

ALGORITMID MUSEAALE UURIMAS: MASINÖPPEMUDELI SÄLLI KASUTAMINE OBJEKTIDE SÄILIVUSE PROGNOOSIMISEKS

Kurmo Konsa, Meri Liis Treimann,
Kristiina Piirisild

Muuseumide ja teiste mäluasutuste üks põhifunktsioone on objektide ja teabe omandamine ning säilitamine tuleviku tarbeks. Eesti Vabariigi muuseumiseadus käsitleb muuseume kui asutusi, mille “ülesanne on koguda, säilitada, uurida ning vahendada inimese ja tema elukeskkonnaga seotud vaimset ja materiaalselt kultuuripärandit”.¹ Rahvusvaheline Muuseumide Nõukogu (International Council of Museums) pakkus 2019. aastal välja uue definitsiooni, tõi küll, selle ümber käivad veel tõsised arutelud. Ka selles rõhutatakse, et “muuseumid talletavad ühiskonna usaldusalustena inimtekkelisi ja looduslikke objekte.”² Mõlemad määratlused rõhutavad muuseumide rolli nii inimtekkeliste kui ka looduslike objektide säilitamisel. See eristab muuseume teistest mäluasutustest, mis koguvad näiteks raamatuid, dokumente, elusolendeid jms. Objekte on kogutud nii nende ilu kui ka väärtuse pärast, samuti selle pärast, mida nad tähendavad, olles

¹ “Muuseum on ühiskonna ja selle arengu teenistuses alaliselt tegutsev kultuuri- ja haridusasutus, mis ei taotle majanduslikku kasumit, mis on üldsusele avatud ning mille ülesanne on koguda, säilitada, uurida ning vahendada inimese ja tema elukeskkonnaga seotud vaimset ja materiaalselt kultuuripärandit hariduslikel, teaduslikel ja elamuslikel eesmärkidel. Muuseum arvestab oma ülesannete täitmisel muu hulgas laste ja puuetega inimeste vajadusi.” “Muuseumiseadus”, Riigi Teataja, redaktsiooni jõustumise kuupäev 01.05.2019. <https://www.riigiteataja.ee/akt/110072013001?leiaKehtiv>.

² “Muuseumid on demokraatiat edendavad mitmehäälised ruumid, mis võimaldavad kriitilist arutelu mineviku ja tuleviku käsitluste üle. Muuseumid talletavad ühiskonna usaldusalustena inimtekkelisi ja looduslikke objekte. Tänapäeva konflikte ja väljakutseid teadvustades ning nendega tegeledes, säilitavad muuseumid erinevaid mälestusi tulevaste põlvetele ning tagavad kõigile võrdse ligipääsu ja õigused pärandile. Muuseumid ei ole suunatud kasumile. Oma tegevuses on nad kaasavad ja läbipaistvad ning töötavad aktiivselt koos erinevate kogukondadega ja nende heaks, et koguda, säilitada, uurida, tõlgendada, eksponeerida ning seeläbi täiustada maailmast arusaamist. Muuseumide eesmärk on toetada inimväarikust ja sotsiaalset õiglust, globaalset võrdsust ja planeedi heaolu.” Agnes Aljas. Uus muuseumidefinitsioon. ICOM Estonia. http://www.icomeesti.ee/konverentsid/2019/muuseumidefinitsioon#_ftn2.

seotud mineviku ja tänapäeva sündmuste, tehnoloogia arengu ning looduskeskkonnaga. Muuseum ei koosne aga ainult objektidest, vaid koon- dab ka inimesi, kes loovad museaalide põhjal huvitavat, kaasakiskuvat ja meelelahutust pakkuvat informatsiooni.

Objektide säilitamine sõltub neile omistatud väärtustest ja tähendus- test. Säilitamise kõige laiem eesmärk on hoida alles objektide, nähtuste ja paikade väärtusi, mis teevad neist kultuuripärandi. Kindlasti tähendab see objektide struktuurse info säilitamist, aga näiteks hoonete korral võib see tähendada ka neid ümbritseva keskkonna, kindla kasutusviisi või nendega seotud ideede alalhoidmist jms. Kõik sõltub ju sellest, millised on need väärtused, mida säilitatakse ja millise informatsiooniga on need seotud.

Artiklis antakse ülevaade masinõppivast otsustusmudelidest Sälli, mis on loodud museaalide kahjustumise prognoosimiseks. Tegemist on tea- daolevalt esimese sellelaadse otsustusmudeliga. Esmalt on aga vaja käsit- leda museaalide kahjustumist, nende kirjeldamise viise ning kahjustuste prognoosimise võimalusi. Kuna masinõppemudel kasutab muuseumide infosüsteemis (MuIS) olevaid andmeid, siis anname ülevaate ka nendest. Otsustusmudelidest parema arusaamise huvides käsitleme selliste mudelite loomise ja hindamise põhimõtteid. Lõpuks kirjeldame loodud prototüübi kasutajate tagasisidet ja anname hinnangu otsustusmudelite kasutusvõi- malustele museaalide vananemise prognoosimisel.

Museaalide kahjustused ja nende uurimismeetodid

Kõik objektid maailmas muutuvad, kuna nendes toimuvad väga erine- vad protsessid. Museaalid ei ole siinkohal mingi erand. See, kui edukalt õnnestub objekte säilitada, sõltub ennekõike nende seisundist ja võima- likest kahjustustest. Esmapilgul võib tunduda, et mingi eseme või ka näi- teks hoone seisund on n-õ objektiivne, loodusteaduslikele analüüsidele alluv olek. Tegelikkus on siiski natuke teine. Seisundi määramine eeldab tervet rida otsuseid, mis kõik on mõjutatud kontekstist ja seotud objekti funktsiooni, eesmärgi ning kasutusega.

Objektides toimuvaid protsesse võime grupeerida füüsikalisteks, keemilisteks, mehaanilisteks ja bioloogilisteks.³ Nimetatud protsessid võivad olla põhjustatud nii objektidele endile omastest teguritest, näi- teks keemiline koostis, materjalide struktuur, kui ka välistest teguritest,

³ Dale Paul Kronkright, "Deterioration of Artifacts Made from Plant Materials", *The Conservation of Artifacts Made from Plant Materials*, toim Mary-Lou E. Florian, Dale Paul Kronkright, Ruth E. Norton (Malibu: The J. Paul Getty Trust, 1990), 139–193.

näiteks temperatuur, õhuniiskus, biokahjustajad, kasutamine, vandalism, õnnetused. Nende protsesside mõjul toimuvad objektides mitmesugused muutused, mis võivad väljenduda nii objektide välimuses (tumenemine, pleekimine, deformatsioonid jms), materjalide füüsikalistes või mehaanilistes omadustes (kaalumuutus, painduvuse kadu, läbipaistvuse muutused jms) kui ka nende keemilises koostises (paberi muutmine happelisemaks, metallide korrodeerumine, naha happeline hüdroloüüs jms).

Muutuste kirjeldamiseks kasutatakse suurt hulka erinevaid termineid, näiteks plekid, määrdumine, hallitus, rebendid, augud, korrosioon jms. Teatud muutusi kutsutakse kahjustusteks. Üldjuhul mõistetakse kahjustuste all selliseid muutusi objektis, mida kindlas kontekstis loetakse ebasoovitavateks.⁴ Muutused on iseenesest tingitud objektiivsetest protsessidest, kahjustuse korral on aga tegemist subjektiivse hinnanguga nende protsesside tekitatud muutustele objektis. Kas mingisugust objekti muutust loetakse kahjustuseks või mitte, sõltub lisaks objektis toimunud muutustele ka objekti väärtustest. Seega ei loeta kõiki objektide muutusi kahjustusteks. Näiteks paatina võib hoopiski suurendada objekti esteetilist või ajaloolist väärtust.⁵

Vaatamata näilisele lihtsusele ja arusaadavusele on inimese loodud objektide, artefaktide kahjustumine kompleksne ja keeruline valdkond. Nagu eelnevalt nägime, on kahjustusprotsessid ise küll looduslikud, kuid nende mõjul objektides toimunud muutuste hindamine (kas objekt on kahjustunud, kas me peame seda konserveerima jne) sõltub sotsiaalsest kontekstist ja kultuuris omaksvõetud väärtustest.⁶ Kõrgema väärtusega objektide juures loetakse ka väiksemad negatiivsed muutused kahjustusteks. See ilmneb hästi üksikute üliväärtuslike objektide, näiteks Ameerika Ühendriikide Iseseisvusdeklaratsioon või Leonardo da Vinci maal “Mona Lisa”, säilitamisel.⁷ Samuti ei ole ülaltoodud kahjustusprotsesside mõju

4 Kurmo Konsa, Anu Lepp, Tiiu Reimo, “Books as Physical Objects: Damage Atlas and Condition Calculator”, *The International Journal of the Book*, 9:4 (2012), 111–123.

5 Vt nt Cesare Brandi, “The Cleaning of Pictures in Relation to Patina, Varnish, and Glazes”, *Burlington Magazine*, 91:556 (1949), 183–189.

6 Vt nt Jonathan Ashley-Smith, *Risk Assessment for Object Conservation* (Oxford, Auckland: Butterworth-Heinemann, 1999).

