aja kokkuhoid	Lahenduste mittetoimivus
		Vastutuse hägustumine
		Avaliku sektori otsustusprotsesside varjatud „erastamine“
		Avaliku sektori legitiimsuse ja usalduse vähenemine

Lisa 4. Õiguslik hinnang

1. Avaliku halduse põhimõtted ja suunised

Põhimõtted ja suuniseid (nn *soft law*) võivad pakkuda mittesiduvaid regulatiivseid suuniseid eetiliste põhimõtete kohta, mida riigiasutused peaksid järgima. Üks selliste võimaluste omapärasid on nende enamasti *mittesiduv iseloom*. Praegu on EL[29] ja mitmed rahvusvahelised organisatsioonid[30] ning mõned riigid[31] kehtestanud tehisintellekti süsteeme käsitlevaid põhimõtteid/suuniseid, mida kohaldatakse ka avalike teenuste suhtes. Majandus-ja Kommunikatsiooniministeerium on avaldanud mitmed juhendid tehisintellekti süsteemide loomiseks.[32] Kuna need suunised keskenduvad siiski peamiselt tehnilistele aspektidele, võiks neid täiustada, lisades teavet eetiliste aspektide ja õiguslike nõuete kohta tehisintellekti süsteemide kasutamisel avalikus halduses ning võtta arvesse Eesti e-riigi hartat, milles on loetletud inimeste õigused suhtlemisel e-riigi ametiasutustega.[33]

Eestil on mitu põhjust kaaluda selliseid tehisintellekti mittesiduvaid lahendusi avalike teenuste valdkonnas. Suunised võivad aidata kohandada Eesti kiiresti arenevat tehisintellekti turgu siseriiklike vajaduste ja tegelikkusega avalikus sektoris. Samal ajal ei loo suunised õiguslikku takistust, vaid on vaid soovitus selle valdkonna reguleerimiseks. Kuigi suunised ja põhimõtted ei saa luua üksikisikutele ja organisatsioonidele rakendatavaid õigusi, annavad nad sisemisi poliitilisi suuniseid algoritmiliste süsteemide arendamiseks ja kasutamiseks. See võib aidata kaasa selliste põhimõtete nagu õiglus, läbipaistvus ja aruandekohustus juurutamisele süsteemi kavandamisel.

Lisaks peaks Eesti valitsus kaaluma nende suuniste õiguslikku staatust, eriti seda, kas suunised - kuigi nad ei ole kodanike poolt jõustatavad - peaksid olema vähemalt haldusasutuste jaoks siduvad. Eesti mittesiduv õigus võiks olla ka eeskujuks tulevaste ELi tasandi õigusaktide jaoks, tuues esile Eesti prioriteedid.

2. Keelud

Üks olulisemaid küsimusi mida tuleb arvestada, kui ML-süsteeme kasutatakse avalikes asutustes on see, mil määral Eesti ühiskond on valmis inimest masinatega asendama: Millal võib inimese otsuse asendada algoritmiga? Milliseid otsuseid peaks igal juhul tegema inimene? Mil määral saab ja peaks haldusotsuseid Eestis automatiseerima?

---

[29] Vt nt High-Level Expert Group on Artificial Intelligence (2019) "Ethics Guidelines for Trustworthy AI".

[30] OECD, "Recommendation of the Council on Artificial Intelligence", OECD/LEGAL/0449 <legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>; OECD (2021), "State of implementation of the OECD AI Principles: Insights from national AI policies", OECD Digital Economy Papers, nr 311, OECD Publishing, Pariis, <https://doi.org/10.1787/1cd40c44-en>.Vt lisaks Ad Hoc Committee on Artificial Intelligence (CAHAI), "Artificial Intelligence in Public Sector,CAHAI-PDG(2021)03 (Provisional)".

[31] Vt näiteks "UK Data Ethics Framework" (2018) <https://www.gov.uk/government/publications/data-ethics-framework>; Canadian Treasury Board Secretariat, "Directive on Automated Decision-Making" (2019) <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>.

[32] Majandus- ja Kommunikatsiooniministeerium, "Juhendmaterjalid" (kratid.ee, 2021) <https://www.kratid.ee/juhendmaterjalid>.

[33] Riigikontroll ja õiguskantsler, "E-riigi harta ehk Igaühe õigused e-riigis" (2018) <https://www.mkm.ee/sites/default/files/content-editors/eng_e-riigi_harta_26.03.2018_lopp.pdf>.

Teine küsimus on seotud ML-algoritmide kasutamisega sotsiaalseks skoorimiseks või järelevalve eesmärgil, nt biomeetrilised identifitseerimissüsteemid (BIS) nagu näotuvastussüsteemid, biomeetrilised kategoriseerimissüsteemid (BKS) ja emotsioonituvastussüsteemid (ETS).[34] Kas selliste süsteemide kasutuselevõtt peaks olema Eestis keelatud? Kas nende süsteemide kasutamine peaks olema piiratud teatud eesmärkidega?

