

TALLINNA TEHNIKAÜLIKOOL
Majandusteaduskond
Majandusanalüüsi ja rahanduse instituut

Raigo Petter

**AUTOMAATSE KINNISVARA HINDAMISMUDELI KASUTUSE
VÕIMALUSED EESTIS**

Magistritöö

Õppekava Ärirahandus ja majandusarvestus, peeriala Ärirahandus

Juhendaja: Ene Kolbre, PhD
Kaasjuhendaja: Ako Sauga, PhD

Tallinn 2021

Deklareerin, et olen koostanud lõputöö iseseisvalt ja olen viidanud kõikidele töö koostamisel kasutatud teiste autorite töödele, olulistele seisukohtadele ja andmetele, ning ei ole esitanud sama tööd varasemalt ainepunktide saamiseks. Töö pikkuseks on 10 026 sõna sissejuhatusest kuni kokkuvõtte lõpuni.

Raigo Petter

(allkiri, kuupäev)

Üliõpilase kood: 106543TARM

Üliõpilase e-posti aadress: raigopetter@hotmail.com

Juhendaja: Ene Kolbre, PhD:

Töö vastab kehtivatele nõuetele

.....

(allkiri, kuupäev)

Kaasjuhendaja Ako Sauga, PhD:

Töö vastab kehtivatele nõuetele

.....

(allkiri, kuupäev)

Kaitsmiskomisjoni esimees:

Lubatud kaitsmisele

.....

(nimi, allkiri, kuupäev)

SISUKORD

LÜHIKOKKUVÕTE	4
SISSEJUHATUS	5
1. KINNISVARA HINDAMINE	8
1.1. Standardid	8
1.2. Hinnatavad väärtused.....	11
1.3. Praktika ja kinnisvara hindamise tulevik	14
1.4. Automaatsed hindamismudelid.....	19
1.5. Automaatsete hindamismodelite kasutamine.....	20
1.5.1. Lihtsad mudelid	22
1.5.2. Hedooniline mudel - regressioonanalüüs	24
1.5.3. Keerulisemad mudelid.....	25
2. MUDELID JA ANDMED.....	29
2.1. Parimad mudelid	29
2.2. Kasutatavad andmed	30
2.3. Kasutatav tarkvara	34
2.4. Valitud mudelid	35
3. EESTISSE SOBIV MUDEL TURUVÄÄRTUSE HINDAMISEKS	37
3.1. Mudelite loomine	37
3.2. Loodud mudelite kirjeldus	40
3.3. Parim mudel.....	42
3.4. Järeldused.....	44
3.5. Ettepanekud	47
KOKKUVÕTE	48
SUMMARY	50
KASUTATUD ALLIKATE LOETELU	53
LISAD	57
Lisa 1. Loodud mudeli skeem	57
Lisa 2. Lihtlitsents	58

LÜHIKOKKUVÕTE

Arvutusvõimsuse kasvu ja masinõppe kasutuselevõtuga arvutites on viimasel kümnendil saanud võimalikuks analüüsida ilma superarvuteid omamata sellises mahus andmeid, milles seoste leidmine ei ole ühelegi inimesele võimetekohane. Paljudes riikides on juba masinõppel põhinevaid erinevaid kinnisvara hindamise mudeleid kasutusel. Sellised mudelid hindavad kinnisvara vähese või üldse ilma inimese sekkumiseta, oodates vaid sisendeid hinnatava vara kohta.

Uurimisprobleemiks on: Kas Eesti väikese kinnisvaraturu tingimustes on võimalik rakendada automaatset hindamismudelit kasutades ametlikku tehinguinfot.

Töö eesmärgiks on leida Eestisse sobivaim mudel turuväärtuse leidmiseks piiratud informatsiooni tingimustes ning testida mudeli täpsust.

Autor uurib maailmas kasutust leidnud ning parimaid tulemusi andnud mudeleid ja loob Maa-ameti ametlikku tehinguinfot kasutades 5 erinevat mudelit: lineaarne regressioon (*Linear Regression*), Bayesi regressioon (*Bayesian Linear Regression*), otsustusmets (*Decision Forest Regression*), närvivõrk (*Neural Network Regression*) ja süvanärvivõrk (*Deep Neural Network Regression*).

Parima tulemuse annab lihtsama närvivõrgu mudeli kasutamine, mille täpsusnäitajad on võrreldavad kirjanduses leitudetega. Sellest järeldub, et Eestis on võimalik rakendada automaatset kinnisvara hindamise mudelit.

Veel saab järeldada, et mudelis kasutatavad ametlikud Maa-Ameti tehinguandmed sisaldavad piisavalt vähe andmevälju, et keerulisemate mudelite eelised lihtsamate ees välja paistaksid, millest tulenevalt süvanärvivõrgu, kui kõige keerulisema mudeli, kasutamine andis hoopis kehvema tulemuse kui mõne lihtsama mudeli kasutamine.

Võtmesõnad: Kinnisvara, automaatne kinnisvara hindamismudel, masinõpe, tehisenärvivõrk

SISSEJUHATUS

Üks rahanduse valdkonna keskseid teemasid on finantsvarade hindamine. Samuti on vaja hinnata vähem likviidseid varasid, seal hulgas kinnisvara. Kinnisvara hindamine on standardiseeritud tegevus, nõuded on mahukad, seetõttu võtab hindamine märkimisväärselt aega, nõuab teadmisi ning suurt infohulka ja vajab eksperti, mis kõik teeb kinnisvara hindamise ka kulukaks.

Siiski on olukordi, kus standardikohane hindamine ei ole vajalik, kuid madal kulu ja kiirus on väga oluline. Sellisteks olukordadeks võib muuhulgas olla ülevaate saamine portfelist, laenuportfellile hinnangu andmine, potentsiaalsete kahjude või kasumi hindamine, investeringute eelotsuste tegemine vms.

Arvutusvõimsuse kasvu ja masinõppe kasutuselevõttuga arvutites on viimasel kümnendil saanud võimalikuks analüüsida ilma superarvuteid omamata sellises mahus andmeid, milles seoste leidmine ei ole ühelegi inimesele võimetekohane. Masinõppe võetakse kasutusele järjest enamatel elualadel, kõige tuntumad näited on isesõitvad autod, käsikirja- ja kõnetuvastus ning kõigis suuremates veebirakendustes töötavad soovitusalgoritmid, mis pakuvad kasutajale tema tegevust analüüsides järgmist võimalikku huvipakkuvat. Kinnisvara hindamise valdkond ei ole jäänud sellest puutumata ning loodud on erinevaid automaatseid kinnisvara hindamise mudeleid.

Sellised mudelid hindavad kinnisvara hindajalt vaid vähest panust oodates või üldse ilma hindaja sekkumiseta, oodates vaid sisendeid hinnatava vara kohta. Mudeleid on loodud väga erinevaid, kuid kõige perspektiivikamateks peetakse tänaseks päevaks masinõppel põhinevaid kinnisvara hindamise mudeleid (Valier, 2020).

Paljudes riikides on juba masinõppel põhinevaid erinevaid kinnisvara hindamise mudeleid kasutusel (edaspidi automaatsed hindamismudelid), neist tuntum on ettevõtte Zillow oma USA-s, mis on arendatud leidma kinnisvara turuväärtust $\pm 20\%$ täpsusega 99%-l juhtudest USA suurlinnades ja keskmiselt 83% juhtudest üleriigiliselt (G. A. Matysiak & TEGOVA, 2018).

Eestis ei ole autorile teadaolevalt kinnisvara hindamise teemat masinõppe vaates uuritud, seega ei ole infot, kas ja mil määral oleks masinõpe Eesti tingimustes kinnisvara hindamisel kasutatav. Nagu selgub pea kõigist kirjandusallikatest, on automaatsete hindamismudelite kasutatavus otseselt seotud andmete hulga ning andmete kvaliteediga, seega on igas riigis kasutusvõimalused erinevad või võib kasutegur hoopiski puududa.

Teema on aktuaalne, kuna masinõppe ja tehisintelligentsi arenguga on jõutud etappi, kus see muudab paljude ametite olemust märkimisväärselt või kaotab need täielikult. Ka kinnisvara hindamine on üks neist mõjutatutest (World Economic Forum, 2020). Ilmselt ei saa hindamine olema täielikult automatiseeritud ning vajab endiselt eksperti, kuid suur osa sellest võib muutuda automaatseks (Royal Institution of Chartered Surveyors, 2017). Üle maailma kasutatakse automaatseid hindamismudeleid, kuid töö autorile teadaolevalt ei ole Eestis saada olevate ametlike andmete alusel toimivaid mudeleid kasutusel. Käesoleva töö kirjutamise ajal on Maa-ametil loomisel mudel maa korralise hindamise teostamiseks. Tegemist on regressioonmudeliga.

Eelnevast tulenevalt on uurimisprobleemiks: kas Eesti väikese kinnisvaraturu tingimustes on võimalik rakendada automaatset hindamismudelit kasutades ametlikku tehinguinfot.

Töö eesmärgiks on leida Eestisse sobivaim mudel turuväärtuse leidmiseks piiratud informatsiooni tingimustes ning testida mudeli täpsust. Eesmärgi täitmiseks on töö autor püstitanud järgmised uurimisküsimused:

1. Milline hindamismudel on parim Eesti piiratud infoga tingimustesse?
2. Milline täpsusklass on võimalik automaatse hindamismudeliga saavutada?
3. Kas Eesti tehinguinfos puuduva, kuid hindamisel olulise seisukorrainfo puudumisel on siiski võimalik hea mudeli puhul leida korteri turuväärtust?

Töö autor plaanib luua automaatse hindamismudeli, mis leiab sisestatava korteri andmete põhjal korteri turuväärtuse parima võimaliku täpsusklassiga. Automaatse hindamismudeli loomiseks kasutab autor Microsoft Azure Machine Learning Studio nimelist masinõppe mudelite loomise veebirakendust. Kuna kirjanduse põhjal on teada, et sellised mudelid ei sobi mittehomoogeensete objektide hindamiseks, valisin mudeli testimiseks kaks Eesti kõige homogeensemata ning piisava tehingute mahuga piirkonda: Tallinna linnaosad Mustamäe ja Lasnamäe.

Magistritöö eesmärgi saavutamiseks viib autor esmalt läbi kvalitatiivse uurimuse teaduskirjanduse abil, selgitamaks maailmas paremaid tulemusi andnud automaatsete hindamismudelite sisu, ülesehitust ning nende meetodite kasutamise võimalikkust ja kitsaskohti Eesti väheste ametlike andmete kontekstis. Sellele järgneb empiiriline uurimus läbi erinevate mudelite loomise ning võrdlemise parima tulemuse (täpsusklassi) saamiseks. Mudelites kasutatavate andmetena kasutab töö autor Maa-ameti ametlikku tehinguinfot, mis on vaieldamatult õige kuid puudulik korteri ja hoone seisukorra info osas, mis on üks olulisemaid turuväärtuse mõjutajaid. Nagu uurimisküsimuses toodud, ongi ühe teemana soov uurida, kas masinõpe aitab vähendada seisukorrainfo puudumisest tekkivat viga.

Töö esimeses peatükis annab autor ülevaate vara hindamise standarditest, teooriast, kasutatavatest meetoditest ning maailmas kasutatavatest automaatsetest mudelitest. Teises peatükis vaatleb, milliseid automaatseid hindamismudeleid oleks võimalik Eestis kasutada ning millised on kasutatavad andmed. Kolmandas peatükis loob autor võimalikke Eestisse sobivaid mudeleid, võrdleb neid omavahel ja leiab parimad.

1. KINNISVARA HINDAMINE

Vara väärtus tugineb vara kasulikkusele ehk selle vara võimele rahuldada inimeste soove ja vajadusi, ütleb majandusteooria. Väärtus on mõiste, mis viitab suhtele, mis eksisteerib saadaoleva vara ning sellest huvitatud olevate ostjate ning müüjate vahel. (Appraisal Institute & EKHÜ, 2008) Väärtuse hindamine võib olla lähtuvalt antud varast lihtsam või keerulisem, kuid väärtus ei ole kunagi fakt, vaid ainult poolte hinnang vara kasulikkusele mingil konkreetsel ajahetkel.

1.1. Standardid

Kinnisvara hindamine on standardiseeritud tegevus, mis tähendab, et standardile vastava hinnangu tellija saab kokkulepitud kvaliteediga hindamistoimingu ja vastava kirjaliku lõpptulemuse.

Standard on üldiseks ja korduvaks kasutamiseks standardiorganisatsiooni poolt vastuvõetud dokument, mis sisaldab tehnopeetsifikaate või juhiseid tegevuse või selle tulemuse kohta ning mille kasutamine on vabatahtlik, kui tehnilises normis ei ole ette nähtud teisiti. (*Tehnilise Normi Ja Standardi Seadus*, 1999).

Standard kirjeldab töövahendeid ja lahendusi, mis aitavad kiiremini, lihtsamalt, ohutumalt ja tõhusamalt tööd teha ning suurendavad seega nende kasutajate usaldusväarsust partnerite ja klientide silmis, standardi järgimine on vabatahtlik, välja arvatud juhul, kui see on õigusaktis viidatud või tehtud kohustuslikuks kliendi, partneri või riigihanke tingimustes.

Kinnisvara hindamisstandardeid koostatakse kolmel tasandil:

1. Riiklikud standardid, Eestis väljaandjaks Eesti Standardikeskus, standard EVS (Eesti Vabariigi Standard).

2. Euroopa standardid, mis on koostatud Euroopa standardimisasutustes. Kinnisvara hindamise puhul EVS (European Valuation Standards) standard, väljaandjaks asutus nimega TEGOVA (The European Group of Valuers' Associations).
3. Rahvusvahelised standardid on koostatud rahvusvahelistes standardimisasutustes. Kinnisvara hindamise puhul on koostajaks IVSC (International Valuation Standards Committee) standardiga, mille lühend on IVS (International Valuation Standard). Levinud on ka teised rahvusvahelisi standardeid koostavate organisatsioonide standardid, nagu Ühendkuningriigi päritolu RICS (Royal Institution of Chartered Surveyors) hindamisstandard nimetusega „Valuation Global Standards“ ning USA päritolu IAAO (International Association of Assessing Officers) hindamisstandardid.

Erinevad standardid ei ole kinnisvara hindamise valdkonnas siiski sisuliselt erinevad. Eesti standardid tuginevad otseselt ja on kooskõlas Euroopa standardiga ning ka IVS standardiga. Euroopa standardid ja IVS kasutavad samuti samu mõisteid ning ei ole vastandlikud. RICS standard on kooskõlas IVS-ga (RICS Valuation Global Standards, 2020), kuid on oluliselt mahukam erinevate varade hindamise ning hindamisvõtete osas. Samuti erineb RICS Euroopa standardist selle poolest, et vaatleb kinnisvara hindamist Suurbritannia õigusruumist lähtuvalt.

Standardite järgimine on küll vabatahtlik, kuid reaalsuses ei ole mittestandardisel kinnisvara hinnangul palju kasutajaid, seega ei ole selline lahendus väljaspool ettevõtete sisest kasutust levinud. Standard kirjutab ette hinnangu valmimiseks vajalikud tegevused (Appraisal Institute & EKHÜ, 2008):

Hindamistellimuse vastuvõtmine ja lähteülesande määratlemine:

- 1) hinnatava vara ja sellega seotud õiguste määratlemine;
- 2) hindamise eesmärgi määratlemine;
- 3) hinnatava väärtuse, väärtuse kuupäeva, hindamise ulatuse ja piiravate tingimuste määratlemine;
- 4) tellimuslepingu sõlmimine.

Andmete kogumine ja analüüs hõlmab järgmisi tööloike:

- 1) hinnatava vara andmete kogumine;
- 2) hinnatava vara ülevaatus;
- 3) turuanalüüs;
- 4) esialgne hinnang kasutatavate hindamismeetodite kohta.