7 Vt nt Mary Lynn Ritzenthaler, Catherine Nicholson, “A New Era Begins for the Charters of Freedom”, *Prologue Magazine*, 35:3 (2003). <https://www.archives.gov/publications/prologue/2003/fall/charters-new-era.html>; Lorenzo Riparbelli, Fabrice Brémand, Paolo Dionisi-Vici, Jean-Christophe Dupré, Giacomo Goli, Frank Hesser, Delphine Jullien, Paola Mazzanti, Marco Togni, Elisabeth Ravaud, Luca Uzielli, Joseph Grill, “Studies for the Mona Lisa conservation: the implementation of its panel’s Digital-Twin”, *Material Properties to Timber Structures. ECCOMAS Thematic Conference*, toim T. K. Bader (Växjö, 2019), 102. <https://hal.archives-ouvertes.fr/hal-03053232/document>.

artefaktidele kunagi rangelt piiritletud. Erisugused protsessid toimivad enamasti koos, kahjustades artefaktide materjale ja struktuuri. Samuti võib üks keskkonnategur põhjustada erisugust tüüpi kahjustusi. Näiteks kõrge suhteline õhuniiskus võib põhjustada nii orgaanilistest materjalidest objektide füüsikalisi muutusi (deformatsioonid) kui ka metallide keemilisi kahjustusi (korrosioon). Kahjustusprotsessid on pöördumatud, kord juba toimunud muutusi materjalides või struktuurides ei ole üldjuhul võimalik olematuks teha.

Kahjustusprotsesside esinemist ja kiirust mõjutab hulk tegureid. Materjalisisestest teguritest on äärmiselt olulised artefaktide valmistamiseks kasutatud materjalide keemilis-füüsikalised omadused, kõikvõimalikud lisandained jms. Oluline on nii materjalide valmistamiseks kasutatud tooraine kui ka tehnoloogilised protsessid. Materjalide ajalugu (hoiutingimused minevikus) mõjutab suurel määral nende käitumist käesoleval hetkel. See, kuidas näiteks paber reageerib keskkonna tingimuste muutumisele või millise kiirusega kulgevad vananemisprotsessid, sõltub paberi varasematest hoiutingimustest.⁸ Väliskeskkonna teguritest on olulisemad järgmised:

- temperatuur,
- õhuniiskus,
- valgus,
- saasteained,
- biokahjustajad,
- magnetväljad (olulised magnetsalvestiste korral),
- vibratsioon.

Kahjustusprotsesse võib käsitleda museaalide erinevatel organisatsioonitasemetel, alates molekulaarsest ning lõpetades kogude ja kollektsioonidega. Eri organisatsioonitasemetel toimuvate kahjustumisprotsesside uurimisel kasutatakse erisuguseid meetodeid. Molekulaarsel tasandil toimuvate muutuste käsitlemisel kasutatakse mitmesuguseid füüsikalisi, keemilisi ja bioloogilisi uurimismeetodeid. Materjali tasandi uurimiseks kasutatakse analüütilisi uurimismeetodeid, mis annavad informatsiooni uuritava materjali koostise, valmistamismeetodite ning seisundi kohta. Objekti tasandi käsitlemisel integreeritakse erinevate materjalide uurimisel saadud informatsioon, arvestades materjalide omavahelist mõju ning keskkonnamõjusid. Oluline on iga konkreetse objekti ajaloo võimalikult

8 Eckhard Ströfer-Hua, "Experimental Measurement: Interpreting Extrapolation and Prediction by Accelerated Aging", *Restaurator*, 11:4 (2009), 254–266.

täpne tundmine. Küllaltki töömahukate uuringute eesmärgiks on koguda võimalikult kõikehaarav teave konkreetse üksikobjekti kohta. Kogude ja kollektsioonide uurimismeetodid sõltuvad eelkõige kogu suurusest, väärtusest ning objektide endi iseloomust. Väärtuslike ning väikesemahuliste kogude uurimisel rakendatakse samu meetodeid kui objektide puhul. Suuremahuliste kogude korral on asendamatud statistilised meetodid.⁹

Museaalide ja teiste mäluasutustes hoitavate objektide säilitamise seisukohalt on oluline teada nende vananemise kiirust. Näiteks oleks võimalik luua kiiresti vananevatele objektidele võimalikult sobivad keskkonnatingimused, mis aeglustaksid vananemiskiirust ja pikendaksid sellega objektide eluiga. Samuti võimaldaks see teha kindlaks, millised museaalid kahjustuvad ja kui suures ulatuses. Nendele teadmistele tuginedes saaks kavandada konserveerimis- ja restaureerimistöid. Lisaks lubaks museaalide vanemiskiiruste teadmine hinnata, kuidas mõjutavad konkreetset säilitustegevused museaalide säilivust.

Kuidas on võimalik saada teavet museaalide vananemise kohta? Esiteks on võimalik ennustada objekti eluiga selle materjali vananemiskiiruse järgi. Kahjuks on teadmised enamike materjalide vananemiskiiruse kohta vägagi puudulikud. Samuti ei tea me sageli objekti valmistamisega ega seniseid keskkonnatingimusi, mis on mõjutanud materjalide vanemiskiiruseid. Lisaks on ju enamik museaale valmistatud erinevatest materjalidest, mis võivad vananeda erineva kiirusega ning mõjutada üksteise vananemist ja kahjustumist.

Mõnede materjalide kohta on sellised kahjustusfunktsioonid teada, kuid nende rakendamine konkreetsetele museaalidele on sageli küsitav. Esimese universaalse kahjustusfunktsiooni pakkus välja Donald K. Sebera, tuginedes tselluloosatsetaadi vananemise sõltuvusele temperatuurist ja suhtelisest õhuniiskusest.¹⁰ James M. Reilly ja Douglas W. Nishimura arendasid seda edasi, töötades välja säilitusindeksi (PI) ja kaalutud säilitusindeksi (TWPI).¹¹ Alusandmeid nad aga ei avaldanud ja vastav tarkvara on Kujutise Püsivuse Instituudi (Image Permanence Institute) omanduses.

Kahjustusprotsesside keerukuse tõttu on enamikul juhtudest uuritud üksikute lagunemisprotsesside mehhanisme kontrollitud

⁹ Kurmo Konsa, *Artefaktide säilitamine* (Tartu: Tartu Ülikooli Kirjastus, 2007), 99–111.

¹⁰ Donald K. Sebera, *Isoperm: An Environmental Management Tool* (Washington, DC: The Commission on Preservation and Access, 1994).

¹¹ James M. Reilly, Douglas W. Nishimura, Edward Zinn, *New tools for preservation, assessing long-term environmental effects on library and archives collections* (Washington, DC: The Commission on Preservation and Access, 1995).

keskkonnatingimustes.¹² Tegelikuses koosnevad aga objektid vägagi erinevatest materjalidest ning neid mõjutavad pidevalt mitmesugused keskkonnatingimused.¹³ Kasutatud materjalid on oma koostiselt ja struktuurilt väga heterogeensed.¹⁴ Kahjustusprotsessid ise on kompleksed ning teadmised nende kineetikast on esialgu veel puudulikud.¹⁵ Seega saab materjalide vanandamisel saadud tulemusi kasutada vaid üksikute objektigruppide (nt filmimaterjalid) korral.

Teiseks on võimalik kasutada vananemise prognoosimisel objektide seisundiuringute andmeid. Objekti seisund on objekti füüsiline olukord mingil kindlal ajahetkel, mis kirjeldab objekti üldist füüsilist olekut, mis on seotud objekti materjalide, konstruktsiooni, struktuuri, väljanägemise, mõõtmete ja kujuga. Museaalide seisundi kirjeldamiseks on vajalik määratleda objektidel esineda võivad kahjustuste tüübid ja kahjustuste ulatus (kahjustuste kategooriad). See, milliseid kahjustuste tüüpe ja kategooriaid eristatakse, sõltub nii kirjeldatavatest objektidest kui ka kirjeldamise eesmärkidest. On ju selge, et näiteks konserveeritava vanaraamatu detailse kirjeldamise meetodika erineb oluliselt statistiliste kogude uuringus kasutatavast meetodikast. Esimesel juhul pannakse kõik kahjustused võimalikult detailselt kirja, teisel juhul peab meetodika olema üldistatud, lihtne ja rakendatav paljudele objektidele. Väga paljusid kahjustuste kirjeldamise süsteeme iseloomustab küllatki suur subjektiivsus ja ebasüsteemaatilisus.¹⁶ Kasutatakse täpselt määratlemata terminoloogiat. Väga sageli määratletakse kahjustusi neid esilekutsuvate protsesside kaudu, mis loob täiendava eksimise võimaluse. Kirjeldava tasemega süsteemi korral ei ole see

¹² Vt nt Robert L. Feller, *Accelerated aging: photochemical and thermal aspects* (Ann Arbor, MI: Edwards Brothers, 1994); Juan Antonio Fafián Labora, Massimo Lazzari, T. López Morán, “Study of the long term stability of an industrial polyamide from a contemporary artwork”, *Science and Technology for the Conservation of Cultural Heritage*, toim Miguel Ángel Rogerio Candelera, Massimo Lazzari, Emilio Cano (London: Taylor & Francis, 2013), 127–130.

¹³ Jonathan Ashley-Smith, “Climate for Culture: Report on newly gathered knowledge on damage functions” (2015). <https://www.climateforculture.eu/index.php?inhalt=furtherresources.projectresults>.