Võrdlusest lähtudes on mõned riigid juba kehtestanud teatavate tehisintellekti süsteemide kasutamise keelud või piirangud. ELis on Itaalia, pidades silmas tehisintellekti süsteemide kavandatud regulatsioone ELi tasandil, kehtestanud õigusliku moratooriumi, mis keelab avaliku ja erasektori asutustel kasutada näotuvastustehnoloogiat avalikes ja avalikkusele avatud kohtades kuni seda küsimust reguleeriva õigusliku raamistiku jõustumiseni, igal juhul aga kuni 31. detsembrini 2023.[35]

ELi tasandil käivad endiselt arutelud teatud tüüpi tehisintellekti süsteemide keelustamise üle. Mitmed organisatsioonid nõuavad "inimõigustega vastuolus oleva" tehisintellekti, sealhulgas eespool nimetatud näotuvastus- ja biomeetriliste süsteemide, keelustamist.[36] Ka Euroopa Parlament on oma mittesiduvas resolutsioonis kutsunud üles piirama näotuvastuse kasutamist teatavatel juhtudel, näiteks keelata politseile sellise tehnoloogia rakendamine avalikes kohtades ja erasektori näotuvastuse andmebaaside kasutamine oma töös, samuti AI kasutamine politsei ennetustöös.[37]

Komisjon on oma ettepanekus AIA (*Artificial Intelligence Act*) kohta teinud ka ettepaneku keelata mõned tehisintellekti süsteemid, nimelt teatavad manipulatiivse tehisintellekti vormid,[38] sotsiaalseks hindamiseks kasutatava tehisintellekti[39] ja biomeetrilised tehisintellekti süsteemid.[40] Siiski jääb ebaselgeks, millist liiki tehnoloogiate keelamine peaks olema keelatud. Lisaks sellele ei selgita AIA,[41] kas ja millises ulatuses võivad liikmesriigid kõrvale kalduda AIA sätetest või kehtestada täiendavaid eeskirju, näiteks keelata muud liiki tehisintellekti süsteeme, mida ei ole nimetatud artiklis 5 AIA või isegi laiendada nimetatud artikli reguleerimisala.[42]

ELi tasandil käimasoleva arutelu tõttu teatavate tehisintellekti süsteemide keelustamise üle, kavandatud artikli 5 AIA ebaselge sõnastuse ja ebamäärase kohaldamisala tõttu peaks Eesti hoiduma täiendavate

[34] Kui BIS tuvastab füüsilised isikud nende biomeetriliste andmete alusel, siis BKS suudab füüsilise isiku liigitada konkreetsetesse kategooriatesse, näiteks soo, vanuse, etnilise päritolu, seksuaalse või poliitilise sättumuse alusel. ETS püüab omakorda tuvastada erinevaid emotsioone, integreerides näoilmete, kehaliigutuste, žestide ja kõne teavet.

[35] "Facial recognition prohibited in Italy: but only for now and with exceptions" (2021) *Italy24News* <https://www-italy24news-com.cdn.ampproject.org/c/s/www.italy24news.com/News/amp/286487>.

[36] "US Corporations Are Talking about Bans for AI. Will the EU??" (2020) *EUROACTIV* <https://www.euractiv.com/section/digital/opinion/us-corporations-are-talking-about-bans-for-ai-will-the-eu/?utm_source=EURACTIV&utm_campaign=f178d07009-RSS_EMAIL_EN_Digital&utm_medium=email&utm_term=0_c59e2fd7a9-f178d07009-116292171>.

[37] Euroopa Parlamendi 6. oktoobri 2021. aasta resolutsioon 2020/2016(INI) tehisintellekti kohta kriminaalõiguses ja selle kasutamise kohta politsei- ja õigusasutuste poolt kriminaalasjades.

[38] Art. 5(1)(c) AIA.

[39] Art. 5(1)(a) ja (b) AIA.

[40] Art. 5(1)(d) AIA.

[41] Martin Ebers jt, "The European Commission's Proposal for an Artificial Intelligence Act - A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)", Journal "J" 2021, 4, 589-603, 591-593, <https://doi.org/10.3390/j4040043>.

[42] Martin Ebers jt (n 46).

keeldude ja moratooriumite vastuvõtmisest riiklikul tasandil, sest seadusandja jaoks on oht, et siseriiklikud sätted tuleb Euroopa Komisjoni võimalike uute õigusaktide valguses hiljem kehtetuks või mittekohaldatavaks tunnistada. Lisaks võivad ELi ja liikmesriikide tasandi vastuolulised õigusaktid põhjustada õiguslikku ebaselgust ja tekitada Eesti seadusandjale täiendavat töökoormust.

Tehisintellekti süsteemide keelustamise asemel seadusandlikul tasandil võiks Eesti loobuda selliste tehisintellekti süsteemide kasutamisest, mille mõjuhinnang[43] näitab, et süsteem tõenäoliselt rikub põhiõigusi. Võimalikud ohud, mis võivad tuleneda tehisintellekti kasutamisest avaliku halduse poolt, võivad olla väga erineva intensiivsusega. Seetõttu tuleks valitsuse tasandil algatada arutelu võimalike piirangute üle, et tagada ühtsed hindamiskriteeriumid ja nende ühtlane kohaldamine.

3. Nõuded kõrge riskiga tehisintellekti süsteemidele

Euroopa Komisjon järgib oma tehisintellekti seaduse ettepanekus nn riskipõhist lähenemisviisi. Kui "lubamatut riski" kujutavad tehisintellekti süsteemid keelatakse, siis "kõrge riskiga" tehisintellekti süsteemide suhtes kohaldatakse rangeid kohustusi enne nende turuleviimist, kasutuselevõttu või kasutamist. Enamik AIA sätteid käsitleb kõrge riskiga süsteeme, sätestades pakkujate, kasutajate ja muude osalejate kohustused kogu tehisintellekti väärtusahelas, kehtestades eelkõige vastavushindamismenetlused, mida tuleb järgida igat tüüpi kõrge riskiga tehisintellekti süsteemide puhul.