Hindamismeetodid on ühe või mitme konkreetse hindamismeetodi kasutamisel põhinev väärtuse hindamise üldisem viis ehk käsitlus (võrdlusmeetod, tulumetod, kulumetod) (Eesti Standard Vara Hindamine EVS 875, 2017; International Valuation Standards, 2020).

Standard näeb ette ka hindamistulemuse esitamise.

Ekspert hinnangus peab olema kindlasti esitatud (Eesti Standard Vara Hindamine EVS 875, 2017):

- a) hindamise eesmärk;
- b) hinnangu number;
- c) tellija nimi;
- d) viide tellimuslepingule;
- e) objekti määratlus: kinnistu number; katastriüksuse number; vara liik; aadress;
- f) vastavuskinnitus standardi nõuetele;
- g) avalikustamise keelustamise klausel;
- h) vastutuse välistamise klausel kolmandate osapoolte ees;
- i) hinnangu kehtivuse eeldused ja piiravad tingimused;
- j) väärtuse liik ja selle mõiste;
- k) hinnatava objekti kirjeldus, sh kõik hinnatava objektiga seotud õigused ja kohustused; tehnilised ja majanduslikud andmed;
- l) andmete kogumise ja objekti ülevaate ulatus: andmeallikad ja hinnang nende usaldusväärsusele; objekti ülevaate andmed;
- m) turuanalüüs: ülevaade majanduslikust taustsüsteemist; hinnatava objekti turusektori analüüs; võrdlusinfo esitus;
- n) kasutatud hindamismetoodika kirjeldus;
- o) hindamistoimingu kirjeldus selliselt, et hindaja mõttekäik oleks selge, arusaadav ja kontrollitav, sh arvutuskäik, mis seob lähteandmeid ja hindamistulemust;
- p) hindamistulemus;
- q) kuupäevad: hindamiskuupäev; väärtuse kuupäev; objekti ülevaate kuupäev;
- r) hindaja(te) nimi, kutsetunnistuse number ja allkiri;
- s) viited kasutatud allhangetele koos selge vastutuse piiritlemisega;
- t) viited kasutatud algallikatele (dokumentidel kuupäev ja nimi);
- u) hinnangu lisad: asukohaplaanid ja -skeemid; fotod; olulised väljavõtted viidatud materjalidest;
- v) muud andmed.

Eelnevast tulenevalt on hindamistoiming väga ajamahukas osa hindamisprotsessis, olenemata sellest, millisele standardile vastavalt seda teha. Seda probleemi aitaksid lahendada automaatsed kinnisvara hindamise mudelid, mis kiirendaksid protsessi oluliselt.

1.2. Hinnatavad väärtused

Väärtuse kontseptsioon peegeldab turu hinnangut tuludele, mida võib saada see, kes vara omab või teenust saab. Väärtust hinnatakse rahas. Hinnatava vara müük ei ole eeltingimuseks, et anda hinnang väärtuse kohta, st hinna kohta, mille eest vara peaks olema võimalik müüa (Eesti Standard Vara Hindamine EVS 875, 2017).

Hinnangut vara väärtusele mõjutavad pakkumine ja nõudlus, konkurents turul, asenduskaubad, varalt tulevikus oodatavad tulud jms. Kõigi nende tahkude ühiseks jooneks on otsene või kaudne mõju vara kasulikkusele omaniku jaoks. Vara kasulikkus peegeldab kõigi vara väärtust mõjutavate turujõudude kombineeritud mõju (Augustyniak et al., 2018). Lisaks on vara hind alati ajas muutuv – nii nagu muutuvad vara väärtust mõjutavad tegurid, nii muutub ka väärtus.

Varadele väärtuse hinnangu andmine on vajalik paljudel erinevatel otstarvetel: maksustamisel, finants-, või muu aruandluse tarbeks, laenu andmisel ning taotlemisel, erinevate otsuste tegemisel ja mujal, kuid kõige tihemini ostu ja müügiga seotud tegevuste puhul. Tahab ju piiratud ressursside tingimustes müüja saada alati maksimaalset ning ostja maksta minimaalset hinda ükskõik milliste varade eest.

Väärtust klassikaliselt hinnata on seda lihtsam, mida homogeensem on vara, mida rohkem on infot ja mida rohkem varaga kaubeldakse. Kinnisvara väärtus on teiste varade väärtustest erinev selle poolest, et kinnisvara on alati seotud oma asukohaga ning kaht identset vara turul ei eksisteeri, seetõttu on väärtuse leidmine keerulisem ja vähem täpne kui näiteks avalikult kaubeldavatel aktsiatel.

Kinnisvara hindamine kui tegevus, ehk hindamistoiming on defineeritud järgnevalt: loogilises järjestuses olevad tegevused, mille eesmärgiks on hindamistulemuseni jõudmine. Hindamistoiming algab Hindamistellimuse saamisest ja lähteülesande määratlemisest, hõlmab

andmete kogumist, analüüsi ning hindamismeetodite rakendust ja lõpeb hindamistulemuse esitamisega tellijale (European Valuation Standards, 2020).

Tegevuste eesmärk on leida võimalikud sotsiaalseid (demograafia, haridus, elustiil, rahvastiku struktuur), majanduslikke (SKT, sissetulekud, laenu kättesaadavus, ehituskulud, konkureeriv pakkumine ja nõudlus), õiguslikke ja keskkonnategurid (asukoht, ümbrus, saastatus), mis mõjutavad kinnisvara väärtus, ning neid analüüsides leida kinnisvara väärtus antud ajahetkel.

Väärtuse hinnangut kasutatakse enamasti ostu- või müüгитеgevuse käigus, laenuotsuste tegemisel, finantsaruandluse tegemisel, maksustamishinna leidmisel, varade jagamisel ja muudel lepinguliste suhete infona. Kasutajateks on enamasti kinnisvaratehingu osapooled, krediidiasutused, riigiasutused, vara omanikud, audiitorid jne.

Kinnisvara hindamisel on oluline täpsustada, millist väärtust hinnatakse (Mackmin, 1999). Saab hinnata investeringuväärtust (väärtus konkreetse investori vaatepunktist vaadatuna) ja teisi, kuid enamasti hinnatakse ning soovitakse teada siiski turuväärtust, seetõttu käesolevas töös autor teisi väärtuse liike ei puuduta.

Eesti varahindamise standardid defineerivad turuväärtuse vastavalt rahvusvahelistele standarditele: „Turuväärtus on hinnangul põhinev summa, mille eest vara peaks väärtuse kuupäeval minema üle tehingut sooritada soovivalt müüjalt tehingut sooritada soovivale ostjale sõltumatus ja võrdsetel alustel toimivas tehingus pärast kõigile nõuetele vastavat müüгитеgevust, kusjuures osapooled on tegutsenud teadlikult, kaalutletult ning ilma sunduseta (Eesti Standard Vara Hindamine EVS 875, 2017). Sama definitsioon on ka euroopa ning rahvusvahelistes hindamisstandardites.

Tulenevalt kinnisvara eripäradest on kinnisvara hindamises väga olulisel kohal ka parima kasutuse põhimõte. Eranditult igakordne turuväärtuse hindamine eeldab vara parima kasutuse analüüsi. Vara parima kasutuse põhimõte on turuväärtuse oluline mõjutaja ja iga hindamise lahutamatu osa: Parim kasutus on vara kõige tõenäolisem kasutus, mis on füüsiliselt võimalik, vajalikult põhjendatud, õiguslikult lubatav, finantsmajanduslikult otstarbekas ning mille tulemusena hinnatav vara omandab kõrgeima väärtuse (International Valuation Standards, 2020).

Kinnisvara hindamisel kasutatakse kõikide rahvusvaheliste hindamisstandardite ja ka Eesti standardi järgi kolme põhimeetodit (European Valuation Standards, 2020),(International Valuation Standards, 2020):

- 1) turupõhine käsitlus, ehk võrdlusmeetod;
- 2) tulupõhine käsitlus, ehk tulumeetod;
- 3) kulupõhine käsitlus ehk kulumeetod.

Turupõhise käsitluse korral võrreldakse hinnatavat vara sarnaste müüdüd varade hindadega, kasutades selleks erinevaid võtteid, kus meetodi aluseks on asenduspõhimõte: potentsiaalne ostja ei ole valmis maksma vara eest rohkem, kui sarnaste varade eest tavaliselt turul makstakse. (Eesti Standard Vara Hindamine EVS 875, 2017).

Piisavalt aktiivse turu olemasolul võib seda meetodit pidada turuväärtuse hindamiseks sobivaimaks ja täpseimaks (Augustyniak et al., 2018). Oluline on märkida, et turuväärtuseni jõudmiseks peavad kõik hindamisel aluseks olevad võrdlustehingud olema tehtud vaba turu tingimustes.

Tulupõhise käsitluse korral hinnatakse vara väärtust prognoosides tuleviku rahavooge ning leitakse oodatava tulu nüüdisväärtus. Tulupõhised käsitlused on tulu kapitaliseerimine ja diskonteeritud rahavoo analüüs (RICS Valuation Global Standards, 2020).

Samu meetodeid kasutatakse ka teiste perioodilist tootlust pakkuvate varade hindamisel, siinkohal on lihtsalt sisendid kinnisvarapõhised.

Turuväärtuse hindamisel on meetodi aluseks asenduspõhimõte: potentsiaalne ostja ei ole valmis maksma vara eest rohkem, kui ostetav vara talle tulu võib toota. (Eesti Standard Vara Hindamine EVS 875, 2017). Samuti kehtib põhimõte, et kõik hindamise aluseks olevad andmed peavad olema tuletatud vabalt turult (European Valuation Standards, 2020).

Kulupõhine käsitlus tähendab, et hinnatakse vara loomise igal etapil tehtud kulusid. Meetodi aluseks on asenduspõhimõte: potentsiaalne ostja ei ole valmis maksma vara eest rohkem, kui on sarnaste varade omandamise kulud (maa ostu ja hoone ehituskulud) (International Valuation Standards, 2020).

Lisaks on võimalik mõnedes olukordades meetodeid omavahel kombineerida.

Meetodi valik turuväärtuse hindamisel sõltub sellest, kuidas varaga vabal turul kaubeldakse.

Standard soovitab kasutada ärihoonel kui potentsiaalsel rendiobjektidel üldjuhul tulumeetodit, korteril kui müügiobjektidel vastavalt võrdlusmeetodit. Puhta kulumeetodi kasutus on valdavalt väga harv ja tihti mitteturuväärtuse hinnangute kasutuses.

Ei saa öelda, et üks meetod oleks parem või täpsem kui teine. Hindaja kasutab hindamisel ühte kuni kolme hindamismeetodit (või nende kombinatsioone). Mitme meetodi rakendamisel tekib mitu erinevat väärtust, mis kohandatakse lõppväärtuseks erinevate väärtuste kaalumisel (Appraisal Institute & EKHÜ, 2008).

See, et enamasti on vajalik turuväärtuse hindamine, annab samuti põhjust pöörduda automaatsete hindamismudelite poole, kuna just turuväärtuse hindamine tähendab vajadust pidevalt analüüsida võimalikult suurel hulgal pidevalt muutuvat infot. Olenevalt turusegmentidest ja muutustest rahaturgudel võib ekstreemsemal juhul ka nädal tagasi analüüsitud info olla hindamise hetkeks vananenud ning tuleb uuesti analüüsida kõiki sotsiaalseid, majanduslikke, õiguslikke ja keskkonnategureid, mis turuväärtust mõjutavad, et turuväärtust leida, ning selles on arvuti tänaseks oluliselt efektiivsem kui inimene.

1.3. Praktika ja kinnisvara hindamise tulevik

Hindamispraktikas kasutatakse meetodeid täpselt nii nagu standard ütleb – vastavalt sellele, kuidas hinnatava varaga turul kaubeldakse, valitakse ka hindamismeetod.

Täpsemalt tähendab see seda, et kõik üksikud korterid, majad ja muud elamispinnad hinnatakse võrdlusmeetodit kasutades. Meetodi kasutus näeb lihtsustatult välja nii, et turuanalüüsi tulemusena leitakse minimaalselt kolm võimalikult sarnast tehingut võimalikult sarnaste objektidega, tuvastatakse väärtust enam mõjutavad tegurid ning võrreldakse hinnatavat ning võrreldavaid tehinguid iga teguri osas. Kui hinnatava ja võrdlustehingu vahel on erinevus, siis rakendatakse kohandusi vastavalt turuanalüüsile.

Kõik investeringu eesmärgil ostetavad varad, nagu erinevad ärihooned, või muud erinevaid üüripindu sisaldavad varad hinnatakse tulumeetodiga kasutades valdavalt diskonteeritud rahavoogude meetodit. Meetodi kasutuse puhul prognoositakse tulevikus tekkivaid rahavooge ning diskonteeritakse need tänasesse päeva saades tulemuseks turuväärtuse.

Arenduspotentsiaali omavad varad hinnatakse residuaalmeetodiga, mis on kombinatsioon tulu- ja kulumeetodist. Meetodi puhul prognoositakse samuti tuleviku rahavooge diskonteeritud rahavoogude tabelisse kuid sinna lisanduvad kulumeetodil prognoositud kulukomponendid (enamasti erinevad ehituskulud).

Kinnisvara hindamise standard kirjutab ette kogu hindamistulemuseni jõudmise protsessi sammhaaval ning ka selle, milline hindamisaruanne peab välja nägema. Seetõttu on töö maht iga hinnangu tegemisel märkimisväärne ja seetõttu on teenus ka üpris kulukas. Töö kirjutamise hetkel algavad kinnisvaraettevõtete kodulehtedel eksperthinnangute hinnakirjad 150 eurost kõige lihtsamate tüüporterite puhul. Samuti kaasneb hindamisega loomulikult ajakulu, mis on väga varieeruv objekti, konkreetse hindaja töömahu ja hindamise keerukuse tõttu, kuid autori kogemuse põhjal ei ole see Eestis kindlasti kunagi alla kahe päeva, enamikul juhtudel siiski oluliselt rohkem. Ajakulu ei ole seotud vaid standardi nõuetega, vaid ka töökorraldusega – peab ju igal hindajal olema optimaalses mahus tööd, mis tähendab et perioodiliselt võib tööd olla palju ja perioodiliselt vähe, kuid juhtub väga harva, et tellimuse esitamise hetkel kohe konkreetse hindamistööga tegelema hakatakse.