¹⁴ Vt nt René Larsen, “Introduction to damage and damage assessment of parchment”, *Improved damage assessment of parchment. IDAP: assessment, data collection and sharing of knowledge*, toim René Larsen (Luxembourg: Office for Official Publications of the European Communities, 2007), 17–21.

¹⁵ Agnes W. Brokerhof, “The deterioration processes of organic objects in museum, historic house and archive environments”, EU MASTER Final Workshop, 10.01.2006, UCL, London. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.3289&rep=rep1&type=pdf>.

¹⁶ Joel Taylor, David Watkinson, “Indexing reliability for condition survey data”, *The Conservator*, 30:1 (2007), 49–62.

õigustatud, kuna iga kord ei ole üheselt selge, mis on kahjustuse põhjus, samuti võib ühte tüüpi kahjustusel olla mitmeid põhjuseid.¹⁷

Kogude uuringute käigus on võimalik hankida teavet ka museaalide vananemise dünaamika väljaselgitamiseks. Selleks on vaja uurida kogude seisundit ühesuguse meetodika alusel kahel erineval ajahetkel.¹⁸ Selliseid andmeid on erinevate kogude kohta küllaltki vähe.

Kolmandaks on võimalik kasutada objektide vananemise prognoosimiseks ekspertide koostatud heuristikaid. Need kujutavad endast senistele objektide uurimise kogemustele tuginevaid eeskirju, mis jagavad objektid vananemiskiiruste järgi gruppidesse. Kõige kiiremini vananevate objektide hulka loetakse näiteks puitmassi sisaldav ja happeline paber,¹⁹ tindikorroosioonist kahjustatud dokumendid,²⁰ tselluloosnitraadist ja tselluloosatsestaadist põhimikul fotomaterjalid,²¹ kromogeensed värvifotod,²² pehmenatud polüvinüülkloriid, polüüretaanvahud, vulkaniseeritud looduslik kummi,²³ punamädanikust (*red rot*) kahjustunud nahk,²⁴ rauaühenditega värvitud (must, pruun) tekstiilid jms.²⁵

Kuidas masinõpe saab appi tulla?

Nagu eeltoodust näha, on museaalide vananemise prognoosimine küllaltki keerukas probleem. Tänapäeva seisuga on selleks kõige sobivam objektide seisundi kirjeldamise meetod. Kahjuks puuduvad meil aga enamike museaalide kohta kindla meetodi järgi vähemalt kahel erineval ajahetkel kogutud andmed. Museaalide kohta on aga olemas hulgaliselt andmeid muuseumide infosüsteemis (MuIS) ning nende andmete kasutamist

17 Joel Taylor, David Watkinson, "Using multiple hypotheses in collection condition surveys", *The Conservator*, 27:1 (2003), 13–22.

18 Kurmo Konsa, "Conservation Strategies: A Deterioration Model of Printed Materials", *Papier Restaurierung*, 9:2 (2008), 29–33.

19 Kurmo Konsa, *Arhivaalide ja trükiste säilitamine* (Tartu: Ajalookirjanduse Sihtasutus Kleio, 2008), 50–52.

20 *Ibid.*, 80.

21 Tim Vitale, *Cold Storage of Cultural Artifacts* (2011). https://cool.culturalheritage.org/videopreservation/library/cold_storage_v17j.pdf.

22 Frank P. Simione, "Low temperature storage", *Preventive conservation: collection storage*, toim Lisa Elkin, Christopher A. Norris (New York: Society for the Preservation of Natural History Collections, American Institute for Conservation of Historic and Artistic Works, Smithsonian Institution, The George Washington University Museum Studies Program, 2019), 473–489.

23 Karoliine Korol, *See "igavene" plastmass. Plastesmete vananemine, säilitamine ja konserveerimine* (Tartu: Eesti Rahva Muuseum, 2019), 125–132.

24 Konsa, *Arhivaalide ja trükiste säilitamine*, 64–66.

25 Konsa, *Artefaktide säilitamine*, 197–204.

museaalide vananemise prognoosimiseks hindasimegi projekti Sälli raames.²⁶ Sälli eesmärgiks on luua masinõppemudel, mis võimaldaks tänasega võrreldes efektiivsemalt läbi viia museaalide inventeerimist ja aitaks ennustada museaalide säilimist. Projektis kasutatakse masinõppe meetodeid, mille korral algoritmid õpivad automaatselt, ilma inimesepoolse otsese juhendamiseteta, kasutades selleks olemasolevaid andmeid. Masinõppivad algoritmid loovad valimi (treenimisandmete) alusel mudeli, mida kasutatakse ennustuse või otsustuste tegemisel.²⁷ Masinõppes kasutatakse erinevaid otsustusmudeleid. Nende arendamine algab lihtsamatest ja kiiremini ülesseatavatest mudelitest ning lõpeb keerulisematega. Üks lihtsamaid ja intuiitvsemaid otsustusmudeleid on otsustuspuu (*decision tree*) – vähe ressursse nõudev mudel, mis jõuab otsuseni sisendi kohta jah/ei-küsimusi küsides. Otsustuspuu õpib suhteliselt kiiresti, kuid jääb mõningate probleemide lahendamisel liiga robustseks ega suuda anda kõige täpsemaid tulemusi. Otsustuspuu suur eelis täpsemate konkurentide ees on selle tõlgendatavus. Puumudeli saab üks-ühele visualiseerida graafina, millelt on lihtne välja lugeda, kuidas mudel otsuseni jõudis.

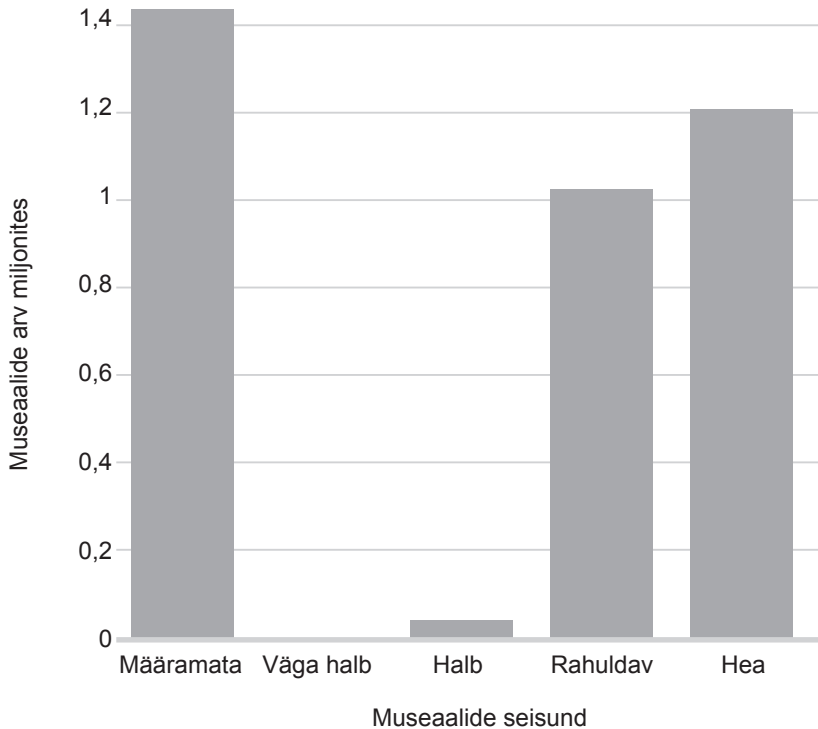
Otsustusmets (*random forest*) ja gradientlaskumine (*gradient boosting*) on ansambelõppemudelid, kus mudeli ennustus pannakse kokku mitme väiksema mudeli ennustustest, näiteks enamushääletuse põhjal. Ansambelõppemudelit on keerulisem tõlgendada kui otsustuspuud, kuid see annab üldiselt paremaid tulemusi kui üksik otsustuspuu. Nii otsustusmetsa kui ka gradientlaskumise algoritmi saab üles seada, kasutades olemasolevaid teke, nagu scikit-learn, XGBoost, LightGBM ja CatBoost. Keerulisemateks otsustusmudeliteks on näiteks tehishäälõppemudelid, mida me selle projekti raames ei kasutanud, kuna tabeli kujul esitatud andmetega ei ole võimalik saada nii häid tulemusi kui eelkirjeldatud lihtsamate mudelitega.

Eduka otsustusmudeli loomise eelduseks on kvaliteetsed ja esinduslikud andmed. Projektis kasutati museaalide seisundimuutuste prognoosimiseks MuISi andmeid. Tegemist on veebipõhise keskkonnaga muuseumikogude üle arvestuse pidamiseks, nende haldamiseks ja muuseumides leiduva informatsiooni kättesaadavaks tegemiseks nii erialaspetsialistidele kui ka kõigile teistele huvilistele.²⁸ MuISiga on liitunud 73 muuseumi ning andmebaasi on kantud ligi 3,7 miljonit museaali 973 muuseumikogust.

²⁶ Tegemist on aastatel 2020–1 läbiviidud Euroopa Liidu Regionaalarengu Fondist rahastatava projektiga, mida juhtis Muinsuskaitseamet. Prototüübi arendus toimus tarkvaraettevõttes STACC (<https://stacc.ee/>).

²⁷ Vt nt Tom Mitchell, *Machine Learning* (New York: McGraw Hill, 1997).