Selle taustal tekib küsimus, kas ka Eesti seadusandja peaks sätestama erireeglid kõrge riskiga tehisintellekti süsteemidele avalikus halduses, nagu algselt oli ette nähtud ka Eesti Justiitsministeeriumi väljatöötamiskavatsuses tehisintellekti süsteemide reguleerimiseks.[44]

Hoolimata sellest, et riskipõhine lähenemine AI-le avalikus sektoris on üldiselt tervitatav, on AIA ettepanekul mitmeid puudusi. Üks neist on kõrge riskiga tehisintellekti süsteemide nõuete väga ebamäärane sõnastus.[45] Lisaks ei ole mõned kohustuslikud nõuded veenvad. Samuti on oluline arvestada, et pärast vastuvõtmist oleks AIA õiguslikust iseloomust tulenevalt AIA-l Eesti õigusaktidele ennetav mõju, mis takistaks Eestil vastu võtta eeskirju, mis on vastuolus AIA ja sellega kehtestatud ühtlustatud standarditega: Liikmesriigid peavad aktsepteerima kõiki kõrge riskiga AI-süsteeme, mis on kooskõlas ühtlustatud standarditega.[46] Neil ei ole neil mitte ainult kohustus jätta kohaldamata kõik siseriiklikud eeskirjad, mis on vastuolus AIA endaga, vaid neil on ka keelatud - ühtse turu läbipaistvuse direktiivi 2015/1535 kohaselt[47] - võtta vastu tehnilisi eeskirju, mis on vastuolus ühtlustatud standarditega. Täiendavate siseriiklikute nõuete kehtestamine toodetele, mis on hõlmatud ühtlustatud standarditega, võib viia isegi rikkumismenetluse algatamiseni liikmesriigi vastu vastavalt artiklile 258 ELTL.[48]

[43] Mõjuhinnangu nõude kohta vt allpool, punktis 8.
[44] "Algoritmiliste süsteemide mõjude reguleerimise väljatöötamise kavatsus ("krati VTK")" 24-25, vt ka eespool p 1.
[45] Ebers jt (n 32) 595-597.
[46] Art. 40 AIA.
[47] Euroopa Parlamendi ja nõukogu direktiiv 2015/1535, 9. september 2015, millega nähakse ette tehnilistest eeskirjadest ning infoühiskonna teenuste eeskirjadest teatamise kord, ELT 2015 L 241/1. Direktiivi kohaselt peavad liikmesriigid teavitama kõikidest siseriiklikest tehniliste normide eelnõudest TRIS-süsteemi, et vältida uute tehniliste tõkete tekkimist. Seetõttu ei ole tõenäoline, et Euroopa Komisjon lubab ühtlustatud standarditega vastuolus olevaid riiklikke tehnilisi eeskirju.
[48] Kohtuasi C-100/13 *Komisjon vs. Saksamaa* ECLI:EU:C:2014:2293.

Eesti jaoks tähendab AIA ennetav mõju, et esialgu on mõistlik hoiduda kõrge riskiga tehisintellekti süsteeme käsitlevate asjakohaste õigusaktide vastuvõtmisest. Selle asemel on soovitatav, et Eesti tagaks oma seisukohtade kuuldavaks tegemise AIA õigusloomeprotsessis ELi tasandil[49] ja võimalusel ka juba Euroopa standardiorganisatsioonide poolt tehtava standardimistöö käigus.

4. Volituste delegeerimise reeglid

Niivõrd, kuivõrd tehisintellekti kasutamine valitsusasutuse poolt oma haldusülesannete täitmisel kujutab endast riigivõimu delegeerimist tekib küsimus, kas Eesti õigus nõuab eraldi õiguslikku alust, mis piiraks valitsusasutuse võimalusi kasutada oma ülesannete täitmisel eri liiki tehisintellekti vahendeid. Nagu õiguslikus uuringus selgitatud[50] on õigusliku aluse olemasolu eriti oluline juhul, kui valitsusasutus kasutab tehisintellekti süsteemi haldusaktide või üksikisikute õigusi ja kohustusi mõjutavate meetmete tegemiseks. Sellisel juhul on tehisintellekti süsteemil iseseisev regulatiivne mõju, mis nõuab seadusandja poolt delegeerimisnormi.

Mõned riigid on sellise delegeerimisnormi juba vastu võtnud. Näiteks võttis Saksamaa 2017. aastal vastu õigusnormi, mille kohaselt "haldusakti võib vastu võtta täielikult automaatselt, kui see on seadusega lubatud ja kui puudub nii kaalutlusõigus kui ka hindamisruum". [51]

Eesti seadusandja kavatseb järgida Saksamaa lähenemisviisi, kuid käesolevate soovituste kirjutamise ajal ei ole ametlikku seaduseelnõud veel avaldatud. Seetõttu ei ole esialgu võimalik anda eelnõu kohta õiguslikku hinnangut ega soovitusi.