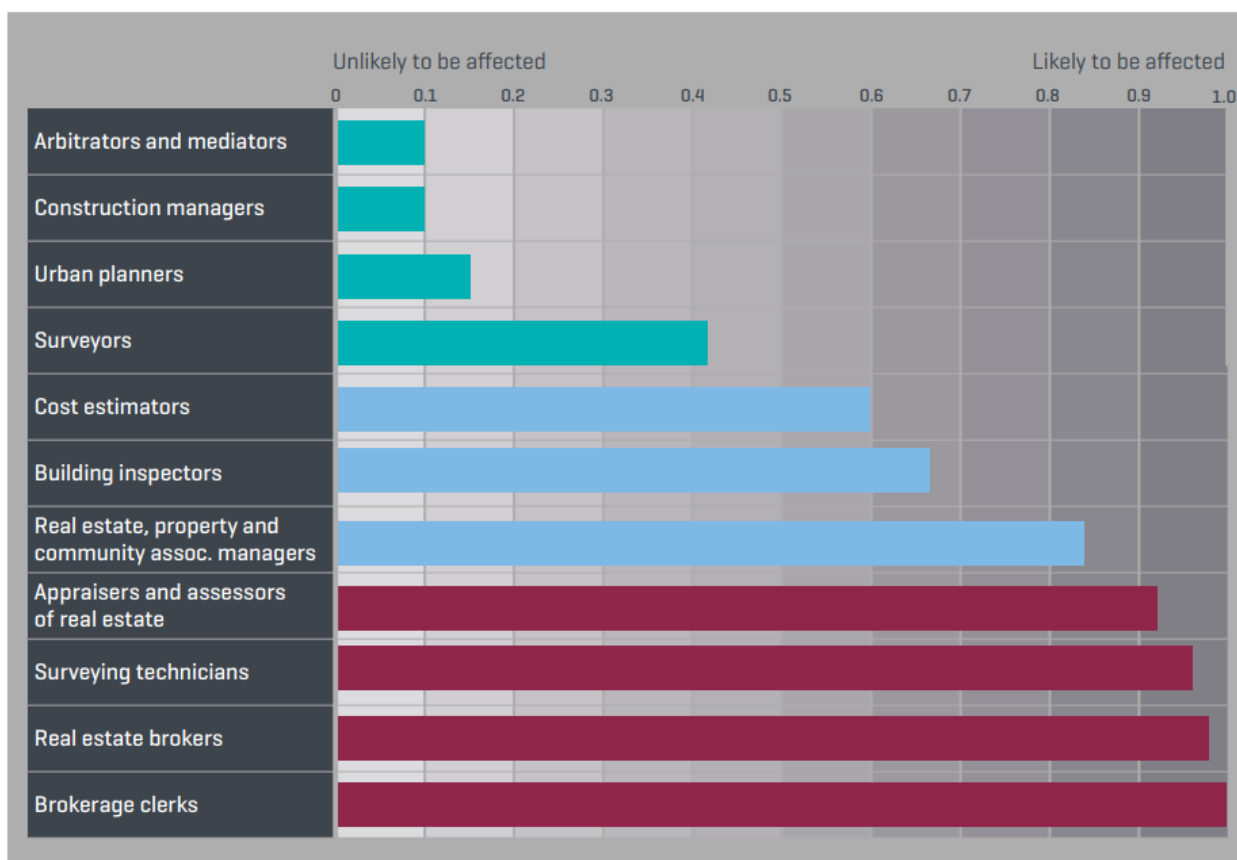
Märkimisväärsest aja- ja rahakulust tulenevalt üritab iga mõistlikult käituv turuosaline kulusid vältida ning kaalub iga kord põhjalikult, kas saadav kasu on seda väärt. Seetõttu on Eestis soojalt vastu võetud teenused, kus pangad laenutagatist ise hindavad, raamatupidamises võimalus juhatusel ise varadele ise hinnangut anda või kasutada sisemisi eksperte või lihtsustatud viise suurtele portfelledele hinnangu andmiseks.

Suurte portfellede puhul on näiteks mõnes krediitiasutuses oma hindamisosakond, mõnes hinnatakse portfelle mõne indeksiga kohandades ja standardikohaseid hinnangud tellivad perioodiliselt vaid need, kellele see on mingitest reeglitest tulenevalt kohustus või muul moel vältimatu nõue.

Alati ei ole vaja teostada standardikohast kinnisvara hindamise protseduuri. Üks lahendus kinnisvara väärtust teada saada on automaatne hindamismudel. Automaatsete hindamismudelite edu kõige tihedamini mainitud põhjus on vähendada kulusid ja kiirendada protsessi. Ka näiteks Ühendkuningriigis võtab standardikohase eksperthinnangu tellimine aega 3-7 päeva, mis on põhjus, miks järjest rohkemates riikides automaatsed hindamismudelid kasutusele võetakse ning juurutamisele kuluv aeg järjest väheneb (Robson & Downie, 2007). Äriprojektid, krediidasutused ja tehingute osapooled konkureerivad teistega ning kiirus annab konkurentsieelise.

Erinevad huvigrupid (RICS, World Economic Forum) on tellinud analüüse, mis uurisid ning üritasid prognoosida, kuidas suurandmete kasutuselevõtt ja laialdasem masinõppe tulek mõjutab erinevaid elualasid.

Järgnevalt toodud joonis näitab, et kinnisvaraga seotud erialad saavad väga suure tõenäosusega olema masinõppe kiire arengu trendist suuresti mõjutatud.



Joonis 1. Tõenäosus, et eriala saab masinõppe tehnoloogia poolt mõjutatud

Allikas: (Royal Institution of Chartered Surveyors, 2017)

Käesoleval hetkel ei oska ilmselt keegi öelda, kas muutused tegevusalal saavad olema sama suured kui olid näiteks taksonduses masinõppe ja IT-lahenduste tulekuga ning kui kaua need muutused aega võtavad, kuid muutused igal juhul tulevad. Ajalisest aspektist saab tuua Küprose näitel välja, et masinõppe küll on turul, areneb ja paraneb, kuid hindajad ei ole siiski väga varmad seda kiiresti praktikas kasutusele võtma (Dimopoulos & Bakas, 2019). Ilmselt oleneb masinõppe kasutusele võtmine siiski rohkem turuolukorrast, seadusandlusest ja riigist kui vaid võimalusi loova tehnoloogia olemasolust turul.

Suurandmete ja masinõppe tulek muudab ka kinnisvaraga tegelevate inimeste maailma vältimatult ja olulisel määral, ning need, kes ei suuda ajaga kaasas käia jäävad maha ning tekivad uued, kaasaegse mõtlemisega ettevõtted, kes viivad turgu suure hüppega edasi. Siiani on olnud palju infot heade töötajate peades ning ettevõtete andmebaasides, suurandmete analüüsi võimekuse tulekuga see muutub ja otsuseid hakatakse tegema analüüsides rohkemaid aspekte aluseks võttes ning vähem tunnetuslikult (Delisle et al., 2020).

Kinnisvaral on muude varadega võrreldes suur hulk eripärasid (Appraisal Institute & EKHÜ, 2008; Eesti Standard Vara Hindamine EVS 875, 2017):

1. Iga vara on unikaalne oma asukohalt, koosseisult, korralduselt.
2. Kinnisvara asukoht ei ole muudetav, ehk see on füüsiliselt immobiilne, seetõttu on väga oluline väärtust mõjutav tegur asukoht.
3. Kinnisvaraturul on palju turusegmente, mille arv on samuti erinevates asukohtades erinev sõltuvalt piirkonnast ja vara liigist.
4. Kinnisvara on väga pikaajaline, kusjuures erinevate osade eluiga on erinev. See tähendab, et maa on igavene, kuna see on hävimatut, maale tehtud parenduste, nagu hoonete, rajatiste, taristu, teede ja muu sellise eluiga on oluliselt lühem ning väga erinev.
5. Kinnisvara pakkumine on alati piiratud, ka siinkohal tuleb täpsustada, et erinevused on maa ja parenduste vahel, kuna maa pindala on limiteeritud, seda ei saa juurde tekkida. Küll aga saab läbi maa parendamise juurde toota ehitisi, kasvavat metsa jm olulisi osasid.
6. Kinnisvaraga on seotud ulatuslik õiguslik regulatsioon, et säilitada tasakaal ühiskonna ja üksikisiku huvide vahel.
7. Kinnisvaraturul ringleb palju raha, tegemist on suurte väärtustega.
8. Kinnisvara kohta teadaolev informatsioon on üldjuhul piiratud, valdkonnas tegutsemine eeldab eriteadmisi ja kogemust.

Seetõttu ei ole kinnisvara kunagi täpselt ühegi teise varaga võrreldav, vajab iga kord, kui soovitakse teada kinnisvara turuväärtust, uut analüüsi ning ühe vara jaoks tehtud analüüs ei anna teise vara turuväärtuse kohta palju infot. Et kõiki ülaltoodud erinevaid kinnisvara eripärasid iga vara puhul uuesti vaadelda ning analüüsida on kasu automaatsetest mudelitest. Arvuti suudab iga kord arvestada vajadusel kõiki toodud erisusi ning analüüsida iga uut vara uuesti ning leida turuväärtus arvesse võttes kogu infot, mida mudel kasutada saab.

Näiteks saab masinõppe tulekuga kasutusele võtta siiani kinnisvara analüüsis vähemkasutatud andmeid nagu (Abidoye et al., 2019; Delisle et al., 2020; Royal Institution of Chartered Surveyors, 2017):

1. Panga poolt omatavad varad piirkonnas.
2. Erinevad kuritegevuse näitajad piirkonnas.
3. Kodanike kaebuste arv politseile, kohalikule omavalitsusele või mõnele muule vastavale ametile piirkonnas.
4. Tehnovõrkude probleemidega seotud statistika piirkonnas.
5. Ehitulubade ja kasutuslubade andmed piirkonnas.
6. Ehitisteta kinnistute arv piirkonnas.
7. Võimalike maksuerinevuste statistika piirkonnas.
8. Erinevad makronäitajad.
9. Majapidamiste suurus piirkonnas.
10. Loodusõnnetuste ohupiirkondade info.
11. Õhu saastatuse info jt.

Näiteid teiste riikide kasutuses olevatest automaatsetest hindamismudelitest ja süsteemidest:

1. Ettevõtte Zillow avalikult kasutatav mudel USA-s, mis on arendatud leidma kinnisvara turuväärtust $\pm 20\%$ täpsusega 99%-l juhtudest USA suurlinnades ja keskmiselt 83% juhtudest üleriigiliselt (G. A. Matysiak & TEGOVA, 2018).
2. 2007. aasta seisuga üle 5% esmastest kinnisvaralaenudest ja üle 60% kinnisvara tagatisel võetud tarbimislaenudest USA-s on väljastatud automaatsete hindamismudelite teostatud hinnangute põhjal ning see on kasvav trend (Robson & Downie, 2007).
3. Rootsis ja Kanadas on pikk ajalugu automaatsete hindamismudelite kasutamisel, neid kasutatakse maksude määramisel (Glumac & Rosiers, 2020a; Kettani & Oral, 2015; G.

Zhou et al., 2018), portfelli hindamisel, tehingute nõustamisel ja vähemal määral laenuandmisel.

4. Juba aastal 2004 avaldatud Basel 2 pangandusnõuded lubavad pankadel kasutada automaatseid hindamismudeleid piiratud ulatuses.

RICS tellitud uuring toob välja, et vaateid tulevikule on erinevaid – ühed arvavad, et automaatsed hindamismudelid saavad jääda vaid hindajat abistavateks tööriistadeks ja teised, et automaatsed hindamismudelid asendavad hindajaid ning eksperti läheb vaja vaid siis, kui mudel ei saa hakkama või muudel erijuhtudel (Robson & Downie, 2007). Käesoleva töö autor kaldub viimase arvamuseavalduse poole ja ei näe mingit põhjust, miks peaks eksperdi kallist aega kulutama tüüpikorterite või eramu hindamisele kui arvuti suudab seda teha sarnase täpsuse, kuid oluliselt suurema efektiivsusega ning arvesse võttes oluliselt rohkem tegureid, kui hindaja suudaks (Wyman et al., 2011). Hindajad peaksid kontrollima mudelite tööd ning hindama keerulisi objekte nagu arendusobjektid, suuremad äripinnad, vähelikviidsed või vähelikviidse turusegimenti varad ja muu selline. Samas ei tahaks autor nõustuda arvamustega, et „Hindaja on dinosaurus“ ja sureb varsti välja (Baxter & Vandell, 2007), sest mõningate varade väärtuse leidmist ei saa autori arvates automatiseerida. Samuti ei usu autor sellele, et tavainimesed sooviksid väga palju panustada kinnisvara hindamisse, täiendades mõne telefonirakenduse kaudu andmeid ja andes tagasisidet nii nagu see toimib hotellivaliku rakendustes, millist tulevikku ennustati ühes teadustöös (Moosavi, 2017).

1.4. Automaatsed hindamismudelid

Automaatne hindamismudel (*Automated Valuation Model*, ehk AVM) on matemaatikal baseeruv arvutitarkvara, mida analüütikud kasutavad turuväärtuse hinnangu andmiseks. Tarkvara kasutab eelnevalt eraldi kogutud ja sisestatud andmeid asukoha, turutingimuste ja kinnisvara iseloomustavate omaduste kohta. Automaatse hindamismudeli eriliseks omaduseks on see, et turuväärtuse hinnang antakse matemaatilise modelleerimise teel (Standard on Automated Valuation Models (AVMs) - 2018, 2018).

Erinevates riikides kasutatavad automaatsed hindamissüsteemid on erinevad vastavalt sellele, kui kättesaadavad ja millise kvaliteediga on andmed, kas mudelite kasutamist on alustanud

riigiasutused, pangad või eraldiseisvad konsultatsioonifirmad jm (Royal Institution of Chartered Surveyors, 2017).

Mudelite katvus geograafiliselt on piiratud neisse sisestatud sisendinfo (tehingute ja objektide andmed) katvusega. Tavapäraselt on katvus parem linnapiirkondades ja oluliselt kehvem hajaasustusega piirkondades. Paljudes riikides, näiteks USA-s ja Austraalias, kus info kättesaadavus on piiratud, tuleb kasutada erinevaid mudeleid erinevate piirkondade jaoks. Riikides, nagu Rootsi, kus info on vabalt kättesaadav ei ole see nii suureks probleemiks (Robson & Downie, 2007).

Üldistatult töötavad mudelid järgmiselt. Mudel valib automaatselt vajaliku turuinfo ja sobiliku meetodi ning kasutab algoritme, mis tavapäraselt sooritavad statistilise analüüsi olemasoleva turuinfo põhjal (G. Matysiak, 2017). Sisendinfona kasutatakse avalikke ja firmasiseseid andmebaase, vanu hindamisaruandeid ja müügiportaalide infot. Idee on pakkuda kasutajale eelist hindamise kiiruses ja kuludes ning vältida hindajapoolset subjektiivsust.

Tüüpiliselt mudeli sisenditeks hinnatava objekti aadress ja standardiseeritud kirjeldus. Mudel identifitseerib selle järgi hinnatava objekti, kontrollib asukohta, suurust ning teisi näitajaid ja otsib siis andmebaasist võrdluseks kõige sobivaimaid tehinguid, mille põhjal hinnatava turuväärtus leida või kasutab muid eelnevalt loodud algoritme (Bergadano et al., 2019). Leitud turuväärtusele lisaks kuvatakse tulemusel mõningatel juhtudel võrreldud tehingute või muu kasutatud info andmed ning tihti teostatud arvutuse täpsusklassi mõõdikud.

Kasutatakse kõiki levinud kinnisvarahindamise meetodeid ning nende kombinatsioone. Areng on aina keerulisemate mudelite suunas, mis suudaksid paremini tabada kinnisvaraturu mittelineaarset olemust ning erinevate mudelite erinevaid tugevusi kasutades mudelite ühendamise ning mudelitest süsteemi loomine.

1.5. Automaatsete hindamismudelite kasutamine

Automaatse hindamismudeli algusajaks võib pidada aastat 1970, kui USA kirjandusse esimest korda tekkis mõiste "*Computerized Assisted Assessments*", paar aastat hiljem võeti see ka laiemalt just maa masshindamise vaates kasutusele. Esimene automaatse hindamismudeli ametlik

definiitsioon ilmus aga alles 2003 (Glumac & Rosiers, 2020b). Seda siis samal kujul kui ülaltoodud 2018 aasta IAAO standardis eelmise peatüki alguses.

Juba 2007 aastal oli automaatsete hindamismudelite kasutamine laialt levinud USA-s, Rootsis ja Kanadas ning paljud teised riigid, nagu näiteks Ühendkuningriik olid kasutust kiirelt laiendamas (Robson & Downie, 2007).

AVM-de kasutus on tänaseks laialdane, neid on kasutatud järgmistel vajadustel (Glumac & Rosiers, 2020b):

1. Kohalike maksude määramine – esimene AVM-i kasutus oli maksu määramise eesmärgil, 1970ndate lõpust on maksustamise eesmärgil kasutanud AVM-e USA, Taani, Holland ja Rootsi.
2. Portfelli riski hindamine.
3. Kindlustusriski hindamine.
4. Laenu andmisel ja riski hindamisel – selgelt kõige rohkem tähelepanu saanud valdkond, kuna võit ajas ja kuludes on märkimisväärne ja igale ostu-müügi tehingu osapoolle tuntav.
5. Vara sundvõõrandamisel ja vahetustehingutel.
6. Kinnisvarainvesteeringute hindamisel.
7. Enampakkumiste alghindade määramisel.