²⁸ Muuseumide infosüsteem MuIS – Muinsuskaitseamet. <https://www.muinsuskaitseamet.ee/et/muuseumide-infosusteem-muis>.



Joonis 1. Museaalide hulk seisundite lõikes viimaste seisundihinnangute põhjal

Kõige enam museaale on MuISi andmetel Eesti Rahva Muuseumi kogudes, kuhu kuulub 18% MuISi kantud museaalidest. Andmebaasis on hulgaliselt teavet nii museaalide, nende konteksti kui ka nendega muuseumides toimuvate tegevuste kohta.

Kuna antud uuringu fookuses on museaalide seisundi muutumine, siis on esmatähtis, kuidas kirjeldatakse MuISis museaalide seisundit. Seda hinnatakse MuISis nelja väärtusega: “hea”, “rahuldav”, “halb”, “väga halb”. Kui seisundit ei määratleta, siis märgitakse väärtuseks “määramata”.

Erinevatele seisunditele on antud järgmised kirjeldused:

“Hea” – objekt on stabiilses seisundis, kasutatav ilma piiranguteta, ei vaja töötlemist. On eksponeeritav.

“Rahuldav” – objekt on rahuldavas seisundis, moonutatud välimusega või kahjustatud, kuid stabiilne. Eksponeerimiseks vajab korrastamist.

“Halb” – objekti seisund on halb, kahjustatud ja/või ebastabiilne, lubatud ainult piiratud kasutamine, vajab töötlemist normaalse seisundi saavutamiseks. Ei ole eksponeeritav.

“Väga halb” – objekti seisund on ülimalt ebastabiilne, struktuur nõrk ja aktiivselt lagunev, mõjutab teisi objekte (nt hallitus, rooste) ja nõuab kohest töötlemist.

Halva või väga halva seisundi korral on vajalik kirjeldada ka objekti kahjustusi.

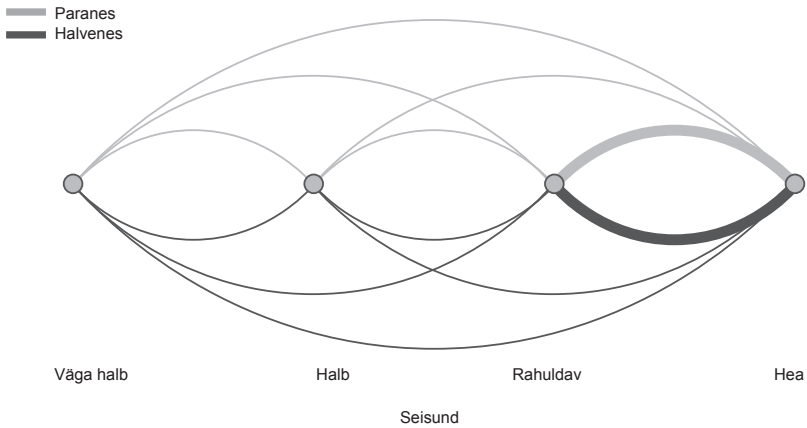
MuSi on kantud peaaegu 3,7 miljonit seisundihinnangut. Vähemalt üks seisundihinnang on MuSis 62% museaalidest, seega rohkem kui kolmandikul museaalidest ei ole kunagi seisundit hinnatud. Viimaste seisundihinnangute põhjal on enamik museaale rahuldavas või heas seisundis (joonis 1).

25% museaalidest on mitu seisundihinnangut, kusjuures üksikutel museaalidel on isegi üle 15 hinnangu ja ühel museaalil on lausa 34 seisundihinnangut. Keskmiselt on museaalide seisundit hinnatud iga 1210 päeva ehk veidi rohkem kui kolme aasta tagant.

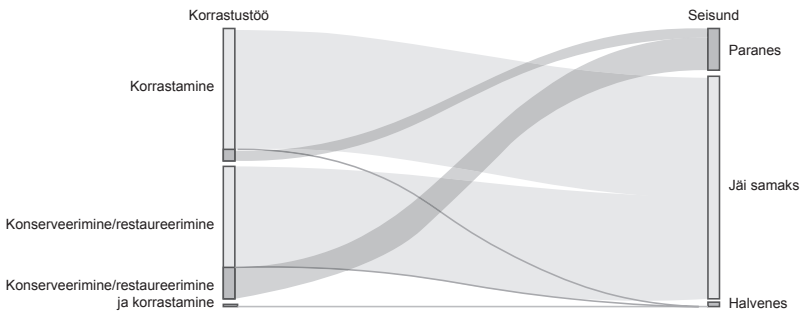
Nende andmete põhjal seisundit prognoosiva mudeli arendamiseks on vaja minimaalselt järjestikuste seisundihinnangute paare, mida saab kasutada tuvastamiseks, kas üks või teine sündmus, museaali omadus või mingi kombinatsioon neist korreleerub seisundi muutumisega. Selliseid paare on mitme seisundihinnanguga museaalide hulgas kokku üle 1,4 miljoni. Nende seast peaaegu 32 000 paari ehk veidi üle 2% koosnevad kahest erinevast seisundihinnangust ehk märgivad seisundimuutust. MuSi kantud andmete põhjal on seisundimuutusi toimunud peaaegu 30 000 museaalil ehk vähem kui 1% kõigist museaalidest (joonis 2).

Nendel museaalidel, mille seisund on muutunud, on see toimunud keskmiselt 2075 päeva ehk peaaegu kuue aasta tagant. Kõige sagedamini on seisund muutunud rahuldavast heaks (peaaegu 14 000 ehk 44% kõigist seisundimuutustest), samas suurusjärgus on toimunud seisundimuutusi vastupidises suunas ehk heast rahuldavaks (üle 10 000 ehk 33% kõigist seisundimuutustest). Vähem on seisundimuutusi teiste seisundipaaride vahel, kuna nendes seisundites ongi suhteliselt vähe museaale.

Enne andmete pealt õppiva mudeli arendama asumist püüdsime selgeks teha, kas MuSis olevatel andmetel üldse on potentsiaali masinõppemudeli arendamiseks. Selleks vaatasime, kuidas see, mida tahame, et mudel ennustaks, korreleerub andmetega, mille põhjal mudel teoreetiliselt ennustuse võiks teha, ehk kas ja kuidas on museaali seisund ja stabiilsus seotud selle omaduste ja sündmustega. Selleks oli meil kasutada 1,4



Joonis 2. Seisundimuutused. Joone paksus märgib muutuste hulka



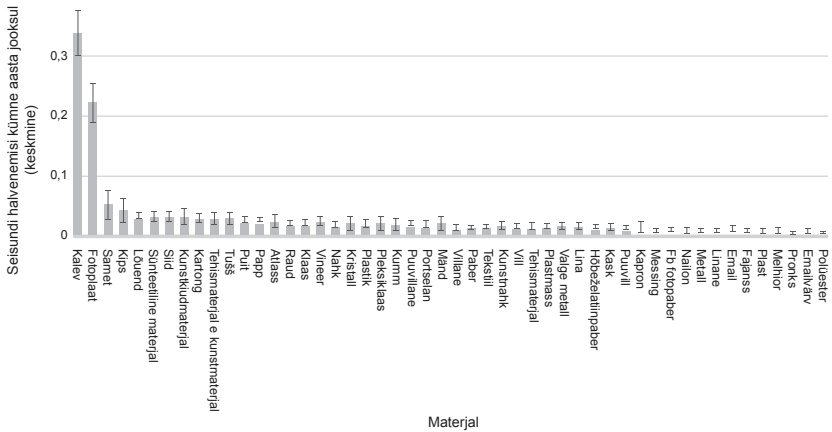
Joonis 3. Seisundi muutumine korrastustööde järel

miljonit järjestikuste seisundihinnangute paari, sh 13 000 seisundi halvenemist. Seisundi paranemist ei ole siinkohal väga oluline käsitleda, kuna museaalide seisukord ei parane iseenesest, vaid see juhtub näiteks konserveerimistööde tulemusena. Analüüsisime esiteks, kuidas muutub seisund erinevate sündmuste (näituste ja korrastustööde) järel, ning teiseks, kas kindlate omadustega museaalid on stabiilsemad kui teised. Kõigist 1,4 miljonist järjestikuste seisundihinnangute paaridest 97,7% seisundihinnang ei muutunud, 1,3% seisundihinnang paranes ja 1% juhtudest halvenes. 1,9% juhtudest saabus museaal enne uue hinnangu saamist näituselt. Näituselt