5. Läbipaistvusmehhanismid

Mõne tehisintellekti süsteemi musta kasti olemuse tõttu ei pruugi arendajad, kasutajad (siinkohal: riigiasutused) ja - mis kõige tähtsam - üksikisikud olla võimelised mõistma algoritmi otsuste taga peituvaid põhjendusi. Praegu ei näe kehtiv ELi õigus ette õigust saada selgitusi tehisintellektipõhiste otsuste kohta. Sellist õigust selgitustele ei saa veel (üheselt) tuletada ka olemasolevast ELi teisesest õigusest (eelkõige AKÜMist), eraelu puutumatuse põhiõigusest ega muudest Euroopa inimõiguste ja põhivabaduste kaitse konventsiooni ja ELi põhiõiguste raamistiku alusel tunnustatud põhiseaduslikest tagatistest.[52]

---

[49] Eesti on juba oma seisukohas AIA kohta juhtinud tähelepanu sellele, et halduskulude ja muude väljaminekute kasv võib olla komisjoni hinnangust oluliselt suurem; Seletuskiri Vabariigi Valitsuse otsuse juurde EL tehisintellekti määruse eelnõu kohta 8-9, 11-12. Lisaks on Eesti märkinud, et tema arvates ei ole kõik AIAs sätestatud kohustused proportsionaalsed. Nagu seisukohas märgitud, on mõningaid kohustusi nende praeguses sõnastuses raske või võimatu täita, näiteks kohustus omada veatuid ja igakülgseid koolitus-, valideerimis- ja katseandmeid.

[50] RITA uuring, üldosa 4. peatükk.

[51] § 35a Verwaltungsverfahrengsgesetz (edaspidi VwVfG), mis on Saksamaa üldine haldusmenetluse seadus. § 35a VwVfG jõustus 01.01.2017, olles osa seadusemuudatusest "Gesetz zur Modernisierung des Besteuerungsverfahrens" G. v. 18.07.2016 BGBl. I lk 1679 (Nr. 35).

[52] Martin Ebers, "Regulating Explainable AI in the European Union. An Overview of the Current Legal Framework(s)" teoses: An Overview of the Current Legal Framework(s), Liane Colonna/Stanley Greenstein (toim.), Nordic Yearbook of Law and Informatics 2020: Law in the Era of Artificial Intelligence, avaldamise järgus <https://ssrn.com/abstract=3901732>.

Praegu sisaldavad erinevad Eesti õigusaktid üldisi läbipaistvusnõudeid, mis kohalduvad ka avalikus sektoris kasutatavatele tehisintellekti süsteemidele. Üks nendest õigusaktidest on Eesti haldusmenetluse seadus (HMS). HMS kohaselt peab Eesti ametiasutuste poolt väljastatud haldusaktis olema esitatud sellise otsuse õiguslik ja faktiline alus ning kaalutlusotsuse puhul kaalutlused, millest haldusorgan on haldusakti andmisel lähtunud.[53] Haldusakti põhjendus peab võimaldama isikul mõista, miks ja millistel õiguslikel alustel haldusakt on antud. Isik peab saama hinnata, kas tema õigusi on õiguspäraselt piiratud.[54] Lisaks on automatiseeritud haldusaktide andmist reguleerivasse seadusmuudatuse eelnõusse kavandatud säte, mille kohaselt haldusakti adressaati tuleb teavitada haldusakti automatiseeritud kujul andmisest.[55]

Eesti õigusaktid ei sisalda konkreetseid sätteid kodaniku õiguse kohta nõuda selgitusi tehisintellektipõhiste otsuste kohta. Seega võiks Eesti seadusandja läbipaistvuse tagamiseks kaaluda võimalust konkretiseerida olemasolevaid sätteid ja/või kehtestada eraldi selgesõnaline õigus saada *tagantjärele* selgitusi avalikus sektoris kasutatavate tehisintellekti süsteemide kohta. Eelkõige võiks seadusandja nõuda üksikasjalikke selgitusi selle kohta, milline on algoritmilise süsteemi loogika ja miks konkreetne otsus on tehtud. Sarnaselt kavatses Eesti justiitsministeerium oma tehisintellekti reguleerimise ettepanekus näha ette läbipaistvusnõuded kõrge riskiga tehisintellekti süsteemidele, sealhulgas kohustuse selgitada süsteemi aluseks olevat loogikat.[56]

AIA ettepanekus on ette nähtud üleeuroopalise registri loomine kõrge riskiga tehisintellekti süsteemide jaoks ilma, et liikmesriikidel oleks keelatud selliseid algatusi ka riiklikul tasandil kasutusele võtta.[57] Sellise registri loomine võiks aidata parandada läbipaistvust ja algoritmilise vastutuse standardeid. Näiteid avaliku halduse tehisintellekti süsteemide registritest leiab mh Prantsusmaal, Soomes ja Kanadas,[58] samuti kogub Eesti Majandus- ja Kommunikatsiooniministeerium teavet avalikus sektoris kasutatavate tehisintellektide kohta ja on teinud avalikult kättesaadavaks olulisemate tehisintellekti süsteemide nimekirja veebilehel https://www.kratid.ee/kasutuslood . Kehtivad Eesti seadused ei näe ette ei kõrge riskiga süsteemide ega muude tehisintellekti süsteemide õiguslikku registreerimise kohustust.[59] Luues registri vaid kõrgema riski kategooria tehisintellekti süsteemidele tuleb samas näha ette kontroll haldusorgani otsuse üle, mis klassifitseerib süsteemi “kõrgema riski kategooriaks”. Otsuse kontrollitavus osutub keeruliseks eelkõige siis, kui haldusorgan on asunud seisukohale, et süsteem ei kujuta endast kõrgema riskiga süsteemi ja seda ei ole registreeritud.