Veel üks suur valdkond kus automaatseid kinnisvara hindamise mudeleid Kanadas, Ameerika Ühendriikides ja Ühendkuningriigis kasutatakse on võimalike pettuste ja vigade avastamine (Robson & Downie, 2007). Pettused või lihtsalt vead nii pankade, muude laenuandjate, hindajate, maaklerite, vahendajate, arendajate ja teiste turuosaliste poolt on lihtsamini avastatavad, kui kogu portfelli ja kõiki tehinguid ning otsuseid kontrollitakse perioodiliselt automaatsete kinnisvara hindamise mudelitega. Pettused või vead võivad ilmneda põhjustatuna kehvadest protsessidest, pahatahtlikest või ka lihtsalt ekslikest inimestest ning automaatsed hindamismudelid aitavad neid märgata ning välja tuua. Loomulikult tuleb tähele panna, et samasugust kontrolli vajavad ka automaatsed kinnisvara hindamismudelid ise, sest ilma pideva kontrollita on vigade ja ka tahtlike pettuste tekkimise võimalus olemas, kuna mudelid hindavad ju alati varasid ilma neid „nägemata“ ning automaatse hindamismudeli kasutaja peab olema see, kes tulemust interpreteerib ning vajadusel objekti üle vaatab.

Üldjoontes näevad mudelid välja järgmised:

1. Vaatlusandmed, ehk kinnisvara hindamise puhul tehinguinfo tuleb reaalselt turult ning moodustab andmebaasi.
2. Andmed andmebaasis koosnevad iga tehingu juures vara tüübi, hinna ja hinda mõjutavate tegurite infost.
3. Andmeid kasutatakse alusinfona empiirilises mudelis või AVM-is, mis kasutab vastavale varale teoreetiliselt sobivat matemaatilist mudelit (võrdlust, lineaarset või mittelineaarset regressiooni, otsustuspuud, närvivõrke või muud) ja arvutab turuväärtuse.

Mudelites kasutatavate andmete kohta on kõigis loetud materjalides selgitus ühene – kasutatakse andmeid, mis olemas on. Kuna erinevates riikides ja isegi riikide erinevates piirkondades on andmete kättesaadavus erinev, siis on ka kasutatavad andmed erinevad. Selge on see, et teadmata toimunud tehingute reaalseid hindu ei saa mudel anda täpseid tulemusi, seetõttu on automaatsete hindamismudelite kasutamine piirkondades, kus hinnainfot ei avaldata, keeruline või piiratud vaid kõige suuremate turuosalistega, kellel on majasiseselt mudeli loomiseks piisavalt tehinguandmeid. Kui tehinguhindade info on avalikult kättesaadav, on võimalik kinnisvara hindamise mudelit luua, kuid selle mudeli tüüp ja täpsus oleneb paljuski andmete mahust ja kvaliteedist.

Mudeleid hinnatakse mudeli seletavusvõime, prognoosivõime ja täpsuse järgi. Mida parem mudel, seda paremini sobib konkreetne mudel konkreetsete andmetega hinnanguid andma.

Tavapärane mudelite hindamise meetod on järgmine: olemasolevad kinnisvaratehingute andmed jagatakse kahte gruppi, treeninggrupp ja testgrupp. Treeninggrupp on tavapäraselt 70-80% andmetest (olenevalt andmete hulgast ka väiksem või suurem), ülejäänud on testgrupp. Järgmises faasis nõ treenitakse algoritmi leidmaks parimaid seoseid treeninggrupi andmetega, seal asuva kinnisvara hinna ja seda mõjutada võivate kinnisvara omaduste vahel. Järele jäävaid andmeid ehk testgruppi kasutatakse leitud mudeli testimiseks, et leida loodud algoritmi täpsust – mida lähemal on mudeli prognoositav väärtus reaalsele testgrupi tehingute hindadele, seda täpsem on mudel.

1.5.1. Lihtsad mudelid

Erinevad indeksid

Ilmselt kõige lihtsamad mudelid on indeksipõhised mudelid. Neid ei kasutata enamasti üksiku kinnisvara hindamisel, vaid näiteks portfelli väärtuse jälgimisel või muul taolisel eesmärgil.

Lihtsustatult tähendab see, et teadaolev väärtus mingist mineviku kuupäevast korrutatakse läbi indeksist tuleneva kordajaga (näiteks aasta tagasi ostetud vara tänase orienteeruva väärtuse teadasaamiseks lisatakse aastatagusele turuväärtusele aastane THI tõus või mõne täpsema kinnisvaraindeksi muutus ning saadakse tänane orienteeruv turuväärtus). Sellist mudelit saab kasutada suuremate portfelli või laenuportfelli hindamisel, kus oluline on turuväärtuse suurusjärk, mitte täpne väärtus ning üksikobjekti tasandile ei minda. Keerulisemal juhul kasutab mudel mitmeid kordajaid või erinevaid kordajaid erinevate portfellis olevate varatüüpide osas. Kui eelnevat hinnangut või tehingut toimunud ei ole, need mudelid enamasti ei tööta, jättes paljud varad hindamata ja seepärast ei peeta neid mudeleid tihti päris automaatseteks mudeliteks. Lisaks on uuritud ka indeksi prognoosimise võimalusi masinõppe meetoditega (Abidoye et al., 2019) kuid see meetod ei ole praktikas levinud.

Selliste mudelite kasuks räägib läbiproovitud ja statistiliselt korrektne lähenemine. Kui õigesti kasutada, annavad sellised lahendused hea tulemuse, kuna piisavalt suure portfelli puhul tekib vigu nii ühele kui teisele poole ning portfelli vaates vead tühistavad üksteist. Kasutuse eelduseks on mingi mineviku kuupäeva seisuga turuväärtuse info olemasolu.

Võrdlusmeetodit matkiv mudel

Veidi keerulisemad mudelid üritavad jäljendada hindaja tegevust võrdlusmeetodi kasutamisel. Mudel kasutab sisestatud tehinguinfot, piirkonna või turusegmendi keskmisi või valib sisestatud info hulgast sobivaima ning rakendab vastavalt hinnatavale objektile mõningaid kohandusi, mis näiteks muudavad keskmise turusegmendi hinna hinnatava objektiga võrreldavamaks (hinnates kehvas seisus Lasnamäe korterit võtab mudel aluseks Lasnamäe keskmise hinna ning lahutab sellest x% lähtuvalt hinnatava korteri seisukorrast). Seda tüüpi mudeleid saab ehitada ka keerulisemaks ja oluliselt põhjalikumaks kuid kuna kohanduste tüüp ja suurus tuleb mudeli kasutaja poolt pidevalt üle vaadata ja turusituatsiooniga vastavusse viia on keerulisemate lahenduste puhul vajalik eksperdi pidev kaasatus, mis teeb mudelist abivahendi eksperdile, mitte iseseisva hindamismudeli.

Selliste mudelite hinnangud sobivad hästi erinevate eelotsuste tegemiseks, kus ei ole teada ajaloolist turuväärtust ega mingit head indikatsiooni turuväärtuse kohta ning turuväärtuse suurusjärgu hinnangust piisab, et otsustega edasi minna.

1.5.2. Hedooniline mudel - regressioonanalüüs

„Regressioonanalüüs on statistiline meetod, mis võimaldab koostada matemaatilise võrrandi, et kvantifitseerida sõltuva muutuja (tulemuse) ja ühe või mitme sõltumatu muutuja (sisendi) seost.“ (Sauga, 2020). Hindamisel on sõltuv muutuja tavaliselt hind või rent ja sõltumatud muutujad vara väärtust mõjutavad füüsilised, majanduslikud ja sotsiaalsed tegurid.

Regressiooni on kasutatud juba aastakümneid, algselt põhiliselt masshindamise meetodina maksustamisel (Appraisal Institute & EKHÜ, 2008). Tänapäevaks on regressioonanalüüs saanud üheks automaatsete hindamismudelite alamhulgaks ning on oma erinevates vormides kasutatav ilmselt kõige laialdasemalt.

Lihtsa regressioonmudeli idee seisneb selles, et seost iseloomustavat punktiparve valitakse esindama selline sirge, millest kõikide üksikpunktide kauguste ruutude summa on minimaalne. Regressioonimudeli põhjal saadud prognoos on üksikobjekti jaoks pea alati (pisut) ebatäpne sest mudel prognoosib vaid keskmist taset. Mudeli “headust” hinnatakse selle põhjal kui tugev on seos sõltumatute tunnuste komplekti ja sõltuva tunnuse vahel, sest see määrab prognoosi täpsuse. Mudeli “headust” väljendab ka mudeli standardviga, mis kirjeldab sõltuva tunnuse väärtuste keskmist kõrvalekallet ehk erinevust prognoosist. Mitme seletava tunnuse puhul kasutatakse mitmest regressiooni. Mudelit saab ka integreerida teiste lahendustega nagu näiteks geoinfosüsteemiga (Xiao-sheng et al., 2011).

Mudeli plussideks on heade sobivusstatistikute olemasolu, vajaliku tarkvara lihtne kättesaadavus ning mudeli läbipaistvus – selgelt on võimalik tuvastada, millistest teguritest väärtust koosneb ning milline tegur väärtust rohkem mõjutab. See on muutnud regressioonmudeli laialdaselt aktsepteeritavaks ning arusaadavaks meetodiks

Mudeli miinuseks on statistiliste teadmiste omamine kasutaja poolt ja mudeli raskused mittelineaarsete seostega (Brotman, 1990). Tuleb märkida, et mittelineaarsete seoste leidmiseks on võimalik kasutada edasiarendust – üldistatud aditiivne mudel (GAM, ehk *General Additive Model*) (Adamczyk, 2010).

1.5.3. Keerulisemad mudelid

Regressioonipuu (*Regression tree*)

Regressioonipuu on otsustuspuu, mis ei otsusta/klassifitseeri vaid annab numbrilise tulemuse. Otsustuspuu loogika on leida sobivaim otsus, klass või number läbi erinevate otsustuskriteeriumite. Alustades juurest (sisendinfo) jagab algoritm andmed kaheks grupiks (haruks), mis omakorda jagunevad kaheks ja nii edasi, kuni jõutakse viimase lehe ehk otsuseni, kus andmed on oluliselt homogeensemad kui alguses. Iga jagunemine, ehk sõlm, kasutab mudelisse sisestatud seletavaid muutujaid ja jagab iga haru väärtusega, mis maksimeerib muutuja müra vähenemise (Gini heterogeensuse indeks).

Mudeli plussiks on lihtsus, arusaadavus ja lihtne jälgitavus.

Sellise mudeli miinuseks on kõrge ülesobitumise risk. Mudel õpib sobituma nii hästi treeningandmetega, et testandmetega sobivus väheneb. Selle probleemi vältimiseks piiratakse õppimisaega või kasutatakse teisi spetsiaalseid tehnikaid (Valier & Micelli, 2020).

Lähimad naabrid (*Nearest neighbors*)

Algoritm ei ürita leida koordinaate, mis selgitaksid kogu andmehulka nagu regressioon vaid tegeleb vaid vaatlusalusele ülesandele kõige sarnasemate andmete analüüsiga. Põhineb sellele, et looduses ei ole peaaegu kunagi ükski vaadeldav grupp lineaarse jaotusega kuid grupi sees on väikeseid alagruppe, mis seda on. Algoritm analüüsib erinevusi vaadeldava ja teiste andmete vahel ning leiab tulemuse vaid kõige sarnasemate põhjal.

Mudel töötab hästi masshindamise puhul ja andmete mediaani lähedal, ekstreemumitele lähenedes prognoosimisvõime halveneb (Valier & Micelli, 2020).

Juhuslik mets (*Random forest*)

Juhusliku metsa algoritm tähendab mitut kokku pandud regressioonipuud, mida eelpool kirjeldatud. Nagu eelmainitud regressioonipuud, otsib ka juhuslik mets oma õigeimat teed juurest leheni. Metsa puhul kasutatakse juhuslikku valikut nii vaatluste kui parameetrite valikul. Selline mudel esindab mitmete puude agregeeritud õppimistulemust ja pakub seetõttu paremat tulemust kui iga puu individuaalselt.

Mudeli plussid on avatus ja jälgitavus nagu ka regressioonipuul, kuid puuduvad regressioonipuudused. Mudel ei vaja suurt andmete eeltöötlust ja ei sobitu üle (Wang, 2019).

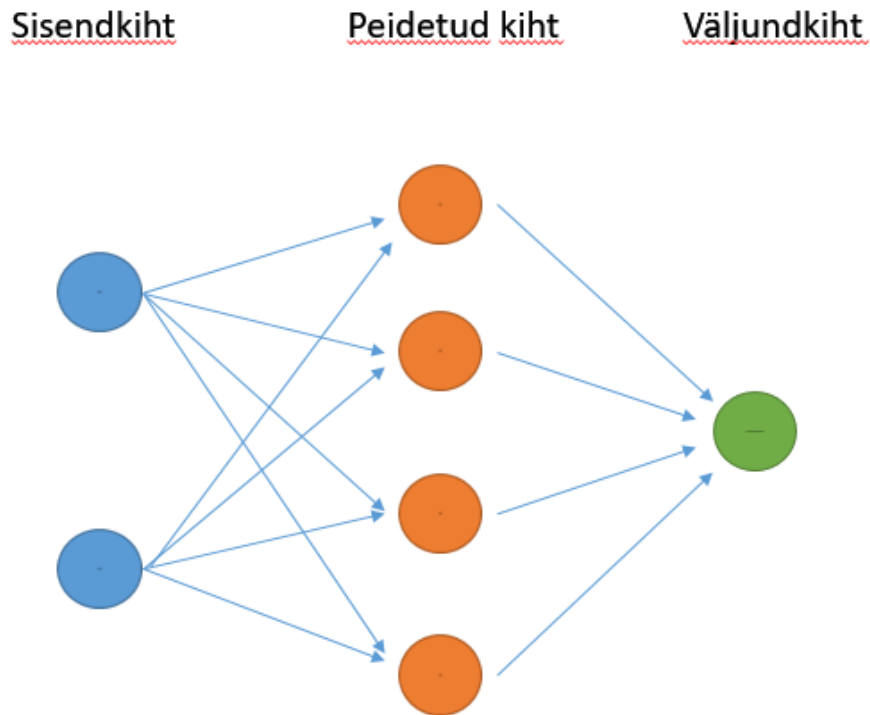
Võimendamine (*Boosting*)

Võimendamine tähendab masinõppes mudelite koondamist gruppidesse, kus nad üksteist täiendavad. Otsustuspuu, mis on võimendatud koostab mitmeid otsustuspuid üksteise järel, nii et iga järgnev puu kasutab eelneva puu poolt leitud prognoosi viga ning parandab seda. Nii saab lõpuks parema tulemuse kui iga puu üksi võiks saada.

Tehisnärvivõrgud (*Artificial Neural Network - ANN*)

Tehislikud närvivõrgud (edaspidi lihtsalt närvivõrgud või ANN) on algoritm, mis töötab hästi mittelineaarsete seoste leidmisel ja on inspireeritud inimaju töötamise loogikast. Närvivõrk koosneb neuronitest, mis meenutavad inimaju neuronite tööpõhimõtet: neuron edastab signaali läbi sünapside teistele neuronitele, milles igaühes toimub info töötlemine ning töödeldud info edastatakse järgmisele neuronile. Tehisnärvivõrgus tähendab signaal mingit arvu ning töötlemine kaalu andmist läbivale infole mõne mittelineaarse algoritmi abil. Õppimise protsessis muudetakse kaalusid, mille tulemusena leitakse lõpuks parim prognoositav väärtus.