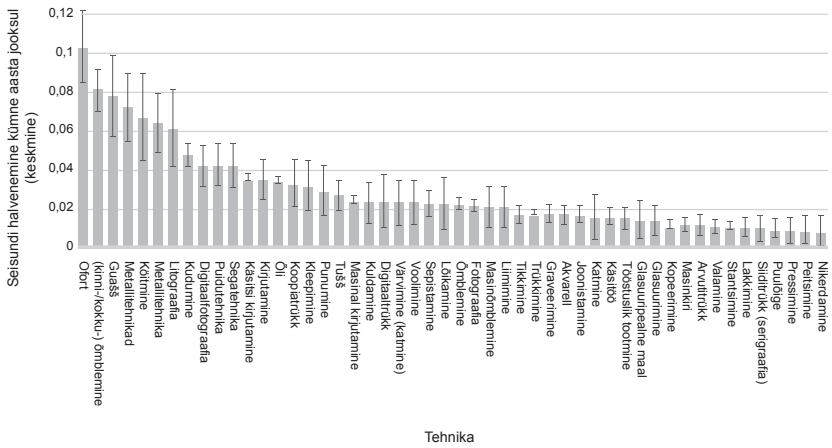
saabumise järel 2,3% juhtudest seisund paranes (vs keskmiselt 1,3%) ja 1,5% juhtudest halvenes (vs keskmiselt 1%) ehk näitusel eksponeerimise järel on keskmisest suurem tõenäosus, et seisund muutub. Näituselt saabunud museaalid võidi enne uut seisundihinnangut korrastada, mille järel nende seisund paranes, või tagastada hoidlasse, kus nende seisund ebasobivate eksponeerimistingimuste tõttu kiiremini halvenes. 0,7% juhtudest saabus museaal enne uue hinnangu saamist mõnelt korrastustöölt. Korrastustöö järel 15,6% juhtudest seisund paranes (vs keskmiselt 1,3%) ja 0,7% juhtudest halvenes (vs keskmiselt 1%) – korrastustöö järel on keskmisest oluliselt suurem tõenäosus, et seisund paraneb ja veidi väiksem tõenäosus, et halveneb (joonis 3). Enne korrastustöö toimumist ei pruugi halvenenud seisund olla MuSiS fikseeritud, mistõttu ei pruugi andmetest välja tulla kõik korrad, kui korrastustöö aitas museaali olukorda parandada. Ootuspäraselt on konserveerimise/restaureerimise mõju seisundile keskmiselt parem kui pelgalt museaali korrastamisel.

Museaalide seisundi püsivuse hindamiseks vaatasime, mitu korda keskmiselt museaali seisund mingi aja jooksul on halvenenud. Museaali seisundit on hinnatud harva (keskmiselt iga kolme aasta tagant) ja muutused on reeglina küllaltki aeglase loomuga. Andmemüra vähendamiseks käsitlesime vaid neid museaale, mille kohta on vähemalt 10 aasta jagu seisundiandmeid (142–242 museaali). Et tulemused oleksid omavahel võrreldavad, skaleerisime seisundi halvenemiste hulga iga museaali kohta 10 aasta peale (kui meil on kaks museaali, mille kohta on vastavalt 10 ja 20 aastat seisundiandmeid ning mõlema seisund on selle aja jooksul ühe korra halvenenud, siis loeme teist stabiilsemaks, kuna selle halvenemine on toimunud pikema perioodi jooksul). Nende 142–242 museaali põhjal analüüsisime, kas stabiilsus sõltub erinevatest omadustest (olemus, materjal, tehnika, dateering). Iga omaduse puhul jätsime välja väärtused, mille kohta oli vähem kui 100 andmepunkti, kuna seisundi halvenemine on harv nähtus ja väheste andmete põhjal ei ole võimalik saada usaldusväärseid tulemusi. Nende piirangutega saime stabiilsuse arvutada 133 olemuse (13%), 74 (13%) materjali ja 82 (5%) tehnika kohta. Samas on tegemist üsna populaarsete väärtustega, kuna neist vähemalt üks (olemus, materjal või tehnika) on fikseeritud 77% kõigist museaalidest.

Joonisel 4 on toodud 50 kõige ebastabiilsemat olemust. Graafikult on näha, et kuigi keskmise määramatust arvestades on paljud olemused sama stabiilsed, on siiski olemusi, mis on selgelt ebastabiilsemad kui teised. Näiteks lint/pael-tüüpi museaali seisund halveneb hinnanguliselt 1,4–3,5 korda kiiremini võrreldes museaalidega, mille olemuseks on märgitud “jakk/pintsak”. Kõige stabiilsemate olemuste hulgas (ilma ühegi seisundi



Joonis 6. Kõige ebastabiilsemad materjalid keskmise seisundi halvenemise sageduse järgi



Joonis 7. Kõige ebastabiilsemad tehnikad keskmise seisundi halvenemise sageduse järgi

keskmiselt sama tempoga. Kuna käsikirja puhul leidis rohkem museaale, mille seisund ei halvenenud, siis võib seda tegelikult lugeda stabiilsemaks kui laualeht/seinaleht-tüüpi museaale.

Joonisel 6 on toodud 50 kõige ebastabiilsemat materjali. Ka siin on näha, et teatud materjalid on tugevalt ebastabiilsemad kui teised (kalev, fotoplaat).

Kõige stabiilsemate materjalide hulgas (ilma ühegi seisundi halvenemiseta) on näiteks nikkel, savi, luu, polüetüleenfotopaber ja metallisulam.

Kui vaadata museaalide valmistamiseks kasutatud tehnikaid (joonis 7), siis tuleb taaskord välja, et mõnes tehnikas valmistatud museaalid on keskmiselt stabiilsemad kui teised. Kõige stabiilsemate tehnikate hulgas (ilma ühegi seisundi halvenemiseta) on näiteks ofsettrükk, tempera, seepia, puugravüür ja käsitsi õmblemine.

Üldiselt leidsime, et MuISi andmed on rahuldava kvaliteediga. Museaalide kirjeid on mõistlikul hulgal, kuid paljude museaalide puhul on mõni väli täitmata. Näiteks pole märgitud rohkem kui poolte museaalide materjali, mis tähendab, et nende säilivust on keeruline adekvaatselt ennustada. Samuti on poolikuks jäänud paljud museaalidega seotud sündmused, näiteks 56% korrastustöödest on nende lõpu- või museaali tagastamise aeg märkimata. Nii seisundihinnangud kui ka muud sündmused on MuISi kantud valdavalt alates MuISi loomisest ehk umbes viimase 10 aasta jooksul, mistõttu ei pruugi seisundimuutuste kui aeglase protsessi modelleerimine hästi välja tulla. Mudel saab anda maksimaalselt nii häid tulemusi, kui hea on sisendandmete kvaliteet. Enne mudelite koostamist tuleb andmed ette valmistada, eemaldada ebavajalikud ja väärad sissekanDED, viia andmed ühtsele ja struktureeritud kujule ning tuletada vajadusel olemasolevate andmete põhjal uusi andmeid.

Otsustusmudelite loomine ja katsetamine

Mudeli treenimiseks kasutatav andmestik kujutab endast suurt tabelit, kus iga rida vastab ühele andmepunktile ja iga veerg ühele tunnusele/omadusele selle andmepunkti kohta. Andmepunktidenä kasutasime iga museaali puhul vähemalt kahte seisundihinnangut, millele lisasime vastava museaali omadused ja muud tunnused, mis aitavad museaali seisundi halvenemist prognoosida. Nende andmete hulka kuulusid museaaliga seotud staatilised andmed:

- muuseum,
- muuseumikogu,
- olemus,
- materjal,
- materjalirühm,
- tehnika,
- eksponeeritavus,
- dateering.

Lisainfona kasutasime museaali ajalugu ehk kokkuvõtet museaaliga seotud sündmustest, arvestades vaid seisundihinnangu andmise ajaks toimunud sündmusi (kuna meil ei ole infot tuleviku kohta):

- mitu päeva on museaal arvel olnud (kuna meil ei ole sellele infole otsest ligipääsu, siis lugesime museaali arvele võtmise ajaks selle esimese seisundihinnangu aja);
- mitu päeva on praegune seisund muutumatuna püsinud;
- kas on näitusel käinud;
- mitu korda on näitusel käinud;
- mitu päeva on näitusel olnud;
- kas on korrastamises (sh konserveerimised/restaureerimised) käinud;
- mitu korda on korrastamises käinud;
- mitu päeva on korrastamises olnud;
- mitu korda on hoidlast välja võetud (näituste ja korrastamiste põhjal);
- mitu päeva on hoidlast väljas olnud (näituste ja korrastamiste põhjal).

Kokku kasutasime 19 tunnust: seisund, 8 museaali omadust ja 10 museaaliga seotud sündmusi kirjeldavat tunnust.

Selleks, et mudeli arendamise ajal selle tulemusi hinnata, jagatakse olemasolevad andmed treening- ja testandmeteks, et imiteerida olukorda, kuhu mudel satub, kui see päriselt kasutusele võetakse, ja et vältida info leket testandmetest treeningandmetesse (*data leakage*). Testandmed jäetakse esialgu kõrvale ja mudelile näidatakse andmetest mustrite õppimiseks ainult treeningandmeid. Seejärel tehakse mudeli põhjal prognoosid testandmetele ja võrreldakse neid tegelikkusega (testandmete kohta on meil see info olemas). See peaks andma üsna hea ülevaate, milliseid tulemusi võib mudelilt oodata, kui see kasutusele võetakse. Testandmete valikul võeti andmete lekke välistamiseks arvesse ka seda, et tegemist on aegri-dadega – mudelit tuleks trennida vanemate seisundihinnangute põhjal ja seejärel hinnata selle osavust “tuleviku” ehk hiljutisemate andmete peal. Seetõttu valisime juhuslikud testandmed vaid museaalide viimaste seisundihinnangute (baasandmesiku loomisel alles jäänute) hulgast, kusjuures testandmestiku suuruseks võeti 20% kogu andmestikust.