[53] HMS § 56 lõiked 1-3; RITA uuring, üldosa peatükk 6.2.1.
[54] RKHKo 22.05.2000, 3-3-1-14-00, p 5.
[55] Vt ka eespool, p 5.
[56] "Algoritmiliste süsteemide mõjude reguleerimise väljatöötamise kavatsus ("krati VTK")" 24-25, vt ka eespool p 1.
[57] Art. 60 AIA.
[58] "Algorithmic Accountability for the Public Sector." (2021) Ada Lovelace Institute, AI Now Institute and Open Government Partnership (n 45) 19 <https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/>.
[59] Majandus- ja Kommunikatsiooniministeerium, "Kasutuslood" (kratid.ee, 2021) <https://www.kratid.ee/kasutuslood>.

6. Diskrimineerimisvastaste õigusaktide selgitamine

Kuigi tehisintellekti kasutamine võib (osaliselt) kõrvaldada subjektiivse inimteguri otsuste tegemisest, näitavad mitmed näited, et tehisintellekti süsteemid võivad mitmel viisil säilitada ja isegi süvendada inimeste eelarvamusi.[60]

Kuigi nii ELis kui ka Eestis on olemas mitmed diskrimineerimisvastased õigusaktid, ei ole algoritmilise diskrimineerimise probleem seni saanud piisavalt tähelepanu.[61] Probleemid tekivad muuhulgas seoses diskrimineerimisvastaste õigusaktide (piiratud) kohaldamisalaga, "kaudse diskrimineerimise" ebaselge mõistega tehisintellekti süsteemide poolt juhitud suurandmete analüüsi kontekstis ning seoses raskustega diskrimineerimise tuvastamisel algoritmiliste otsuste tegemisel, mis on tingitud algoritmide tööprotsesside läbipaistmatusest.[62] Problemaatiline on ka puuduv teadlikkus diskrimineerimisvastastest seadustest ning asjaolu, et siseriiklikele kohtutele esitatakse vähe diskrimineerimisjuhtumeid.

Sellest tulenevalt tekib küsimus, kas Eesti peaks vastu võtma täiendavad seadusmuudatused kaitsmaks kodanikke masinõppel põhinevate diskrimineerivate otsuste eest. Kuna ELi diskrimineerimisvastane õigus ei takista liikmesriikidel sätestada rangemaid eeskirju inimeste kaitseks diskrimineerimise eest,[63] on Eesti seadusandja vaba kaaluma erinevaid õiguslikke võimalusi, et ära hoida diskrimineerimist tehisintellekti kasutamisel avalikus sektoris.

Näiteks võiks Eesti (i) töötada välja "andmete kvaliteedi" kriteeriumid, et vältida andmekogumite kallutatust algusest peale; (ii) võtta kasutusele meetodid algoritmide läbipaistvuse ja selgitatavuse suurendamiseks, et kodanikud oleksid võimelised ära tundma võimalikku diskrimineerimist, ja/või (iii) otsustada keelata teatavate tehisintellekti süsteemide kasutamine haldusotsuste tegemisel, kui esineb tõsine diskrimineerimise oht ja/või otsustada neid süsteeme mitte kasutada.

Algoritmilist diskrimineerimist aitab vähendada ka tehisintellekti süsteemide järelevalve tugevdamine, eelkõige andmekaitseasutuse ja võrdõiguslikkuse voliniku rahaliste vahendite suurendamine ja seeläbi diskrimineerimisvastaste nõuete parema täitmise tagamine.[64] Teine võimalus on luua kollektiivne diskrimineerimisvastane jõustamismehhanism ja asjakohane sertifitseerimise/järelevalve süsteem.

---

[60] Vt lähemalt teema kohta: RITA uuring, üldosa peatükk 7.

[61] Vt lähemalt: RITA uuring, üldosa peatükk 7.5. Vrd ka Vrd Janneke Gerards ja Raphaële Xenidis, "Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law", Eriaruanne Euroopa Komisjoni jaoks (Euroopa Liidu Väljaannete Talitus 2021) 9, <https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1>.

[62] Vt täpsemalt RITA uuring, üldosa peatükk 7.5.

[63] Üldjuhul sõnastavad kõik asjaomased ELi direktiivid diskrimineerimisvastase õiguse valdkonnas ainult miinimumnõuded; vt eelkõige Euroopa Liidu direktiiv 2000/43/EÜ, millega rakendatakse võrdse kohtlemise põhimõte sõltumata isikute rassilisest või etnilisest päritolust, raamdirektiiv 2000/78/EÜ, millega kehtestatakse üldine raamistik võrdseks kohtlemiseks töö saamisel ja kutsealale pääsemisel, direktiiv 2004/113/EÜ, meest ja naiste võrdse kohtlemise põhimõtte rakendamise kohta seoses kaupade ja teenuste kättesaadavuse ja pakkumisega ja direktiiv 2006/54/EÜ meeste ja naiste võrdsete võimaluste ja võrdse kohtlemise põhimõtte rakendamise kohta tööhõive ja elukutse küsimustes (uuestisõnastamine). Seetõttu lähevad paljud riigid direktiividest kaugemale. Ülevaate saamiseks vt nt Bell, "Anti-Discrimination Law and the European Union" (OUP, 2002) 145-190.