Neuronid on tavapäraselt paigutatud kihtidena, närvivõrk koosneb sisendkihist, väljundkihist ning nende vahel olevatest ühest kuni mitmest niinimetatud peidetud kihist. Erinevad peidetud kihid võivad erinevaid ülesandeid täita, sisend- ja väljundkiht täidavad oma rolli vastavalt nende nimetustele sisendinfo andmiseks ning tulemuse väljutamiseks. Närvivõrke saab kasutada nii klassifitseerimisülesannete kui regressiooni puhul, tuntumad kasutusala on arvuti poolt tekstist arusaamine ning piltidel oleva tuvastamine.



Joonis 2. Lihtsa närvivõrgu skeem

Allikas: Autori koostatud

Närvivõrkude tugevuseks on uue info lisandumisel seda arvesse võtta, tunda ära mittetäielikke, keerulisi ja vähenähtavaid mustreid andmetes (Abidoye & Chan, 2018).

Närvivõrkude nõrkuseks on varjatud kihis toimuvate protsesside keerukus, nende protsesside täpse kirjeldamise ja põhjendamise võimatus ning sellega seoses raskused tulemuse kontrollimisel ning kinnitamisel. Lisaks sobivusstatistikute puudumine (Standard on Automated Valuation Models (AVMs) - 2018, 2018).

Tugivektor-masin (*support vector machine, SVM*)

Meetodi taga on idee, et oma ruumis asuv probleem sisaldab elemente, mida on raske üksteisest eraldada, kuid lisades ruumile mõõtmeid muutub ülesanne lihtsamaks. Tugivektor-masin otsib nii-öelda tugivektoreid, mille abil on võimalik teha eristusi.

Tugivektor-masinate plussiks on mittelineaarsete probleemidega hästi hakkama saamine, seetõttu on mõned uurijad pidanud seda parimaks kinnisvara väärtuse leidmise algoritmiks kuid selle teemalisi uuringuid on veel liiga vähe (Valier & Micelli, 2020).

Mudelid mudelite sees

Keerulisemate mudelite hulka kuuluvad ka kõigi eelmainitud mudelite kombinatsioonid. On võimalik luua mudeleid, mis koosnevad mitmetest mudelitest, mille kasutamise valiku teeb kõrgemal asuv mudel kas vastavalt sisenditele või muudele eelnevalt sisestatud parameetritele. Samuti saab jagada andmeid erinevate mudelite vahel, mis parandab tulemust (kinnisvara hindamisel sobiv näide võiks olla, et hindamise mudel on üks, kuid selle sees hindab 1-toalisi kortereid teine mudel, 2-toalisi kolmas jne) või kasutada ühe mudeli sees korraga mitmeid mudeleid ning väljastada tulem vaid suurima usaldusväärsusega mudelist, keskmisena või muul moel optimaalseimat lahendust pakkudes. Samuti kasutada avalike andmebaaside ja pildiinfot (You et al., 2017).

Selliste mudelite miinused on jällegi raske jälgitavus, tihti suure andmehulga vajadus ning probleemne kontrollitavus.

2. MUDELID JA ANDMED

Kõigi eeltoodud mudelite kasutamine analüüsis ei ole võimalik, seega vaatleb autor käesolevas peatükis, millised on kirjanduse põhjal parimaid tulemusi andnud mudelid, et nende hulgast valida töösse parimad. Samuti vaatleb kasutatavaid andmeid ning tutvustab kasutatavat tarkvara.

2.1. Parimad mudelid

A.Valier teostas 2019 aasta juuli seisuga uuringu, mille käigus ta otsis üles ning analüüsis läbi kõik selleks hetkeks avaldatud teaduslikud artiklid kinnisvara turuväärtuse hindamise teemal, kus võrreldakse kinnisvara hindamiseks loodud hedooniliste mudelite tulemusi masinõppe mudelite tulemustega samadel andmetel. Erinevaid eeltoodud tingimustele vastavaid artikleid leiti uuringusse 53 (Valier, 2020).

Teostatud analüüsi tulemusena leiti, et ei saa väita kindlalt, milline masinõppe mudel on parem kui teised sama eesmärgi täitvad masinõppe mudelid, kuna valim ei olnud kõigi mudelite osas piisavalt suur. Küll aga saab antud töö põhjal väita, et masinõppe mudelid on enamasti parema prognoosimisvõimega kui klassikalised hedoonilised regressioonmudelid. Valier toob ka välja, et hedoonilistel mudelitel on paremad seletavad omadused, masinõppel aga paremad prognoosivad omadused.

53 erinevat teadustööd analüüsid Valier, et regressioonmudel oli küll uuringus mõnest masinõppe mudelist parem 13-l korral, kuid masinõppe mudelid olid regressioonist paremaks tunnistatud 57-l korral (paljudes vaadeldud töodes oli kasutatud mitut mudelit, mistõttu on mudelite arv suurem kui teadustööde arv). Samuti tuli uuringust välja, et regressioonipuud on teistest mudelitest kehvemad. Ka tavalisest regressioonist olid paremad vaid juhusliku metsa vormis mudelid, mitte aga regressioonipuud ise. Lähimate naabrite (Nearest Neighbors) mudel oli mõnest teisest mudelist parem vaid ca 20% töödest, mis seda mudelit sisaldasid. Parimate mudelitena paistavad vähesest valimist olenemata lisaks närvivõrkudele välja SVM tüüpi mudelid ja mitmetest mudelitest kokku pandud hübriidmudelid (Valier, 2020).

SVM mudeleid on kahjuks veel vähe uuritud ja seetõttu ka täpsemat infot väga vähe, närvivõrkusid on aga hindamise vaates piisavalt vaadeldud. Närvivõrkude ühendamisel teiste algoritmidega on saadud veelgi paremaid tulemusi kui lihtsalt närvivõrke kasutades, näiteks on saadud paremaid tulemusi geneetilise algoritmi ja närvivõrgu ühendamisel (Sun, 2019), võimendamise ning närvivõrgu ühendamisel (Bin et al., 2017), lähimate naabrite ning närvivõrkude ühendamisel (Bin et al., 2019), tekstiõppe ja närvivõrkude ühendamisel (X. Zhou et al., 2019), erinevat tüüpi närvivõrkude nagu näiteks pildituvastuse närvivõrgu ning hindamise närvivõrgu ühendamisel (Poursaeed et al., 2018). Samuti on häid tulemusi saavutatud erinevate võimendusmeetodite ühendamisel (Kim et al., 2020).

Enamasti on keerulisemad mudelid paremad just suuremate andmehulkadega kasutamisel, üle 2000 väga detailse tehinguinfo reaga uurimisel Hispaanias olid närvivõrgud ja regressioonipuud teiste mudelitega võrreldes oluliselt paremad (Baldominos et al., 2018).

2.2. Kasutatavad andmed

Mudeli loomiseks kasutatavaks infoks võiks teoreetiliselt olla avalikult kättesaadavaid kanaleid vaadates kinnisvaraportalides esitatav pakkumisinfo ning vähem kättesaadavana piisavalt suure turuosaga ettevõtete andmebaasides olev tehinguinfo nende poolt müüdud objektide kohta. Kõige täpsem on aga avalikkusele üldse mitte kättesaadav Maa-ameti detailsem tehinguinfo, mis on siiski kättesaadav Maa-ameti litsentsi omavatele kinnisvara hindajatele.

Kriteeriumid, millise info alusel maailmas mudeleid luuakse, on et info on kvaliteetne, täpne, saadud usaldusväärsetest allikatest ning piisav mudeli loomiseks.

Kinnisvaraportalides olev info on piisavalt mahukas, kuid ei ole kunagi täpne müügihinna osas, kuna tegemist on vara eest küsitava ehk pakkumishinnaga. Samuti ei pruugi kinnisvaraportalidest leitav info olla piisavalt kvaliteetne, kuna ka ülejäänud andmetes võib olla vigu, arvestades, et igasugune kontroll andmete sisestamise üle puudub, andmevälju on väga palju ning enamik andmeväljade täitmine ei ole kasutajale kohustuslik.

Kinnisvaraga tegelevate ettevõtete andmebaaside info on kindlasti täpsem tehinguhinna osas, kuna maakler on tehingu juures lõpuni ning teab teostatud tehingu lõplikku hinda. Samuti on see info piisavalt detailne, kuna maakler teab kõiki müüdud vara üksikasju ja ka piisavalt mahukas ning tõenäoliselt oluliselt paremini sisestatud ja kontrollitud kui kinnisvaraportaalide info. Kahjuks ei ole aga kinnisvaraga tegelevate ettevõtete info teistele, väljast poolt konkreetset ettevõtet, kättesaadav. Kõige suuremaks probleemiks on info aga liiga vähese mahuga kogu turu või turusegmendi üleste järelduste või analüüside tegemiseks. Seda põhjusel, et Eestis ei oma ükski kinnisvaraettevõtte piisavat turuosa, et teha mõnes turusegmendis kogu tehingute arvust kas või pooledki tehingud.

Kõige paremini vastab enamusele eelpool nimetatud tingimustele Maa-ameti kogutav tehingute info. Maa-ameti info sisaldab absoluutselt kõiki vastavas turusegmendis tehtud tehinguid, on ametlik ja täpne tehingu hinna ning kõigi muude andmeväljade osas, kuna Maa-amet koondab tehingute infot, mida neile saadavad notarid. Ainuke probleem on see, et Maa-ameti info puhul on piisavuse tingimus täidetud mööndustega, kuna Maa-ameti info ei ole väga detailne ning ei sisalda väga palju andmevälju. Lisaks on mitmed sisalduvad andmeväljad juriidilise iseloomuga, mis kirjeldavad tehingut, mitte niivõrd müüdud vara selle tehingu taga. Automaatse kinnisvara hindamismudeli loomisel on Maa-ameti tehinguinto suurimaks probleemiks see, et see info ei sisalda tehingu objekti seisukorra kirjeldust ega muud sarnast infot, nagu näiteks pilte varast. See probleem muudab väljatöötatava mudeli kasutuse komplitseerituks eripärase hoonestusega piirkondades ning autori arvates väga raskeks eramute hindamisel, kuna homogeenseid eramutega piirkondi on kogu eesti peale üksikuid ja needki on väikesed. Võrreldes erinevaid andmeallikaid on siiski andmete täpsus, kvaliteetsus ja õigsus mudeli loomisel olulisem kriteerium kui andmeväljade piisavus, kuna ilma turutehingute täpsete tehinguhindade infota ei ole võimalik turuväärtust leida.

Kuna kirjanduse põhjal töötavad mudelid kõige paremini võimalikult homogeensetes piirkondades, siis valis autor mudelite loomiseks kaks kõige suurema tehingute arvu ning kõige homogeensemata piirkonda: Lasnamäe ja Mustamäe linnaosa Tallinnas.

Kasutatavate andmete probleemiks on hoone ja korteri seisukorra puudumine. Teadusartiklite põhjal on leitud, et seisukorrainfot omamata on võimalik saada aktsepteeritava täpsusega tulemusi, kui selle asemel kasutada hoone vanust (Oshodi et al., 2019). Oshodi leidis sellist kasutust 15nes teadusartiklis. Käesoleva töö autor ei ole siiski väga kindel, et hoone vanuse kasutamine mudelites

Eesti tingimustes oleks väga hea seisukorrainfo asendaja, kuna sellist võimalust kasutanud uurimused on teostatud stabiilse arenguga riikides, kus hooneid on aastakümnete vältel samadel põhimõtetel hooldatud ja remonditud. Eestis, lähtuvalt (paljude hoonete puhul mitmetest) riigikorra muutustest, ei ole läbi aja pidevalt stabiilset ehitustegevust ega remonte toimunud, mis tähendab, et on väga levinud aastakümneid täiesti hooldamata hooned ning pikaajalisest hooldamatusest tulenevalt ka oma kasutusea vaates liiga vara kapitaalselt rekonstrueeritud hooned. Näiteks võib igast eesti asulast leida kõrvuti asetsevad 60ndatel ehitatud eramud, millest üks võib olla kvaliteedilt võrreldav uue hoonega kuid teine sobib vaid lammutuseks. Kahjuks ei ole Maa-ameti tehinguandmetes toodud hoone ehitusaastat, seetõttu ei saa testida erinevusi.

Mudeli loomisel kasutab autor Maa-ameti tehinguandmeid aastast 2014, 01.08 kuni 31.10 ehk kolm kuud. Andmed sisaldavad vaadeldava kolme kuu jooksul tehtud kõiki korteriomanditega tehtud tehinguid Mustamäe ja Lasnamäe linnaosades. Vaadeldava kolme kuu jooksul tehti vaadeldavates linnaosades 706 tehingut. Andmete hulk ei ole väga suur kuid peaks olema piisav töö eesmärgi vaates, kuna mudeleid on loodud ka oluliselt väiksema tehinguinfo hulga pealt (väikseim 30 tehingut (Giudice et al., 2017)) ning saadud häid tulemusi.

Kasutatavad Maa-ameti tehinguandmed sisaldavad järgmist infot: asukoht maakonna, linna ja linnaosa täpsusega, asumi nimetus, hoone aadress, maja number, korteri number, tehingu liik (kas tegu oli ostu-müügi, enampakkumise, kinke või mõne muu tehinguga – kasutatavate andmete puhul ost-müük), objekti liik (kinnistu korteriomand, hoonestusõigus või korterihoonestusõigus - kasutatavate andmete puhul kõik korteriomandid), omandiosa suurus, tehingu aeg (kuupäev), tehingu summa, käibemaksu sisaldumise märge (sisaldas või ei sisaldanud), katastritunnus, registriosa number, kinnistute arv, kinnistu pindala, kinnistu hoonestatuse märkus (vaadeldaval juhul kõik hoonestatud, kuna tegemist korteritega), korteri pindala, koormatiste olemasolu, müüjate arv ja liik (eraisik, juriidiline isik, välismaalane), ostjate arv ja liik (eraisik, juriidiline isik, välismaalane), maa sihtotstarve ja märkuste lahter.

Kuna Maa-ameti tehinguinfo sisaldab Andmekaitse Inspektsiooni hinnangul isikuandmeid, ei saa töö autor loodavate mudelite aluseks olevaid andmeid käesoleva töö lisana tutvumiseks esitada vaid esitab kokkuvõtted andmetest ning analüüsi tulemused.

Kasutatavate andmete kokkuvõtte ülaltoodud info alusel:

- 5% tehingutest sisaldasid käibemaksu;

- 13% müüjatest olid juriidilised isikud;
- 10% ostjatest olid juriidilised isikud;
- 7% tehingutest asusid vähemalt osaliselt ärimaal;
- 9% tehingutest teostati uute korteritega

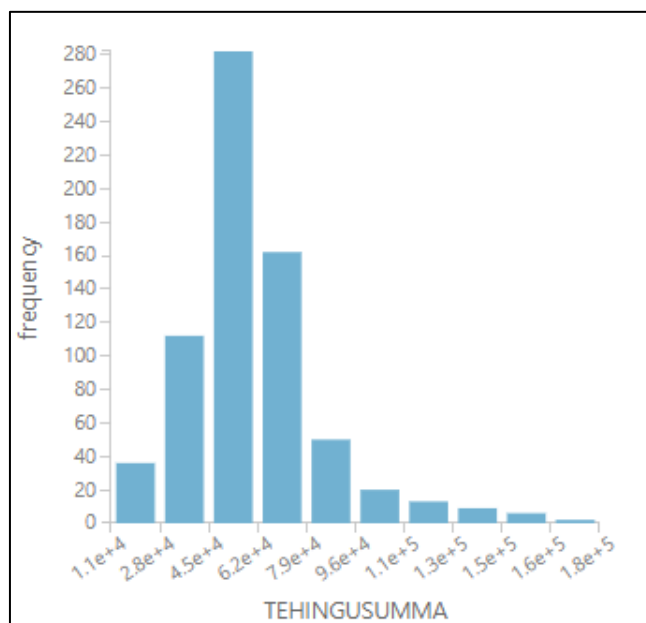
Järgnevas tabelis on toodud statistilised näitajad prognoositava tehingu hinna põhjal.