Klassifikatsioonimudelite puhul algab mudeli hindamine tavaliselt veamaatriksist (*confusion matrix*, *error matrix*, joonis 8), mis visualiseerib, kui hästi vastavad prognoosid tegelikkusele ja millest on tuletatud erinevaid näitajaid klassifikatsioonimudelite hindamiseks. Ideaalis võiksid kõik tulemused olla helehallilt markeeritud diagonaalil.

Sälli puhul on kõige olulisem, et mudel tuvastaks võimalikult palju halvenevaid museaale, st et oleks võimalikult vähe valenegatiivseid tulemusi.

Prognosis → Tegelikkus ↓	Negatiivne (ei halveneb)	Positiivne (halveneb)	
Negatiivne (ei halveneb)	Tõene negatiivne, TN (ingl <i>true negative</i>)	Valepositiivne, FP (ingl <i>false positive</i>)	
Positiivne (halveneb)	Valenegatiivne, FN (ingl <i>false negative</i>)	Tõene positiivne, TP (ingl <i>true positive</i>)	→ saagis = $\frac{TP}{TP+FN}$
		↓ täpsus = $\frac{TP}{TP+FP}$	

Joonis 8. Veematriks

Selle hindamiseks sobib hästi saagis (*recall*), mis näitab, kui suure osa kõigist positiivsetest tulemustest mudel üles leiab. Näiteks Sälli puhul tähendaks saagis 0,9, et mudel leiab üles 90% halvenevatest museaalidest ja 10% jääb leidmata (valenegatiivsed). Saagise kõrval võetakse teiseks kriteeriumiks harilikult täpsus (*precision*), mis näitab, kui suur osa positiivsetest prognoosidest on ka tegelikult positiivsed. Näiteks Sälli puhul tähendaks täpsus 0,9, et kõigist mudeli poolt halvenevaks prognoositud juhtudest 90% ka tegelikult halvenevad ja 10% ei halvene (valepositiivsed). Nii saagise kui ka täpsuse puhul on suurem väärtus parem. Arvestades halvenevate museaalide hulka ja seda, et oluline on tuvastada võimalikult palju halvenevaid museaale, võime täpsuses veidi järele anda, kui see tähendab suuremat saagist. Leppisime kokku, et täpsus ei või olla madalam kui 0,5 ehk Sälli võib anda valehoiatuse kuni pooltel juhtudel.

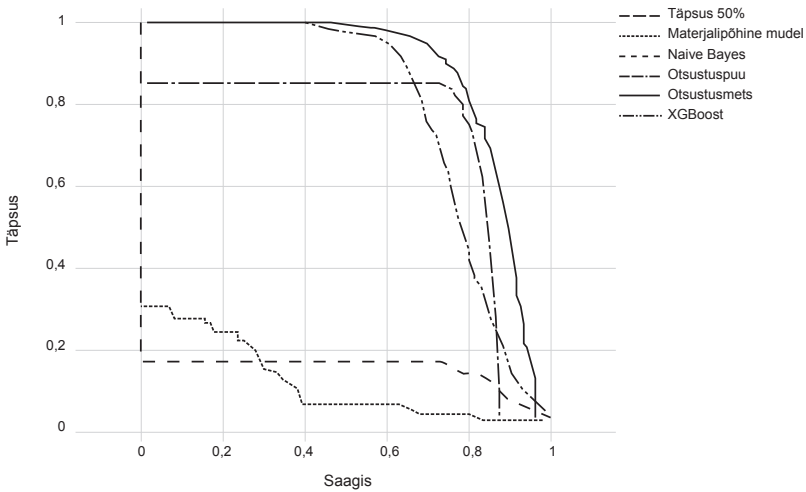
Esimese eksperimendina püüdsime välja selgitada, millise perioodi kohta oleks mõistlik prognoose anda. Eelanalüüsi käigus andmeid uurides sai selgeks, et MuISis ei ole piisavalt pikemaids seisundihinnangu aegridu, millelt oleks võimalik tuvastada, kuidas erinevate museaalide seisundid ajas muutuvad, ja anda detailne prognoos selle kohta, milline on museaali seisund mingil konkreetsel ajahetkel tulevikus. Kuna nii detailne prognoos ei ole ka praktikas oluline, siis vormistasime säilivuse prognoosimise ülesande veidi robustsemaks. Sälli ülesanne on prognoosida, kas museaali seisund halveneb järgmise n aasta jooksul, kus kõige paremaid tulemusi andev n selgitatakse välja eksperimentide käigus. Tegemist on kahe klassiga (“halveneb” ja “ei halvene”) klassifikatsiooniprobleemiga ja selle jaoks saab museaalide aegridade asemel mudeli treenimiseks kasutada järjestikuste seisundihinnangu paare, mida on rohkem.

Kuna MuISi loomisest on möödas veidi üle 10 aasta ja varasemast on sinna kantud vähem andmeid, siis arvestasime, et maksimaalselt saame luua mudeli, mis prognoosib seisundi halvenemist järgmise 10 aasta jooksul. Prognoose ei ole mõistlik anda ka väga pika ajaperioodi kohta, mis muudab museaali seisundi hindamise ja konserveerimismenetluste planeerimise kaootiliseks. Sellest tulenevalt katsetasime perioode 1–10 aastat. Eesmärgiks oli leida periood, mille kohta annab mudel kõige paremaid tulemusi.

Eksperimendi viisime läbi otsustuspuu algoritmiga. Kõige paremate tulemusteni nii saagist kui ka täpsust arvestades jõudis tavaline otsustuspuu 10 aasta prognoosimisel: testandmestikust leidis mudel üles 75% halvenevatest juhtudest (saagis) ja kõikidest halvenemise prognoosiga juhtudest 77% olid õiged (täpsus). Lõpliku perioodi valikul lähtusime ka praktilisusest ehk millise perioodi prognoosid toetaksid muuseumide tegevuste planeerimist. Muuseumid teevad pikaajalisi tööplaan aastaid ette ja ühishoidla kontekstis koostatakse isegi 7 aasta tööplaan. Sellest lähtuvalt otsustasime prognoosida museaalide säilivust 10 aasta peale.

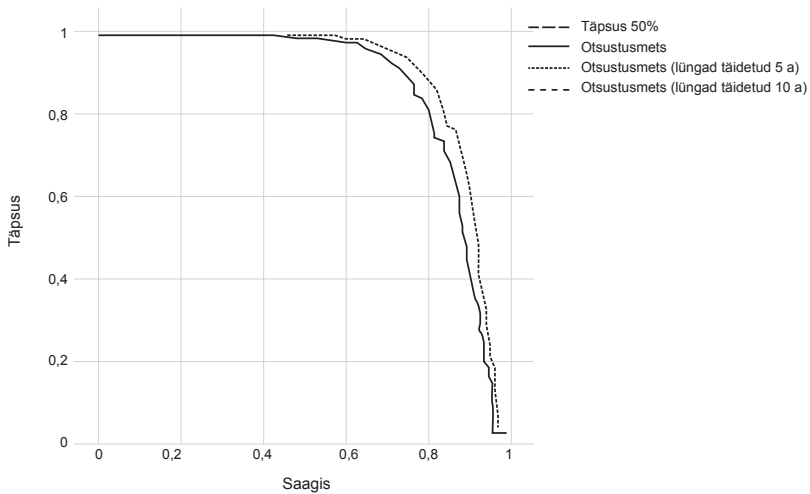
Järgmisena oli vaja leida mudeli optimaalne otsustuslävend. Mudel leiab tõenäosuse, et järgmise n aasta jooksul museaali seisund halveneb. Kui halvenemise tõenäosus on suurem või võrdne seatud lävendiga, siis annab mudel vastuseks “halveneb”. Vaikimisi on masinõppemudelite otsustuslävendiks 50%, kuid olenevalt lahendatavast probleemist, algoritmist ja sellele seatud tingimustest võib optimaalseim lävend olla ka madalam või kõrgem.

Madalama lävendiga on võimalik kätte saada rohkem halvenevaid museaale (saagis paraneb), kuid sellega kaasneb ka kõrgem valepositiivsete hulk (täpsus langeb). Kuna leppisime kokku, et täpsus võib olla minimaalselt 0,5, siis saame saagise tõstmiseks nihutada otsustuslävendit allapoole, nii kaua kuni täpsus on veel suurem kui 0,5. Optimaalseima otsustuslävendi leidmisel kasutasime 10 aasta pikkust ennustusperioodi ehk treenisime otsustuspuud prognoosima järgmise 10 aasta seisundihalvenemisi. Kasutades erinevaid otsustuslävendeid leidsime saagise-täpsuse kõverad (joonis 9). Testisime järgmiseid algoritme: otsustuspuu, XGBoost, materjalipõhine mudel, Naive Bayes ja otsustusmets. Madalama lävendiga on kõrgem saagis, ent madal täpsus (paremal all), lävendit tõstes täpsus paraneb, kuid saagis väheneb. Üldiselt on mudel seda parem, mida suurem on kõvera alla jääv pindala. Kuna oleme kokku leppinud, et minimaalne täpsus on 0,5, võime vaadelda ka joonisel seda piirkonda ja leida, milline mudel annab selle täpsuse juures kõige parema saagise (triipjoone piirkonnas kõige kaugemal paremal). Nagu jooniselt näha, annab parima tulemuse otsustusmetsa mudel (*random forest*), mis 50% täpsuse juures tuvastab testandmetel 90% halveneva seisundiga museaalidest.