[64] Frederik Zuiderveen Borgesius, "Discrimination, Artificial Intelligence and Algorithmic Decision-Making" (Council of Europe 2018) 61-65.

Sellised kollektiivsed õiguskaitsevahendid võiksid eelkõige kompenseerida individuaalse õiguskaitse puudusi.[65]

Pehme, kuid suhteliselt kergesti rakendatava ja rahaliselt soodsa meetmena võib soovitada suurendada Eestis teadlikkuse tõstmise meetmeid algoritmilise diskrimineerimise kohta ja pakkuda eetikakoolitusi nii ametnikele, sealhulgas kohtunike, kui ka IT-sektori töötajatele. Kodanikke ja tarbijaid tuleks samuti teavitada diskrimineerimise riskidest.

7. Mõju hindamine

Algoritmiline mõjuhinnang võib aidata tuvastada ja leevendada võimalikke ühiskondlikke riske ja ära hoida kahju enne tehisintellekti süsteemi kasutusele võtmist. Siinkohal võib eeskuju võtta AKÜMist, milles nõutakse andmekaitsealase mõjuhinnangu koostamist, kui tegevus "tõenäoliselt toob kaasa suure riski füüsiliste isikute õigustele ja vabadustele", eriti uute tehnoloogiate kasutamisel.[66] Sellise mõjuhinnangu kehtestamine - koos kohustusega jälgida intelligentsete süsteemide riske nende kasutamise ajal - võiks tugevdada vajalikku dialoogi poliitikakujundajate, riigiasutuste ja ML-algoritme arendavate ettevõtete vahel ja aidata kaasa üldise vastutuskultuuri arengule tehnoloogiatööstuses.[67]

Algoritmilise mõjuhinnangu kehtestamine aitaks Eesti haldusasutustel ennetada tehisintellekti kasutusest tulenevaid ohte. Rakendades seda koos auditeerimisega (vt punkt 9), oleks võimalik jälgida algoritmi ohutust nii enne kui ka pärast selle kasutuselevõttu.

8. Auditid, regulatiivsed kontrollid ja välised, sõltumatud järelevalveasutused

Tehisintellekti süsteemide parema vastutuse tagamiseks võib Eesti kasutusele võtta ka algoritmilise auditeerimise. Algoritmilist auditeerimist võib määratleda praktikate kogumina, mille abil on võimalik "kontrollida konkreetset algoritmilist süsteemi, et mõista selle toimimist ja hinnata seda mõne eelnevalt määratletud normatiivse standardi suhtes".[68] Lisaks tuleks otsustada, kas enesehindamine on piisav või tuleks kehtestada sõltumatu välise järelevalveorgani audit.

Nii auditeerimine kui ka väline järelevalve võivad olla kasulikud mitte ainult algoritmide läbipaistvuse, erapooletuse ja tõhususe tõendamiseks, vaid ka selleks, et suurendada oluliselt üldsuse usaldust selliste süsteemide kasutamise suhtes avalikus sektoris. Alustuseks tuleb märkida, et ELi õigusaktid, sealhulgas kavandatav AIA, ei sisalda üksikasjalikke sätteid auditeerimise ja välise järelevalve kohta ega nende keelamise või piiramise kohta, mis jätab nende kehtestamise liikmesriigi, st ka Eesti seadusandja otsustada. Kuid arvestades, et praegu on Eestis avalikus sektoris kasutusel üle 50 ML-lahenduse, millest osa on kasutusel tundlikes sektorites, tasub kaaluda nii auditi kui ka välise järelevalve süsteemi kavandamist. Justiitsministeerium tegi tehisintellekti reguleerimise väljatöötamiskavatsuses ettepaneku

[65] Kollektiivse hüvitamise eeliste kohta vt Christopher Hodges, "The Reform of Class and Representative Actions in Europe. A New Framework for Collective Redress in Europe", Oxford/Portland, Oregon 2008.
[66] Artikkel 35 lõige 1 AKÜM.
[67] Sellise algoritmilise mõjuhinnangu lisaväärtus võrreldes artikkel 35 AKÜM kohase menetlusega võiks seisneda eelkõige selles, et peale andmekaitse oleks võimalik analüüsida ka muid olulisi aspekte.
[68] ibid 24.

reguleerida tehisintellekti süsteemide riiklikud järelevalvemeetmed ja määrata kaebuste käsitlemise eest vastutav haldusasutus.[69]

Auditeerimisasutuse võiks luua kas eraldi seisvana või olemasolevate asutuste osana. Selle peamised ülesanded võiksid seisneda regulatiivsete kontrollide läbiviimises, et tagada algoritmsüsteemide vastavus kehtiva Eesti ja Euroopa õigusega. Arvestades Eesti praktikaga vältida tõhususe huvides eraldi uute institutsioonide loomist, võiks sellise funktsiooni ühendada algoritmilise auditiga ja määrata näiteks Eesti Majandus- ja Kommunikatsiooniministeeriumile, luues ministeeriumis eraldi osakond või järelevalve eest vastutav asutus. Siiski ei tohiks siinkohal tähelepanuta jätta huvide konflikti ohtu. Probleemseks on peetud ka Eesti Andmekaitse Inspektsioonile määratud erinevaid ülesandeid, kuna inspektsioonile on lisaks isikuandmete kaitsele pandud ka avaliku teabe seaduse alusel ülesanne tagada juurdepääs avaliku ülesande täitmise raames kogutud ja avalikuks kasutamiseks mõeldud teabele. Lisaks sellele ei tohi jätta tähelepanuta, et kui täiendavad ülesanded määratakse ilma vastavate ressursside suurendamiseta, muudab institutsioonide üle koormamine võimatuks ülesannete rahuldava täitmise.[70]