Tabel 1. Andmete statistilised näitajad „Tehingu summa“ põhjal (eurot).

Näitaja	Turuväärtus €
Aritmeetiline keskmine	60 600
Mediaan	57 500
Minimaalne	11 000
Maksimaalne	180 000
Standardhälve	23 953
Unikaalseid väärtusi	261

Allikas: Autori arvutused

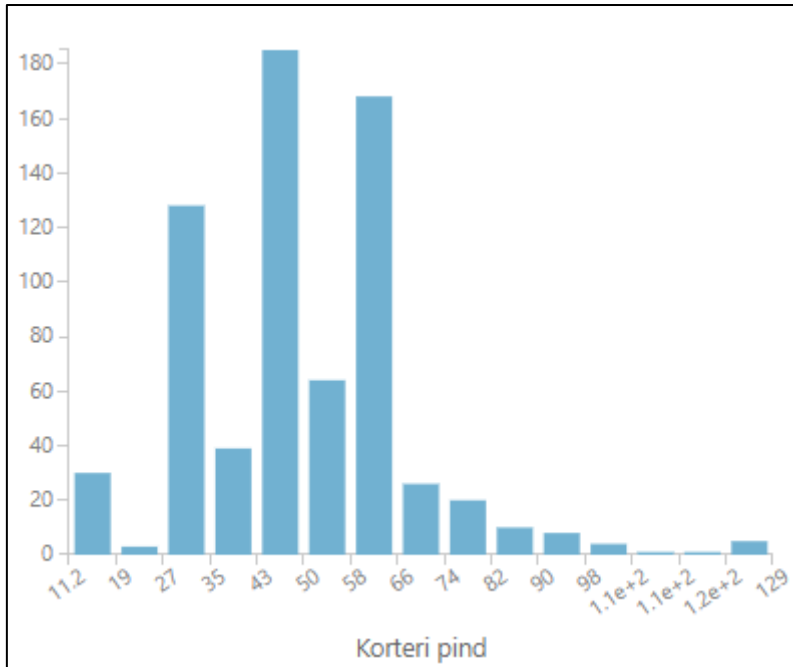
Järgneval joonisel (vt Joonis 3) on näidatud andmetes oleva tehingu summa varieeruvus ning jaotumine (tehingusumma kümnetes tuhandetes eurodes, sagedus (*frequency*) esinemise kordadena).



Joonis 3. Andmete kirjeldus „Tehingu summa“ põhjal

Allikas: Autori arvutused

Järgneval joonisel (vt Joonis 4) on näidatud andmetes oleva korteri pinna varieeruvus ning jaotumine (korteri pind ruutmeerites, sagedus esinemise arvuna).



Joonis 4. Korteri pinna suuruse jagunemine andmetes

Allikas: Autori arvutused

2.3. Kasutatav tarkvara

Autori poolt mudelite loomiseks kasutatav tarkvara on Microsoft Azure Machine Learning Studio nimeline masinõppe mudelite loomise veebirakendus, mis lubab ilma programmeerimiskeelt kasutamata luua ja võrrelda erinevaid mudeleid. Antud tarkvara eelis võrreldes statistilistele mudelitele suunatud programmidega on võimalus luua ka oluliselt keerulisemaid masinõppe mudeleid, mis on üks selle töö eesmärke. Eelis keerulisemate masinõppe mudelite loomiseks tavapäraselt kasutatavate programmide ees on, et Azure tarkvara ei nõua märkimisväärselt programmeerimiskoodi kirjutamist. Nimelt luuakse keerulisemaid mudeleid enamasti mõnes programmeerimiskeeles (levinuim Python), mis nõuaks laialdasi teadmisi programmeerimisest ja matemaatikast ning tähendaks ka loodud programmi kontrollivajadust, mis ei ole käesoleva töö eesmärk. Valitud tarkvaras on mudelite programmeerimistöö professionaalide poolt ära tehtud

ning kontrollitud, kasutaja peab oskama vastavalt vajadusele neist „legoklotsidena“ rakenduse luua.

Kasutatava programmi miinuseks on, et mudelit üles ehitades ei saa seda lõpmatult mugandada ja muuta ning ei ole võimalik luua mudel-mudeli sees lahendusi.

2.4. Valitud mudelid

Erinevates riikides loodud mudeleid analüüsides on selge, et mudelite kasutusvõimalus ja täpsus sõltub kasutada olevatest andmetest, seetõttu ei saa kindlalt väita, et üks mudel on alati parem kui teine. Seetõttu on autor loonud mitmeid mudeleid, et näha nende sobivust Eestisse ja sobivust Eestis kasutada olevate ametlike andmetega (Maa-ameti tehinguinfoga) turuväärtuse hindamisel. Et mudeleid võrrelda, soovis töö autor kindlasti ühe mudelina kasutada klassikalist regressioonianalüüsi, millega on mõningate andmehulkade puhul saadud isegi paremaid tulemusi kui keeruliste meetoditega (Valier, 2020). Kindlasti soovis töö autor analüüsida keerulisemate mudelite, nagu närvivõrgud, toimimist, mis on tihti saavutanud kinnisvara hindamisel väga häid tulemusi. Seega on analüüsi võetud kaks erinevat keerukuse ja sügavusega närvivõrkude mudelit, et näha, kas varjatud kihtide lisamine ning keerukuse tõstmine annab eelise vähese infoga (ilma seisukorrainfoga) tingimustes. Lisaks otsustas autor võrdluseks lisada otsustusmetsa ning Bayesi regressiooni algoritmid. Ka otsustusmets on mudel, mis on mõnes uuringus andnud paremaid tulemusi kui närvivõrgud, seda just juhul kui andmed on puudulikud ja tavalise lineaarse regressiooni jaoks halvasti sobivad (Antipov & Pokryshevskaya, 2012).

Väga keerulisi mudeleid ei loodud kahel põhjusel: esiteks ei ole käesoleva töö eesmärk luua väga keerulist, mudel-mudelis süsteemi, mis otsiks lisaandmeid internetist, kategoriseeriks tehinguinfot ka piltide ja teksti järgi või muud sellist, vaid tuvastada, kas Maa-ameti info põhjal on võimalik piisavat täpsusklassi saavutada. Teiseks ei võimalda kasutatav tarkvara ilma programmeerimiskeelt mahukalt kasutamata mudel-mudelis süsteeme luua.

Mudelite võrdlemiseks kasutab autor erinevaid statistilisi näitajaid. Käesoleval juhul viite erinevat näitajat loodud mudelite omavaheliseks võrdlemiseks ning parima välja selgitamiseks. Parima välja selgitamiseks kasutab autor nn astakute süsteemi, kus järjestab eraldi iga näitaja alusel mudelid paremusjärjestusse ja seejärel liidab kõik järjenumbrid, et saada lõplik järjestus.

Lisaks võrdleb autor parimat mudelit teistes teadustöodes leitud tulemustega sarnaste mudelite loomisel. Selleks kasutab autor keskmist suhtelist absoluutviga protsentides (*Mean absolute percentage error*) ehk lühendina MAPE.

Mudelite võrdlusel kasutatavad erinevad näitajad koos valemitega on toodud järgnevas tabelis.

Tabel 2. Mudelite võrdlemiseks kasutatavad statistilised näitajad ja nende valemid (valemite tähistus: A_j – tegelikud väärtused, \bar{A} – tegelike väärtuste keskmine, P_j – prognoositavad väärtused, $e_j = A_j - P_j$ – viga, n – väärtuste arv)

Tähis	Inglisekeelne nimetus	Eestikeelne nimetus	Valem
MAE	Mean absolute error	Keskmine absoluutviga	$MAE = \frac{1}{n} \sum_{j=1}^n e_j $
RMSE	Root mean squared error	Juuritud keskmine ruutviga	$RMSE = \sqrt{\frac{\sum_{j=1}^n e_j^2}{n}}$
RAE	Relative absolute error	Suhteline absoluutviga	$RAE = \frac{\sum_{j=1}^n e_j }{\sum_{j=1}^n A_j - \bar{A} }$
RSE	Relative squared error	Suhteline ruutviga	$RSE = \sqrt{\frac{\sum_{j=1}^n e_j^2}{\sum_{j=1}^n A_j - \bar{A} }}$
R ²	Coefficient of determination	Determinatsioonikordaja	$1 - \frac{\sum_{j=1}^n (P_j - A_j)^2}{\sum_{j=1}^n (A_j - \bar{A})^2}$
MAPE	Mean Absolute Percentage Error	Keskmine suhteline absoluutviga	$MAPE = \frac{100}{n} \sum_j \frac{ e_j }{ A_j }$

Allikas: Autor, (Botchkarev, 2018)

3. EESTISSE SOBIV MUDEL TURUVÄÄRTUSE HINDAMISEKS

Kuna kirjandus ei ütle üheselt, millised on parimad mudelid, vaid mudeli sobivus sõltub kasutada olevatest andmetest, loob autor testimiseks mitmeid mudeleid. Käesoleva töö empiirilises osas esitab töö autor kirjelduse loodud kinnisvara turuväärtuse hindamise mudelitest, omavahelistest erinevustest, mudelite võrdluse ning tulemused. Esmane andmetöötlus ja kitsendused viidi läbi enne mudeli koostamist. Edasised valikud parameetrite osas ning kitsendused rakendati juba mudeleid analüüsides.

3.1. Mudelite loomine

Enne andmete kasutamist viis autor läbi andmete analüüsi, et eraldada kasutatavad ja mittekasutatavad andmed ning eemaldada vigased andmed:

1. Andmed sisaldavad ka mitte vabaturutehinguid (seotud osapoolte vahelised tehingud, sugulaste vahelised tehingud, varade liigutamine sama omaniku ettevõtete vahel ja sarnased). Kui see info on tehingu märkuste lahtris ära toodud on tehing analüüsiks kasutatavatest andmetest eemaldatud.
2. Analüüsi eesmärgi vaates on selgelt ebaolulised tehingu identifitseerimiseks vajalikud andmed nagu tehingu kood, aadress, korteri number, katastritunnus, registrios, maja number, ostjate ja müüjate arv – see info on samuti analüüsiks kasutatavatest andmetest eemaldatud.
3. Analüüsis kasutatavatest andmetest on välja jäetud ka andmed, mis vaadeldavas kogumis üksteisest ei erine – objekti liik (kõik on korteriomandid), tehingu liik ja iseloom (kõik ostu-müügi tehingud), kinnistu hoonestatus (kõik hoonestatud), kinnistute arv (kõigil üks), ning info mis ei ole täielikult sisestatud – koormatiste olemasolu.
4. Analüüsis kasutatavad tekstilised väärtused on kodeeritud kvalitatiivseteks väärtusteks: linnaosa, asum, müüja tüüp (juriidiline või füüsiline isik – andmetes tähistatud „Muuja“),

- ostja tüüp (juriidiline või füüsiline isik – andmetes tähistatud „Ostja“), ärimaa sisaldus (sisaldab või ei sisalda), käibemaksu sisaldus (sisaldab või ei sisalda – andmetes tähistatud „KM“) ning märkus „uus“ (kas teadolevalt on tegemist uue korteriga või mitte).
5. Kuna asumeid on kasutatavates andmetes rohkem kui kaks kuid mõningate mudelite sisendiks sobivad kvalitatiivsed väärtused vaid on/ei ole kujul, siis on Asumi väärtused teisendatud kõik sellele kujule, millega tekkis iga asumi kohta üks uus andmetulp.
 6. Lisaks on kasutatavate andmete põhjal arvatud välja iga tehingu ruutmeetrihind ning omandiosa suurus on välja toodud hoone suurus ruutmeetrites, kus müüdüd korter asub.

Analüüsis kasutatavateks andmeteks jäävad seega esialgu: linnaosa, asum, hoone suurus, tehingu toimumise kuupäev (andmetes „Kuupäev“), hoonealuse kinnistu pindala (andmetes „Krunni suurus“), korteri pind, uue korteri märged (andmetes „Uus“), müüja ja ostja tüüp (andmetes „Muuja“ ja „Ostja“), ärimaa sihtotstarbe sisalduse märged, käibemaksu märged (andmetes „KM“), tehingu hind ruutmeetrihinna ning koguhinnana.

Peale mitteturutehingute eemaldamist jäi alles 693 tehingut. Esialgu plaanis autor kasutada muutuja väärtusena hinda ruutmeetri kohta kuid esimesed loodud mudelite kontrollid näitasid, et olulisemalt täpsemad tulemused saab kasutades muutujana väärtust „Tehingu hind“ ehk koguhinda.

Mudelite toimimisel olid olulised järgmises tabelis toodud vastastikuse informatsiooni korrelatsioonikoefitsiendi põhjal toodud tegurid.

Tabel 3. Statistilise olulisuse koefitsient teguritel (koefitsient mõõtühikuta)

Tegur	Koefitsient
Tehingu summa	1.000
Korteri pind	0.428
Uus	0.137
Hoone suurus	0.117
Krundi suurus	0.109
KM	0.077
Muuja	0.060
Asum-17	0.044
Kuupaev	0.041
Arimaa	0.027
Asum-5	0.022
Ostja	0.017
Asum-11	0.014
Asum-4	0.012
Asum-18	0.012
Asum-9	0.010
Asum-2	0.010
Asum-7	0.009
Asum-20	0.008
Asum-6	0.008
Asum-10	0.006
Asum-3	0.006
Asum-19	0.005
Asum-1	0.005
Asum-8	0.003
Asum-16	0.002

Allikas: Autori arvutused

Olulisuse põhjal eemaldati esialgu kõik alla 0,005 olulisusnivooga tegurid ja testiti mudeleid. Erinevate korrelatsioonitestide tulemuste põhjal on eemaldatud käibemaksu märke, mis on väga otseselt seotud uue korteri märkega, lisaks muutus eelpoolnimetatud tegurite eemaldamisega väga vähe oluliseks ka tehingu kuupäeva info, mis sel põhjusel samuti eemaldati. Kuna närvivõrkude mudelid töötavad ainult väikeste numbritega, oli lisaks vajalik kõigi numbriliste väärtuste, peale korteri ruutmeetrite, läbijagamine 1000-ga. Samuti lõigati kasutatavatel andmetel ära tehinguhinna ekstreemumid 1% ulatuses tipust ja 1% ulatuses põhjast.

Autori poolt kasutatavates andmetes on statistiliselt kõige olulisemad tegurid korteri ruutmeetrid ja uut korterit näitav tegur, mis mõningal määral asendab seisukorrainfot.

3.2.Loodud mudelite kirjeldus

Et mudeleid võrrelda ja leida parim võimalik mudel, lõi autor erinevaid automaatseid kinnisvara turuväärtuse hindamise mudeleid. Kindlasti soovis töö autor ühe mudelina kasutada klassikalist regressioonanalüüsi. Analüüsimaks keerulisemate mudelite, nagu närvivõrgud, toimimist, on analüüsi võetud kaks erineva keerukuse ja sügavusega närvivõrkude mudelit ning lisaks otsustas autor võrdluseks lisada otsustusmetsa ning Bayesi regressiooni algoritmid.

Mudeleid on kokku viis ja need on järgmised:

1. Lineaarne regressioon (*Linear Regression*);
2. Bayesi regressioon (*Bayesian Linear Regression*);
3. Otsustusmets (*Decision Forest Regression*);
4. Närvivõrk (*Neural Network Regression*);
5. Süvanärvivõrk (*Deep Neural Network Regression*).