Joonis 9. Katsetatud algoritmide saagise-täpsuse kõverad, mis on saavutatud erinevaid otsustusländeid kasutades. Mida suurem on kõveraalune pindala ehk mida kaugemal paremal ülal nurgas kõver asub, seda paremaks üldiselt mudel loetakse (nii saagis kui ka täpsus on keskmiselt kõrgemad)

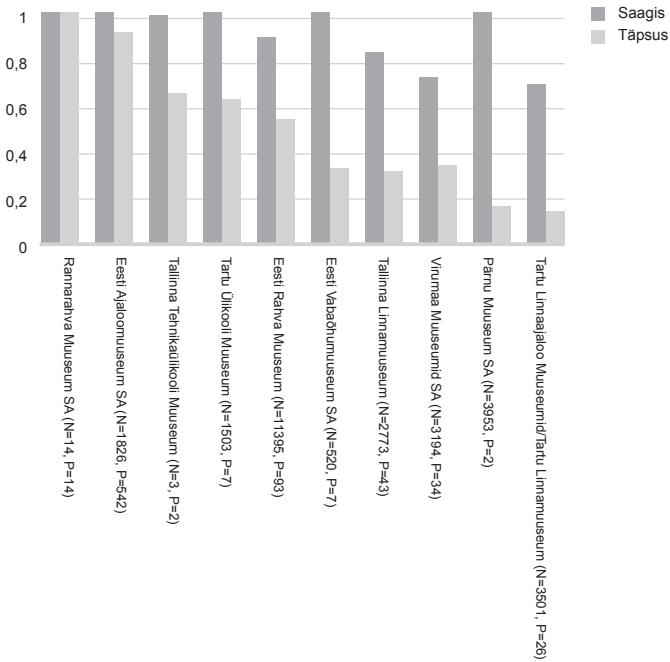
Järgnevalt katsetasime otsustusmetsa algoritmi erinevate andmete ettevalmistamise viisidega. Esmalt lisasime sarnaselt materjalirühmadele ka olemuste rühmad (kokku 97 rühma). Olemuste rühmade lisamine mudeli tulemusi ei muutnud. Teiseks lisasime kunstlikult museaalidele seisundihinnanguid juurde, nii et kahe seisundihinnangu vahele ei jääks rohkem kui n aastat. Lisatud andmepunkt sai viimase seisundihinnangu väärtuse. Nii jäid treeningandmetesse ka need andmepunktid, millele originaalis ei saanud leida silti, kas seisund halvenes või mitte, kuna järgmine seisundihinnang oli rohkem kui 10 aastat hilisem. Kuna selline andmete ettevalmistamise protsess oli küllaltki ajakulukas ja mudeli treenimine võttis lisatud andmete hulga tõttu samuti kauem aega, siis testisime kahe n -väärtusega: 5 ja 10 aastat (joonis 10). Kõige paremad tulemused saime, kui lisasime andmetesse seisundihinnanguid juurde, nii et kahe seisundihinnangu vahele ei jääks rohkem kui 5 aastat. Nii oli 50% täpsuse juures võimalik testandmetes tuvastada 92% halvenevatest museaalidest (2% parem kui andmeid juurde lisamata).



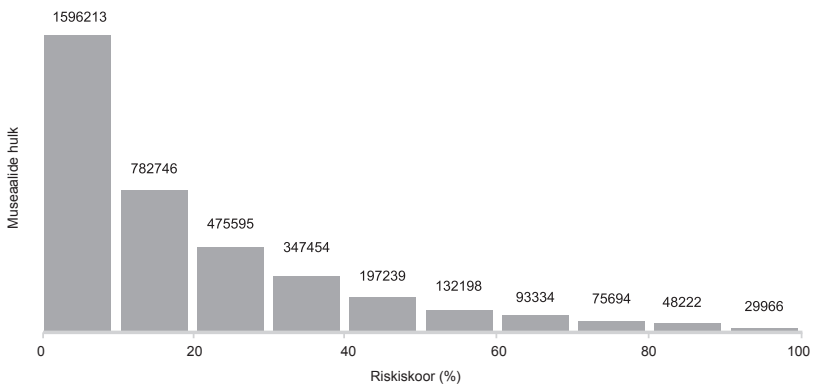
Joonis 10. Otsustusmetsa algoritmiga tulemused erinevatel andmete ettevalmistamise viisidel

Sälli otsustusmudeli hindamine

Eksperimentide käigus saame mudeleid valideerida vaid testandmete peal, mis on väikseim osa kõikidest kasutada olevatest andmetest. Kui nendes andmetes esineb ebamäärasusi või puudujääke, siis esinevad need ka testandmetes. Joonisel 11 on toodud viimase mudeli tulemused pilootmuuseumides testandmetel, kui jääme otsustuslängendi juurde, mille korral on mudelil üle kõigi testandmete 50% täpsus (otsustuslängendi väärtus 6%). Sõltuvalt muuseumist andis mudel erinevaid tulemusi. Mudel annab suhteliselt häid tulemusi näiteks Eesti Ajaloomuuseumis, kus see tuvastas testandmetest korrektselt 541 halvenevat museaali 542st (saagis peaaegu 100%), andes seejuures vaid 41 ehk 7% valepositiivseid (täpsus 93%). Mõõdikute põhjal saavutas mudel kehvad tulemused mitmes suure museaalide hulgaga muuseumis, sh Tallinna Linnamuuseum, Virumaa Muuseumid, Pärnu Muuseum ja Tartu Linnamuuseum, kus üle 60% ja isegi üle 80% positiivsetest prognoosidest olid valepositiivsed. Samas ei ole see nii suur probleem, kui halvenevate museaalide hulk on väike. Näiteks Pärnu Muuseumist oli testandmetes kaks halvenevat museaali ja mudel leidis need mõlemad üles, lisaks prognoosis mudel seisundi halvenemist veel üheksale museaalile (muidugi kokku on Pärnu Muuseumis umbes 40 korda rohkem museaale ja 400 valepositiivse museaali kontrollimine



Joonis 11. Lõpliku mudeli tulemused pilootmuuseumides testandmetel. N – museaalide hulk testandmetes, P – halvenevate museaalide hulk testandmetes



Joonis 12. Sälli mudeli antud riskiskooride jaotus (03.06.2021 MuI Si andmete seisuga)

võtab kauem aega). Tulemused sõltuvad tõenäoliselt olulisel määral muuseumide kogude koosseisust.

Museaali kahjustumise tõenäosus järgmise 10 aasta jooksul on museaali riskiskoor. Sälli mudeli antud riskiskooride jaotus (03.06.2021 MuISi andmete seisuga) on toodud joonisel 12.

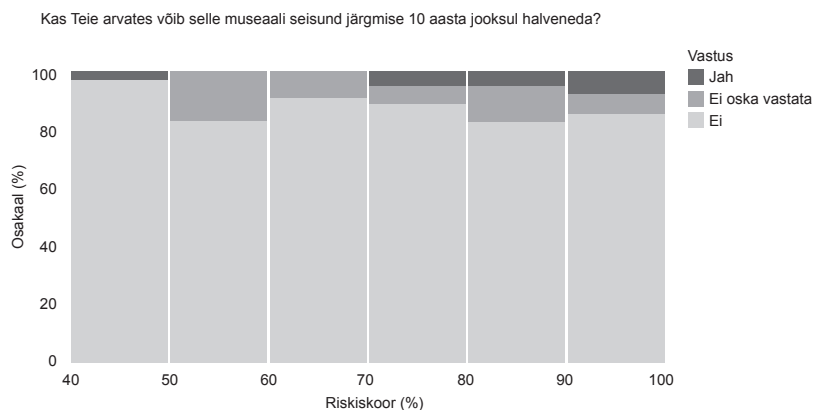
Kõige kõrgema riskiskooriga (80–100%) objektide hulk ei ole suur, moodustades 2% analüüsitud museaalidest. Materjalidest on kõrgema riskiskooriga plastikud, arheoloogilised leiud, aga ka fotomaterjalid ja klaas. Osaliselt langevad need andmed kokku eelmainitud kogemustele tuginedes koostatud heuristikatega materjalide vanemiskiiruste kohta. Samas tuleb arvestada, et masinõppemudel vaatleb enamaid tunnuseid ja suuremat arvu objekte ning võtab arvesse just käsitletavate kogude eripärasid.

Masinõppemudeli prognooside kasulikkuse ja kasutatavuse lõplikuks hindamiseks lõime veebirakenduse, kust muuseumitöötajad näevad oma muuseumi tuhandet kõige kõrgema riskiskooriga museaali, mille seisukorda tuleks esmajärjekorras kontrollida.²⁹ Nimekirjast jätsime välja 2021. aasta jooksul seisundihinnangu saanud museaalid. Museaalide kontrollimisel kogusime kasutajatelt tagasisidet, küsides nende hinnangut museaali halvenemise kohta.³⁰ Museaali kontrollimisel palusime rakenduses vastata küsimusele “Kas teie arvates võib selle museaali seisund järgmise 10 aasta jooksul halveneda?” (joonis 13). Kontrollitud museaalidest 5% ehk 160 kohta arvasid kontrollijad, et seisund võib järgmise 10 aasta jooksul halveneda, 85% juhtudest arvati, et seisund ei halvene järgmise 10 aasta jooksul, ja 10% juhtudest ei osatud vastata. Vahemikus, kuhu langesid kontrollitud museaalide riskiskoorid (48–100%), oli mudeli täpsus testandmetel (mudeli arendamise etapis) 90% ehk otsustuslävendi 48% korral 90% museaalide seisund halvenes ja 10% olid valepositiivsed. Testandmete põhjal oleks seega eeldanud oluliselt suuremat hulka jah-vastuseid. Samas ei saa sellest järeldada, et mudel ei ole võimeline museaalide halvenemist adekvaatselt hindama, kuna vastajate hinnang on subjektiivne ja kuna me ei tea, mis museaalidega järgmise 10 aasta jooksul päriselt juhtub.