Välise järelevalve osas tasub arvesse võtta teiste riikide kogemusi[71] ja kaaluda üksikasjalikult vastava võimaliku asutuse juhtimissüsteemi, selle ekspertiisi (täidesaatev või nõuandev) ja samuti õigusi ja kohustusi seoses teiste asutustega, mis tegelevad samamoodi masinõppe süsteeme käsitlevate küsimustega avalikus sektoris. Võimalik on ka auditeerimise ja välise järelevalve ühendamine, mis tagaks avalikus sektoris kasutatava tehisintellekti suurema tõhususe ja läbipaistvuse.

9. Lõplikud soovitused individuaalsete MVPde kohta

9.1. MVP1 Tööturg

Töötuse riski hindamiseks on vajalik koostada isiku profiil, mille pinnalt on võimalik järeldada tööstuse riski võimalikkust. Vastavalt Euroopa Liidu andmekaitsemäärusele tuleb Eestis kehtestada sobivad meetmed andmesubjekti õiguste kaitseks ning vabaduste ja õigustatud huvide kaitseks.

Täpsemalt tuleb Eesti õiguses reguleerida järgmised vastutuse aspektid:

- Kasutaja ja tootja vastutuse tase töötuse riski vältimise eest.
- Osaliselt autonoomsete AI-süsteemide vastutuse tase.
- Süsteemi auditeeritavus.
- Turvalisus ja töökindlus.
- Vastupidavus küberrünnakutele.

Konkreetsemad soovitused töötuse riski hindamise MPV-ga seonduvalt:

[69] "Algoritmiliste süsteemide mõjude reguleerimise väljatöötamise kavatsus ("krati VTK")" 26, vt ka eespool p 1.
[70] Vt ka: Tupay Paloma Krõõt ja Mikiver Monika, "Der estnische E-Staat - Zukunftsweisendes Vorbild oder befremdlicher Einzelgänger" (2015) Osteuropa Recht, 9 f.
[71] "Algorithmic Accountability for the Public Sector." (2021) Ada Lovelace Institute, AI Now Institute and Open Government Partnership (n 45) 57-64 <https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/>.

a. Isiku töötuse riski profiili koostamiseks vajalike andmete vastutav töötlemine tuleks anda Töötukassale. Töötukassa saab õiguse juurde pääseda profiili koostamiseks vajalikele andmetele. See eeldaks muudatuse tegemist Tööturuteenuste ja toetuste seaduses andmekogu pidamise osas.
b. Profiili koostamine eeldab GDPR kohaselt konkreetse inimese nõusolekut. Tagamaks isiku nõusoleku autentsust, saaks sellist profiili koostada Töötukassa kohalikus esinduses klienditeenindaja abiga. Eelduslikult tähendab see Töötukassale täiendavaid tehnilisi investeeringuid. Kui isik sisestab vajalikud andmed ise, saab eeldada tema nõusolekut. Kui vastavad andmed võtab Töötukassa muudest registritest, siis saaks isiku profiili luua Töötukassa esinduses tagades niisugusel kujul ka isiku nõusoleku autentsuse. Sellega seotud muudatused tulevad kehtestada tööturuteenuste ja toetuste seadusega.
c. Õigusaktides tuleb täpsustata vastutuse reegleid kui MVP poolt koostatud profiil eksib töötuse riski hindamisel ning sellest tulenevalt ei osutata vajalikke tööturuteenuseid.

9.2. MVP2 Tulekahjud

a. Selleks, et määrata kindlaks, milliseid ettevaatusabinõusid tuleb võtta andmekaitse ja mittediskrimineerimise seisukohast, on vaja täpsemalt määratleda nii tehisintellekti süsteemi adressaadid, kasutajad kui ka juurdepääsu õigused asjaomastele andmetele (Eesti Päästeametil on üldjuhul õigus töödelda vastavaid andmeid oma ülesannete täitmise raames, samas kui teistel kasutajatel võib selline õigus puududa).

b. Selleks, et adekvaatselt hinnata võimalikke mõjusid, mis ei ole esialgu ilmsed, tuleb hinnata vastavust andmekaitse ja mittediskrimineerimise õigusnormidele mitte ainult enne süsteemi kasutuselevõttu, vaid ka selle nende kasutamise käigus (muuhulgas selleks, et hinnata isiku kaudse tuvastamise võimalust, mis on otsustav andmete isikuandmeteks kvalifitseerimisel).

c. MVP 2 õiguslik analüüs näitas, et tihe koostöö tehniliste ja õigusalaste töörühmade vahel on otsustava tähtsusega. Täpne õiguslik analüüs on võimalik vaid tingimusel, et tehnilised üksikasjad on piisavalt teada ja selged. Edasises etapis, kui on ambitsioon arendada selle MVP prototüübi alusel tehisintellekti süsteemi, oleks vaja kaasata analüüsi kolmanda osapoolena Eesti Päästeamet, et hinnata tehniliselt väljatöötatud ja õiguslikult lubatud lahenduse praktilist kasutatavust ning teha ettepanekuid vajalike muudatuste tegemiseks.