Mudelite sisendid andmete osas olid kõigil loodud mudelitel samad – kasutatavad andmed on jagatud juhusliku valiku printsiibil 70% treening-grupiks ning 30% suuruseks testgrupiks. Treenimiseks mõeldud andmetega treenib tarkvara mudelit ära tundma väärtust mõjutavate tegurite panuseid turuväärtusesse ning loob sellega mudeli. Järgmises etapis kasutab tarkvara saadud algoritmi tulemuste prognoosimisel testgrupi andmete põhjal, mida mudel ei ole enne näinud. Sellest testimisest tekivad mudeli statistilised hinnangud täpsuse ning usaldusväärsuse kohta – mida lähemal on iga üksik mudeli prognoositav väärtus reaalsele testgrupi tehingu hinnale, seda täpsem on mudel. Samad etapid läbivad kõik loodud mudelid paralleelselt (skeem mudelite toimimisest on toodud lisa 1), misjärel koondab tarkvara kõigi mudelite leitud tulemuste põhjal tekkinud statistilised näitajad võrdlustabelisse, kus erinevaid mudeleid on võimalik erinevate näitajate alusel omavahel võrrelda.

Lineaarne regressioon (*Linear Regression*)

Kasutatud tavalist vähimruutude meetodit programmi standardseadetes.

Bayesi regressioon (*Bayesian Linear Regression*)

Kasutatud mudelit programmi standardseadetes.

Otsustusmets (*Decision Forest Regression*)

Programmil on lastud valida parimad parameetrid tulemise saamiseks, mille tulemusena loob programm 8 otsustuspuud, maksimaalse sügavusega 32, okste jagunemisega 128 korda.

Järgneval joonisel (vt Joonis 5) on näide mudeli poolt loodud otsustumetsa ühest otsustuspuust.



Joonis 5: Näide loodud otsustumetsa otsustuspuust, puu nr 2.

Allikas: Autori loodud mudeli väljavõte

Närvivõrk (*Neural Network Regression*)

Loodud lihtne ja standardne kolmekihiline närvivõrk ühe varjatud kihiga. 1. kiht sisendkiht 13 sõlmega, 2. varjatud kiht 100 sõlmega ja 3. kiht väljundkiht 1 sõlmega. Kõik kihid on täielikult omavahel ühendatud, õpisamm 0,005.

Süvanärvivõrk (*Deep Neural Network Regression*)

Programmi piiranguid arvestades on loodud R-programmeerimiskeele abil 5-kihiline närvivõrk kolme varjatud kihiga. Autor katsetas erinevate varjatud kihtide (2-4) arvu ja erinevate sõlmede arvuga kihtides (3-100 sõlme) kuid arvestades, et närvivõrkude erinevate lahenduste vahel on leitud vähemate andmemahutude juures kasutegurit väiksema neuronite arvuga mudelites (Mach, 2017) jõudis autor järgmise lahenduseni: 1. kiht on sisendkiht 13 sõlmega, 2. 3. ja 4. kiht on varjatud kihid, vastavalt 9, 7 ja 5 sõlmega ning 5. kiht on väljundkiht ühe sõlmega.

Kõigi mudelite skemaatiline ülesehitus on toodud Lisas 1.

3.3. Parim mudel

Mudelite tulemused on kokkuvõtvalt toodud järgnevas tabelis.

Tabel 4. Mudelite tulemused ja paremusjärjestus, kus MAE - keskmine absoluutviga, RMSE - juuritud keskmine ruutviga, RAE - suhteline absoluutviga, RSE - suhteline ruutviga, R^2 - determinatsioonikordaja (ühikuta suhtarvud).

Algoritm	MAE	Astak	RMSE	Astak		Astak		Astak		Astak		Parim (väikseim parim)
				2	RAE	3	RSE	4	R^2	5		
Süvanärvivõrk	70.828	4	96.798	4	0.461	4	0.192	4	0.808	4	20	
Närvivõrk	58.209	1	84.633	1	0.379	1	0.147	1	0.853	1	5	
Bayesi regressioon	67.239	2	95.911	3	0.437	2	0.188	2	0.812	2	11	
Lineaarne regressioon	67.746	3	96.514	2	0.441	3	0.191	3	0.809	3	14	
Otsustusmets	72.075	5	98.347	5	0.469	5	0.198	5	0.802	5	25	

Allikas: Autori arvutused

Kuna kasutada on erinevad näitajad mudelite võrdlemiseks, on kasutatud süsteemi, kus erinevate näitajate kohta loodi paremusjärjestus astakute kaudu ja nende põhjal tekitati kokkuvõttev paremusjärjestus. Tabelist on näha, et kõige kehvema tulemuse annab otsustusmetsa mudeli kasutamine.

Parimaks mudeliks on tabelist selgelt näha Närvivõrgu mudel. Heade mudelitena paistavad välja veel kaks suhteliselt võrdset mudelit. Nendeks on lineaarne ja Bayesi regressioon.

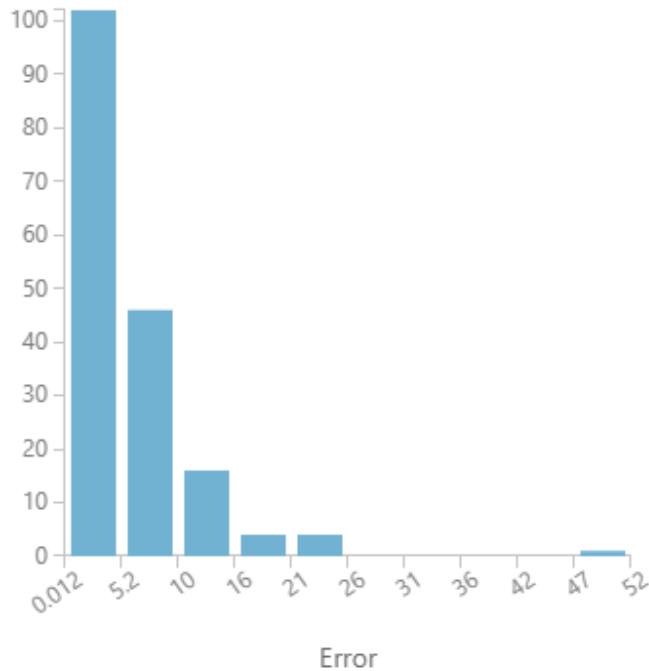
Mudelite täpsuse kontrollimiseks kasutab autor lisaks eelnevatele tabelis toodutele ka keskmist suhtelist viga, põhjusel et keskmine suhteline viga näitab, kas mudel ala- või ülehindab väärtusi süstemaatiliselt. Keskmist suhtelist viga arvutatakse prognoositud väärtus jagatud tegeliku väärtusega.

Keskmine suhteline viga närvivõrgu mudeli puhul on +4%, ehk mudel hindab keskmiselt 4% üle, ning Bayesi regressiooni puhul +3%, mis on väga hea tulemus.

Mudeli kirjeldusvõime (determinatsioonikordaja) on parimatel mudelitel üle 0,8, närvivõrgul lausa 0,85, mis näitab, et mudel kirjeldab ära üle 80% hinda mõjutavatest teguritest.

Vigade suuruste graafikut (näitena närvivõrk) vaadates selgub, et testgrupi andmete 173 tehingu hulgast oli üle 100 prognoositava väärtuse viga alla 5 200 € ja 86% juhtudest alla 10 000 €. Arvestades, et keskmine tehingusumma oli ca 60 000 €.

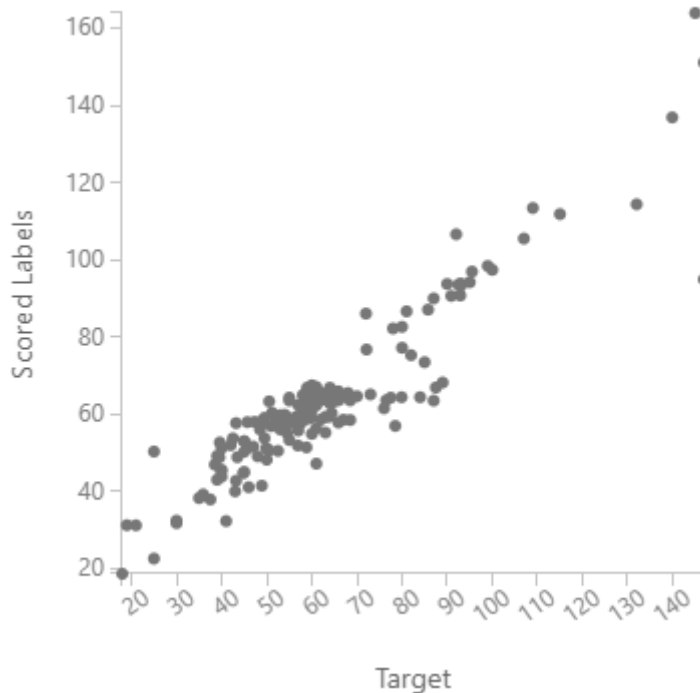
Järgneval joonisel (vt. Joonis 6) on toodud vigade jaotussagedus (horisontaalteljel vigade (Error) suurus kümnetes tuhandetes eurodes ning vertikaalteljel vigade esinemise arv)



Joonis 6. Närvivõrgu mudeli vigade jaotusgraafik

Allikas: Autori arvutused

Pannes järgneval joonisel (vt. Joonis 7) ühele graafikule tegelikud ja närvivõrgu poolt prognoositud väärtused on näha, et suurim erinevus tekib kõige kallimate tehingute osas (horisontaalteljel tegelikud tehinguhinnad (Target) vertikaalteljel prognoosid (Scored Labels), mõlemad tuhandetes eurodes).



Joonis 7. Närvivõrgu prognooside (Scored Labels) ja tegelike tehinguhindade (Target) graafik.

Allikas: Autori arvutused

Lineaarse regressiooni puhul saab ära tuua ka turuväärtuse arvutamise valemi, ehk mudeli lõpliku kuju:

$$Y = 193,11 \times \text{Uus}_1 - 116,71 \times \text{Uus}_0 + 75,53 \times \text{Ostja}_0 + 45,34 \times \text{Muuja}_1 + 40,49 \times \text{Arimaa}_1 + 35,92 \times \text{Arimaa}_0 + 34,69 \times \text{Asum}_{18} + 31,07 \times \text{Muuja}_0 - 26,79 \times \text{Asum}_5 + 24,42 \times \text{Asum}_{11} - 9,01 \times \text{Asum}_{17} + 8,96 \times \text{Korteri pind} + 8,19 \times \text{Asum}_4 + 0,87 \times \text{Ostja}_1 + 0,80 \times \text{Krundi suurus} - 0,02 \times \text{Hoone suurus} + 76,4$$

3.4.Järeldused

Kõige olulisem järeldus on see, et Eestis on võimalik ametliku tehinguinfo baasil automaatset kinnisvara hindamise mudelit luua ja seda erinevate lahendustega.

Mudeli headust vaatleb autor kahest aspektist:

1. Milline on mudeli täpsus võrreldes teiste loodud mudelitega ning korterite hindamisel standardkohase eksperthinnanguga kasutatava täpsusklassiga, mis on praktikas korterite puhul tihti 10%.

2. Milline on mudeli täpsus võrrelduna teadustöodes saadud tulemustega sarnaste mudelite kasutamisel.

Autori loodud parimat automaatset kinnisvara turuväärtuse hindamise mudelit vaadeldes nähtub, et mudel suudab leida korterite turuväärtusi kiiresti, efektiivselt ja ilma suure panuseta hindaja poolt, seega on väga efektiivne, mis ongi automaatsete hindamismudelite kasutamise põhjus. Mudel suudab hinnata homogeensel turul, nagu seda on Tallinna Mustamäe ja Lasnamäe linnaosad, korteri turuväärtust keskmise suhtelise veaga +3% ning keskmise absoluutveaga alla 10%. Vigu täpsemalt vaadeldes on vaid 14% juhtudest viga suurem kui 20% ning vaid 1% puhul on viga suurem kui 50%, mida võib pidada autori arvates väga heaks tulemuseks.

Erinevates avaldatud teadustöodes on teistes riikides automaatseid kinnisvara turuväärtuse hindamise mudeleid korduvalt loodud ning testitud. Autor kasutab teadustöodes leitud tulemusi enda poolt loodud mudelite täpsuse kohta võrdlusena, nägemaks, kas loodud kinnisvara hindamise mudel on hea või mitte.

Järgnevalt on toodud tabel teadustöödest, kus on käesoleva tööga sarnaseid mudeleid loodud ning arvatud keskmine suhteline absoluutviga, protsentides (MAPE). Näitaja MAPE on võrdluseks valitud seetõttu, et see on teadustöodes enim kasutatud näitaja. Autori loodud parima mudeli (Närvivõrk) MAPE on 11%.

Tabel 5. Parima mudeli keskmise suhtelise absoluutvea (MAPE) võrdlus teadustöödega.

Viide teadustööle	MAPE
(Giudice et al., 2017)	6,6%
(Kim et al., 2020)	4,9 - 7,7%
(Guan et al., 2011)	10%
Käesolev töö	11%
(X. Zhou et al., 2019)	12%
(Kok et al., 2017)	14%
(Alexandridis et al., 2019)	16,5%

Allikas: Autori arvutused

Võrdlus näitab, et loodud mudeli keskmine suhteline absoluutviga on samas suurusjärgus, mis teadustöodes on saavutatud ning loodud mudelit võib seega pidada hea täpsusega mudeliks.

Võrreldes kõiki loodud mudeleid vigade ja determinatsioonikordaja põhjal tuleb välja, et kasutatud andmetega ei ole keerulisematel mudelitel suurt eelist lihtsamate mudelite ees. Eelis küll on, ent mitte väga suur – ka lihtne lineaarne regressioon andis sarnaseid tulemusi ning ei ole vigade ja determinatsioonikordaja poolest väga palju kehvem kui närvivõrkude mudel. Samuti ei suutnud süvanärvivõrk teha täpsemaid prognoose kui lihtsam, ühe varjatud kihiga närvivõrk ning on tulemuste poolest isegi kehvem kui lineaarse regressiooni mudel. Sama tulemuseni on jõutud ka teadusartiklites (Mach, 2017) ja (Guan et al., 2011), leides et keerulisemate mudelite paremus tuleb välja suurte andmehulkade ja paljude andmeväljade puhul. Autori hinnangul võiks ka Eesti tingimustes keerulisemad mudelid toimida paremini, kui lisada tegureid, mida Maa-ameti info ei sisalda, nagu päringud ehitisregistrist jms.

Vaatlusalustes korterelamupiirkondades kompenseerib seisukorrainfo puudumist tegur, mis ütleb, kas on tegemist uue korteriga või mitte. Välja töötatud mudeli puhul see lahendus töötab ning selle teguri ning teiste tegurite koosmõjul on siiski võimalik korterite turuväärtust leida.

Seega tuleneb tehtud tööst vastuseks uurimisküsimustele, et:

1. Eesti tingimustes ja Maa-ameti ametlikku infot kasutades on võimalik automaatse hindamismudeli loomine kuid Maa-ameti infot kasutades ei anna keerulisemad mudelid olulist eelist võrreldes lineaarse regressiooniga.
2. Parimate mudelitega on võimalik saavutada üle 80% prognoosimise täpsus homogeensetes korterelamupiirkondades.
3. Seisukorrainfo puudumisel on siiski võimalik luua automaatne hindamismudel.