Joonisel 13 on toodud vastuste jaotus erinevate riskiskoori vahemike lõikes. Vastajad olid kõikide kontrollitud museaalide suhtes valdavalt seisukohal, et nende seisund järgneva 10 aasta jooksul ei halvene, kuid on näha

²⁹ Eksperimentides kasutasime MuISi andmeid seisuga 27.01.2021. Treenisime prototüübi jaoks parima mudeli (seisundihinnangute lisamine + otsustusmets), kasutades MuISi andmeid seisuga 03.06.2021. Lõpliku mudeli võimekus 50% täpsuse juures jäi 92% saagise juurde.

³⁰ Testisime rakendust 30 testkasutajaga 10 pilootmuuseumist, kes kontrollisid üle rohkem kui 3000 museaali.



Joonis 13. Testijate antud hinnangud seisundi halvenemise tõenäosuse suhtes erinevates riskiskoori vahemikes

trendi, et kõrgemate riskiskooride juures hinnati seisundi halvenemist tõenäoliselt suuremal osal museaalidest. Joonis 13 iseloomustab ka väga ilmekalt museaalide seisundi halvenemise prognooside hindamise keerukust.

Kokkuvõte

Loodud masinõppemudel analüüsib kõikide museaalide seisundihinnangute ja -muutuste ajalugu ning leiab, milliste omaduste ja kasutamise ajaloo museaalide seisund on sagedamini halvenenud. Lõpliku mudeli võimekus 50% täpsuse juures jääb 92% saagise juurde. Projekti käigus hindasime masinõppemudeli võimekust ennustada museaalide säilivust pikema ajaperioodi vältel. Ennustusmudel ei keskendu museaali hetkeseisundi määramisele, vaid üritab erinevaid MuISi andmeid arvesse võttes ennustada museaali säilivust ehk võimalikke seisundihalvenemisi järgmise 10 aasta vältel. Sälli toob esile just need museaalid, mis vajaksid eelisjärjekorras tähelepanu ja mille seisundikontrollini ei pruugita planeeritud tegevuste käigus jõuda. See võimaldaks muuseumidel planeerida erinevaid säilitus- ja konserveerimistööd riskigrupi museaalidega pikema aja vältel ja tegeleda tulevikus pigem ennetamise kui avariilukordadega. Sälli ei dubleeri MuISis olevaid andmeid museaalide seisundi kohta, vaid prognoosib olemasolevatele andmetele tuginedes museaali seisundit tulevikus (10 aasta perspektiivis). See tähendab, et ka näiteks praegu heas seisukorras olev museaal võib olla ohustatute nimekirjas ja samal ajal ei pruugi nimekirjas

olla praegu halvas või väga halvas seisukorras museaale, kui Sälli ei prognoosi nende seisukorra halvenemist. Kvaliteetsete MuI Si andmete korral on Sällil võimalik anda muuseumitöötajale üsna täpne ülevaade riskigrupi museaalidest ja muuseumitöötaja ei pea ressursi raiskama nende museaalide monitoorimisele, mille seisund on stabiilne (umbes 95% museaalidest). Sälli kontekstis jääb museaali seisundi järjepidev hindamine vajalikuks, kuid stabiilsete museaalide puhul oleks võimalik seisundikontrolli intervalli pikendada. Samuti oleks võimalik prioriseerida seisundikontrolli järjekorda, st alustada kõige kõrgema vananemisriskiga museaalidest.

Sälli masinõppemudel võiks abiks olla pigem suurte kogude ja väheste töötajatega muuseumides, samuti muuseumides, kus puuduvad säilitusspetsialistid. Sällit võiks kasutada näiteks tervete muuseumikogude või hoidlate inventeerimise kavandamiseks keskmise riskiskoori alusel või näitusele minevate museaalide ülevaatamise prioriseerimiseks. Sälli masinõppemudeli üks probleem on selle keeruline tõlgendatavus, kuna ei ole võimalik piisavalt lihtsalt ja lühidalt põhjendada, miks mudel ühe või teise skoori andis. Samas puuduks vajadus masinõppemudeli järele, kui me teaksime täpselt, millistel põhjustel ja kui kiiresti museaalid vananevad. Kahjuks on teadmised enamike materjalide vanaemiskiiruste kohta puudulikud. Olemasolevatest andmetest mustreid otsiv masinõppemudel on hetkel ainus võimalus sedalaadi teabe hankimiseks.

MÄRKSONAD: muuseum; museaalide säilitamine; masinõpe; otsustusmudelid

KURMO KONSA on Tartu Ülikooli ajaloo ja arheoloogia instituudi arhiivinduse kaasprofessor ja Kõrgema Kunstikooli Pallas professor.*

MERI LIIS TREIMANN on andmeanalüütik ettevõttes STACC (Software Technology and Applications Competence Centre).

KRISTIINA PIIRISILD on Eesti Rahva Muuseumi konserveerimisosakonna juhataja.

* Kirjavahetus: Ajaloo ja arheoloogia instituut, Tartu Ülikool, Jakobi 2, 51003, Eesti. E-post: kurmo.konsa@ut.ee

ABSTRACT: Algorithms study museum objects: using the machine learning model Sälli to assess the durability of museum objects

The primary task of museums is to preserve museum objects in the form of physical objects. Despite its apparent simplicity and comprehensibility, damage to man-made objects – artefacts – is a complex and complicated field. Damage processes are grouped as being physical, chemical, mechanical, and biological. In most cases, different processes work together, damaging the materials and structure of the artefacts. A number of factors, the most important of which are the composition and structure of materials, environmental conditions, and human impacts, affect damage processes. It is very difficult, and in most cases impossible, to take all these factors into account. At the same time, modelling the aging of museum objects is especially important for their successful preservation. Modelling of damage processes makes it possible to assess the extent of such processes (which objects have been damaged and what the degree of damage is), the speed of damage processes, and thereby changes in the number of damaged objects over time, and finally, the effectiveness of possible management measures.

In this article, we discuss the machine learning model Sälli, which predicts the durability of museum objects. For this purpose, the machine learning model uses data from MuIS (Estonian Museum Information System). The condition of objects is assessed in MuIS with four values: ‘good’, ‘satisfactory’, ‘poor’, and ‘very poor’. Almost 3.7 million condition assessments have been entered into MuIS. The development of a condition prediction model based on these data requires at least pairs of consecutive condition assessments in order to attempt to determine what correlates with the change in condition, whether it be one or another event, or a property (nature, material, age, techniques) of a museum object, or some combination of such factors. There are more than 1.4 million such pairs among the museum objects with several condition assessments. Almost 32,000 of them, or a little over 2%, consist of two different condition assessments, i.e., they indicate a change in condition. According to the data entered in MuIS, almost 30,000 museum objects, i.e., less than one percent of all museum objects, have been subject to a change in condition.

As data points, we used at least two condition assessments for each museum object, to which we added the characteristics of the respective museum object and other features that help to predict the deterioration of the condition of the museum object. These data included static data related to the museum object: museum, museum collection, nature, material,

material group, technology, exhibitability, and dating. As additional information, we used the history of the museum object, i.e., a summary of the events related to the museum object, taking into account only the events that took place during the condition assessment (because we do not have information on the future). The model finds the probability that the condition of the museum object will deteriorate in the next n years. If the probability of deterioration is greater than or equal to a set threshold, the model responds with ‘deterioration’. In finding the optimal decision threshold, we used a 10-year forecast period, i.e., we trained the decision-makers to predict deterioration over the next 10 years.

The best results were obtained using the decision forest algorithm, which was able to identify 92% of deteriorating museum objects with 50% accuracy. This model was also used to create the Sälli prototype. The task of the Kratt Sälli prototype is to draw the attention of museum staff to museum objects, the condition of which may deteriorate in the next 10 years and the situation of which should therefore be reviewed. For testing, a prototype of the 1,000 highest-risk museum objects from that museum was added to each test museum. To test the usefulness and usability of the machine learning model predictions, we created a simple web application that was tested in pilot museums. We found that the available data have the potential to predict deterioration, but the data still need to be improved and the model trained on them is not yet mature enough.

KEYWORDS: museum; preservation of museum objects; machine learning; decision models

KURMO KONSA is an Associate Professor at the Department of Archival Studies at the University of Tartu, and Professor of Conservation at the Pallas University of Applied Sciences.*

MERI LIIS TREIMANN, STACC (Software Technology and Applications Competence Centre), Data Analyst.

KRISTINA PIIRISILD, Estonian National Museum, Head of the Conservation Department.

* Correspondence: Institute of History and Archaeology, University of Tartu, Jakobi 2, 51003, Estonia. E-mail: kurmo.konsa@ut.ee