9.3.MVP3 E-tervis

Tehisintellekti, sealhulgas süvaõppe integreerimist Eesti tervishoiusüsteemi peetakse paljulubavaks, kuigi sellega kaasnevad erinevad riskid ja väljakutsed. Rahvusvaheline kogemus näitab, et tehisintellekti kasutamine tervishoius võib ohustada patsiente, kui kasutatavate andmete hulk on piiratud ja võib seetõttu põhjustada ravivigu (Topol, 2019; Ross &Swetlitz, 2018). Lisaks võib terviseandmete töötlemine riivata turvalisuse ja privaatsusega seotud õigusi.

Kõnealuse projekti tulemused näitavad, et MVP-ga 3 saab optimeerida personaliseeritud tervishoiuteenuseid Eestis, kui võtta arvesse järgmisi soovitusi:

a. Andmeid on vähe, nad on piiratud ja nõuavad tõendamist. Enne MVP 3 rakendamist patsientide ravis tuleb seda valideerida reaalses kliinilises olustikus statistilise metoodika ja analüüsi abil.

b. Selleks, et MVP 3 osutuks edukaks, peavad andmed olema kättesaadavad vastavalt vajadusele ja kooskõlas artikliga 5 GDPR. Üldiselt peavad terviseandmed olema arstide jaoks kättesaadavad identifitseeritaval viisil ning teadlastele ja tervishoiupoliitika otsuste tegijatele mitteidentifitseeritaval viisil.

c. Praegu säilitatakse terviseandmeid erinevates andmebaasides, mis muudab vajalike terviseandmete töötlemise keeruliseks. Seetõttu on vaja teha muudatusi asjaomastes valitsuse otsustes, mis käsitlevad juurdepääsu, kasutamist ja säilitamist jne.

d. Põhimõtteliselt on riigiasutustel lubatud kasutada automatiseeritud individuaalset otsustamist terviseandmete põhjal ainult siis, kui andmesubjekt on andnud selleks nõusoleku või kui töötlemine on lubatud liikmesriigi õiguse alusel olulise avaliku huvi eesmärgil (artikli 9 lõike 2 punktid a ja g GDPR; artikkel 22 lõige 4 GDPR). Automaatne otsuste tegemine, sealhulgas terviseandmetel põhinev profiilianalüüs ei ole praegu ühegi Eesti õigusaktiga lubatud. Selles osas tuleks kaaluda vastavaid seadusemuudatusi. Vastavalt artiklile 22 lõikele 4 GDPR tuleb kehtestada sobivad meetmed andmesubjekti õiguste ja vabaduste ning õigustatud huvide kaitseks. Lisaks sellele on vastutavad töötlejad kohustatud teostama andmekaitsealaseid mõjuhinnanguid, kui andmetöötlustoimingud kujutavad endast tõenäoliselt suurt ohtu andmesubjektide põhiõigustele ja -vabadustele (artikkel 35 lõige 1 GDPR).

e. Tagamaks, et MVP 3 töötleb terviseandmeid kooskõlas artiklis 5 GDPR sätestatuga, on soovitatav vaadata üle kehtivad kindlustatuse ja kahju hüvitamise regulatsioonid.

9.4. MVP4 Küberjulgeolek

MVP 4 õiguslikuks hindamiseks analüüsiti Eesti avalikel e-teenustel põhinevaid kasutusjuhtumeid ja MVP 4 tehniliste aspektide kirjeldust. Põhirõhk oli võimalike valede väljunditega seotud riskidel (valepositiivsed või valenegatiivsed hoiatused, mis võivad kaasa tuua õiguslikke tagajärgi) ja turvanõuete täitmisel riigi- ja kohalike asutuste infosüsteemide haldamisel.

Me leidsime, et:

- Sõltumata tehisintellekti süsteemi kasutuselevõtu kontekstist (kriminaal- või tsiviilõiguslik kontekst) ei ole selge, kes ja kuidas määrab/hindab turvameetmeid käsitlevate kehtivate õigusnõuete täitmist; tehisintellekti tehnoloogiate käsitlemine ISKE julgeolekusüsteemi rakendamise juhendis on väga vähene.
- MVP4-süsteemide valed väljundid võivad põhjustada tagajärgi, mis riivavad kasutajate põhiõigusi. See näib siiski sõltuvat vastava infosüsteemi funktsioonist (funktsioonidest), milles MVP4 süsteemi rakendatakse.

Soovitused:

- Määratleda selged protseduurireeglid ja tagada konkreetsete suuniste vastuvõtmine tehisintellektile toetuvate süsteemide turvalisuse hindamiseks, järgides riskipõhist lähenemisviisi.
- Tagada suurem läbipaistvus (kasutajatele kuvatavate teadete abil) võimalike meetmete kohta, mida võidakse võtta vastuseks hoiatuste/julgeolekujuhtumite puhul, eelkõige juhul, kui see võib mõjutada

kasutajate põhiõigusi. Tagada, et järelevalve MVP 4 kohaldamisega seotud küsimustes on kättesaadav ja tõhus.