3.5. Ettepanekud

Autor jõudis tulemuseni, et Eesti piiratud info tingimustes on võimalik turuväärtust automaatselt hindavat mudelit luua. Järgmine samm võiks olla mudeli või mudelite parandamine, täiendamine ning laiendamine. Loodud mudelit ennast saaks veel oluliselt laiendada, lisades andmetesse lisaks tegureid – näiteks hoone ehitusaasta automaatne päring ehisregistrist on tehniliselt võimalik. Samuti tuleks uurida mudeli kasutusvõimalusi laiemalt – kõigepealt pikema ajaperioodi vaates – sisestada mudelisse mitme aasta andmed samade piirkondade kohta ning testida mudeli toimimist. Teiseks luua mudel rohkemate piirkondade vaates – kas ja kuidas toimib mudel vähem homogeensetes linnaosades ning edasi ka teist tüüpi varade vaates kui vaid korterid. Suurim lahendamist vajav probleem teiste kinnisvaraturu segmentide turuväärtuse hindamisel on seisukorrainfo puudus, mis on näiteks eramute puhul veelgi olulisem kui korterite puhul. Samuti ei annaks ka uut ja vana eramut eraldav tegur „Uus“ nii head tulemust kui see andis magalapiirkonna korterite puhul, kuna ka uute eramute turuväärtused võivad kvaliteedist tulenevalt erineda väga suurel määral. Samuti muutuvad oluliseks asukohad asumi sees (vaated, mere lähedus jms), millist probleemi aitaks lahendada geoinfo lisamine mudelisse. Autori arvates on aga tähtis, et ka Eestis tekiks huvi automaatsete hindamismudelite vastu ja hakataks ka reaalsel turul automaatseid hindamismudeleid kasutama, et kaasas käia maailmas toimivate arengutega selles vallas ning võib-olla jõuda ka esirinda.

Edasiseks uurimiseks aitaks kirjanduse põhjal mitmeid probleeme lahendada ka mudel-mudeli sees lahenduse kasutamine (Crosby et al., 2015), mida käesolevas töös ei vaadeldud.

KOKKUVÕTE

Töö eesmärgiks oli leida Eestisse sobivaim mudel turuväärtuse leidmiseks piiratud informatsiooni tingimustes ning testida mudeli täpsust. Eesti tingimustes on teemat vähe uuritud kuid masinõppe kiire arenguga on ka kinnisvara hindamine kindlasti tulevikus muutumas ning uusi meetodeid tuleb katsetada ning kasutusele võtta.

Kinnisvara hindamine on standardiseeritud tegevus ja sel kujul ilmselt täielikult ei kao, kuna professionaali teostatud eksperthinnang tagab läbipaistvuse, kontrollitavuse, selguse, standardiseeritud tegevused ja ühetaolise tulemuse. Tihti on aga kuluv aeg ning töö hind liiga suured, et sellist lahendust kasutada. Juba on välja töötatud ka automaatsete hindamiste esimesed standardid ning täpsuse ja läbipaistvuse suurenemisel võtavad automaatsed mudelid ilmselt mitmed hindamistöö aspektid ja tüüphindamised üle.

Automaatsete hindamismudelite loomisel tuleb lähtuda alati kättesaadavast infost, mis on igas riigis erinev. Seetõttu ei ole välja kujunenud ühte ja parimat mudelit vaid erineva info tingimustes on parimad ka erinevad mudelid. Seetõttu vaatles autor mitmeid kirjanduses häid tulemusi saavutanud mudeleid ning tõi need Eesti konteksti.

Autor lõi 5 erinevat hindamismudelit:

1. Lineaarne regressioon (*Linear Regression*);
2. Bayesi regressioon (*Bayesian Linear Regression*);
3. Otsustusmets (*Decision Forest Regression*);
4. Närvivõrk (*Neural Network Regression*);
5. Süvanärvivõrk (*Deep Neural Network Regression*).

Teadusartiklite analüüs indikeeris, et ka Eesti tingimustes võiks olla võimalik kasutada kõiki loodud mudeleid. Kõik loodud mudelid töötavadki ja annavad parema täpsusklassi kui $\pm 50\%$. Loodud viie mudeli puhul olid otsustusmetsa ning süvanärvivõrgu mudelite tulemused märgatavalt kehvemad kui teistel.

Parima tulemuse andis lihtsama närvivõrgu mudeli kasutamine. Seega sobib kasutatavate andmetega kõige paremini närvivõrk. Samas näitavad tulemused, et tavalise lineaarse regressiooni mudel ei jäänud parimale mudelile suurel määral alla. Sellest järeldub, et mudelis kasutatavad ametlikud Maa-ameti tehinguandmed on piisavalt väikese mahuga, et keerulisemate mudelite kasutamine ei anna olulist eelist ning süvanärvivõrgu, kui kõige keerulisema mudeli, kasutamine andis hoopis kehvema tulemuse kui mõne lihtsama mudeli kasutamine.

Vastuseks uurimisküsimustele saab järeldada, et:

- Automaatset hindamismudelit on võimalik Eesti ametliku info pealt luua.
- Parim mudel on närvivõrgu mudel, kuid keerulisemad ja vähem läbipaistvad mudelid ei oma vaid ametliku infot kasutades väga suurt eelist oluliselt lihtsamate ja läbipaistvamate mudelite ees.
- Automaatse hindamismudeliga on võimalik saavutada 86% tõenäosusega viga alla 17% üksikobjekti hinnates ning alla 5% viga suuremat portfelli hinnates. Ka Eesti piiratud infoga tingimustes on võimalik saavutada sarnane täpsusklass, mida on saavutatud teadustöodes üle maailma.
- Seisukorrainfo puudumine ei takista automaatse hindamismudeli loomist ja turuväärtuse leidmist, kuid mudel vajab selleks võimalikult palju teisi argumente, mis aitavad seisukorrainfot tuletada.

Edasiseks uurimiseks saab luua keerulisemaid mudeleid, kus kasutatakse rohkem sisendeid kui vaid ametlikku tehinguinfot, mille tulemusena on tõenäoline paremate tulemuste saamine, kuna suuremate andmehulkade puhul töötavad keerulisemad mudelid paremini kui lihtsad. Samuti tuleks uurida mudel mudeli sees lahendust ning mudelite loomist teiste piirkondade, pikemate ajaperioodide ning teiste kinnisvaraturu segmentide põhjal.

Kõige olulisem on autori arvates aga hakata ka reaalsel turul automaatseid hindamismudeleid kasutama, et mitte jääda maailmas toimuvatest arengutest maha.

SUMMARY

AUTOMATED REAL ESTATE VALUATION MODEL USAGE POSSIBILITIES IN ESTONIA

Raigo Petter

Real estate appraisal is a standardized field of activity, there are many requirements, and it takes a considerable amount of time, requires knowledge, a large amount of information and needs an expert, all of which makes real estate appraisal costly and time consuming.

However, there are situations where a valuation accordance to the standard is not necessary, but low cost and speed are essential. Such situations may include, but are not limited to, obtaining an overview of the portfolio, assessing the loan portfolio, assessing potential losses or gains, making investment decisions, and so on.

Machine learning is being introduced in more and more areas of life. The field of real estate appraisal has not remained untouched and various automatic real estate appraisal models have been created, that give fast and cheap evaluations. Such models assess real estate with little or no human intervention, they only require some inputs on the property under valuation. There are different models, but those based on machine learning are considered to be the most promising (Valier, 2020).

Many countries and companies already use various machine-based property valuation models, the best known one is company named Zillow in the US, which has developed an model capable to find real estate market value with + -20% accuracy in 99% of US cities and an average of 83% nationwide (GA Matysiak & TEGOVA, 2018).

In Estonia, the topic has not been studied in terms of machine learning, so there is no information on whether and to what extent it would be usable in Estonian conditions. As can be seen from almost all literature sources, the availability of automated estimation models is directly related to the amount and quality of data, so the possibilities for use of automated models vary from country to country or may not be usable at all. The issue is relevant because the development of machine learning and artificial intelligence has reached a stage where it significantly changes the nature of

many professions or eliminates them altogether. Real estate valuation is also one of those affected (World Economic Forum, 2020)

Due to the above, the research problem is risen: is it possible to use automatic valuation model in the conditions of the small Estonian real estate market and using only official transaction information.

The aim of the work is to find the most suitable model for finding market value in Estonia, considering limited information availability and to test the accuracy of the model.

In order to achieve this goal, the author of the work has raised the following research questions:

1. Which evaluation model works best in limited information conditions of Estonia?
2. What accuracy can be achieved with an automatic valuation model in Estonia?
3. Is it possible to find the market value of an apartment with a good model considering the absence of condition information of apartments in Estonian transaction information?

The author of the work created an automatic valuation models that find the market value of an apartment with the best possible accuracy class based on the data of the apartment to be entered. To create the automated assessment model, the author used a web-based machine learning model creation application called Microsoft Azure Machine Learning Studio. As literature shows that such models are not suitable for the evaluation of non-homogeneous objects, author selected two most homogeneous regions in Estonia with sufficient volume of transactions to test the model: Mustamäe and Lasnamäe districts in Tallinn.

The automated evaluation models are always based on available information in current country. Therefore the best model fitting all countries has not been found, but it is proven, that availability of different information proves different models to be the best. Therefore, the author created several models that showed good results in the literature and applied them to the Estonian market.

The author created 5 different evaluation models:

1. Linear Regression;
2. Bayesian Linear Regression;
3. Decision Forest Regression;
4. Neural Network Regression;
5. Deep Neural Network Regression.

The analysis of research articles indicated that even in Estonian conditions it could be possible to use all created models. All created models work and give an accuracy class better than $\pm 50\%$. In the case of the five models created, the results of the decision forest and deep neural network models were significantly worse than in the others.

The best result was obtained using a simpler neural network model. Thus, the neural network is best suited to the data used. However, the results show that the conventional linear regression model did not lag far behind the best model. It follows that the official transaction data of the Land Board used in the model is small enough that the use of more complex models does not provide a significant advantage, and the use of the deep neural Network, as the most complex model, performed much worse than the use of simpler models.

In response to the research questions, it can be concluded that:

- An automatic evaluation model can be created on the basis of official Estonian information.
- The best model is the neural network model, but more complex and less transparent models do not only have a huge advantage over simpler and more transparent models using only official information.
- With an automatic valuation model, it is possible to achieve an error less than 17% of an individual object and error less than 5% error when valuing a larger portfolio in probability of 86%. Despite the limited information conditions in Estonia, it is possible to achieve a similar accuracy class that has been achieved in research around the world.
- The lack of condition information of the apartment does not prevent the creation of an automatic valuation model and the finding of market value, but the model needs as many other arguments as possible to help derive condition information.

As a further development, more sophisticated models can be created that use more inputs than just formal transaction information, which is likely to lead to better results, as more complex models work better than simple ones for larger amounts of data.

KASUTATUD ALLIKATE LOETELU

- Abidoye, R. B., & Chan, A. P. C. (2018). Improving property valuation accuracy : a comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal*, 5921, 1–13. <https://doi.org/10.1080/14445921.2018.1436306>
- Abidoye, R. B., Chan, A. P. C., Abidoye, F. A., & Oshodi, O. S. (2019). Predicting property price index using artificial intelligence techniques: Evidence from Hong Kong. *International Journal of Housing Markets and Analysis*, 12(6), 1072–1092. <https://doi.org/10.1108/IJHMA-11-2018-0095>
- Adamczyk, T. (2010). Application of GAM Additive Non-Linear Models to Estimate Real Estate Market Value. *Geomatics and Environmental Engineering*, 4(2), 55–62.
- Alexandridis, A. K., Karlis, D., Papastamos, D., & Andritsos, D. (2019). Real Estate valuation and forecasting in non-homogeneous markets: A case study in Greece during the financial crisis. *Journal of the Operational Research Society*, 70(10), 1769–1783. <https://doi.org/10.1080/01605682.2018.1468864>
- Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments : An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems With Applications*, 39(2), 1772–1778. <https://doi.org/10.1016/j.eswa.2011.08.077>
- Appraisal Institute, & EKHÜ. (2008). *Kinnisvara hindamine* (13th ed.). Appraisal Institute.
- Augustyniak, H., Laszek, J., Olszewski, K., & Waszczuk, J. (2018). Property valuation for mortgage purposes in Poland. *Property Management*, 36(2), 234–247. <https://doi.org/10.1108/PM-09-2016-0049>
- Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied Sciences (Switzerland)*, 8(11). <https://doi.org/10.3390/app8112321>
- Baxter, J., & Vandell, K. D. (2007). Expanding the academic discipline of real estate valuation: A historical perspective with implications for the future. *Journal of Property Investment & Finance*, 25(5), 427–443. <https://doi.org/10.1108/14635780710776657>
- Bergadano, F., Bertilone, R., Paolotti, D., & Ruffo, G. (2019). Learning real estate automated valuation models from heterogeneous data sources. *ArXiv*, 1, 1–14.
- Bin, J., Gardiner, B., Li, E., & Liu, Z. (2019). Peer-Dependence Valuation Model for Real Estate Appraisal. *Data-Enabled Discovery and Applications*, 3(1). <https://doi.org/10.1007/s41688-018-0027-0>

- Bin, J., Tang, S., Liu, Y., Wang, G., Gardiner, B., Liu, Z., & Li, E. (2017). Regression model for appraisal of real estate using recurrent neural network and boosting tree. *2017 2nd IEEE International Conference on Computational Intelligence and Applications, ICCIA 2017, 2017-Janua*, 209–213. <https://doi.org/10.1109/CIAPP.2017.8167209>
- Botchkarev, A. (2018). Evaluating performance of regression machine learning models using multiple error metrics in Azure Machine Learning Studio. *SSRN E-Journal*, 1–16.
- Brotman, B. A. (1990). Linear and Nonlinear Appraisal Models. *The Appraisal Journal*, 2(58), 249–253.
- Crosby, H., Davis, P., & Jarvis, S. A. (2015). Exploring new data sources to improve UK land parcel valuation. *Proceedings of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics, UrbanGIS 2015*, 32–35. <https://doi.org/10.1145/2835022.2835028>
- Delisle, J. R., Grissom, T. V., & Beach, F. (2020). The big data regime shift in real estate. *Journal of Property Investment & Finance*, 38(4), 363–395. <https://doi.org/10.1108/JPIF-10-2019-0134>
- Dimopoulos, T., & Bakas, N. (2019). Sensitivity analysis of machine learning models for the mass appraisal of real estate. Case study of residential units in Nicosia, Cyprus. *Remote Sensing*, 11(24). <https://doi.org/10.3390/rs11243047>
- Eesti Standard Vara Hindamine EVS 875, Pub. L. No. EVS875 (2017).
- Giudice, V. Del, Paola, P. De, Forte, F., & Manganeli, B. (2017). Real Estate Appraisals with Bayesian Approach and Markov Chain Hybrid Monte Carlo Method : An Application to a Central Urban Area of Naples. *MDPI.Com Sustainability*, 9(2138), 1–17. <https://doi.org/10.3390/su9112138>
- Glumac, B., & Rosiers, F. Des. (2020a). Practice briefing – Automated valuation models (AVMs): their role, their advantages and their limitations. *Journal of Property Investment & Finance*, 7. <https://doi.org/10.1108/JPIF-07-2020-0086>
- Glumac, B., & Rosiers, F. Des. (2020b). Towards a taxonomy for real estate and land automated valuation systems real estate and. *Journal of Property Investment & Finance*, ahead of p(ahead of print). <https://doi.org/10.1108/JPIF-07-2020-0087>
- Guan, J., Levitan, A., & Zurada, J. (2011). A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context. *Journal of Real Estate Research*, 33(3), 349–387.
- Standard on Automated Valuation Models (AVMs) - 2018, 1 (2018).
- International Valuation Standards, (2020).
- Kettani, O., & Oral, M. (2015). Designing and implementing a real estate appraisal system : The case of Quebec Province, Canada. *Socio-Economic Planning Sciences*, 49, 1–9. <https://doi.org/10.1016/j.seps.2014.12.003>
- Kim, Y., Choi, S., & Yi, M. Y. (2020). Applying comparable sales method to the automated estimation of real estate prices. *Sustainability (Switzerland)*, 12(14), 1–19.

<https://doi.org/10.3390/su12145679>

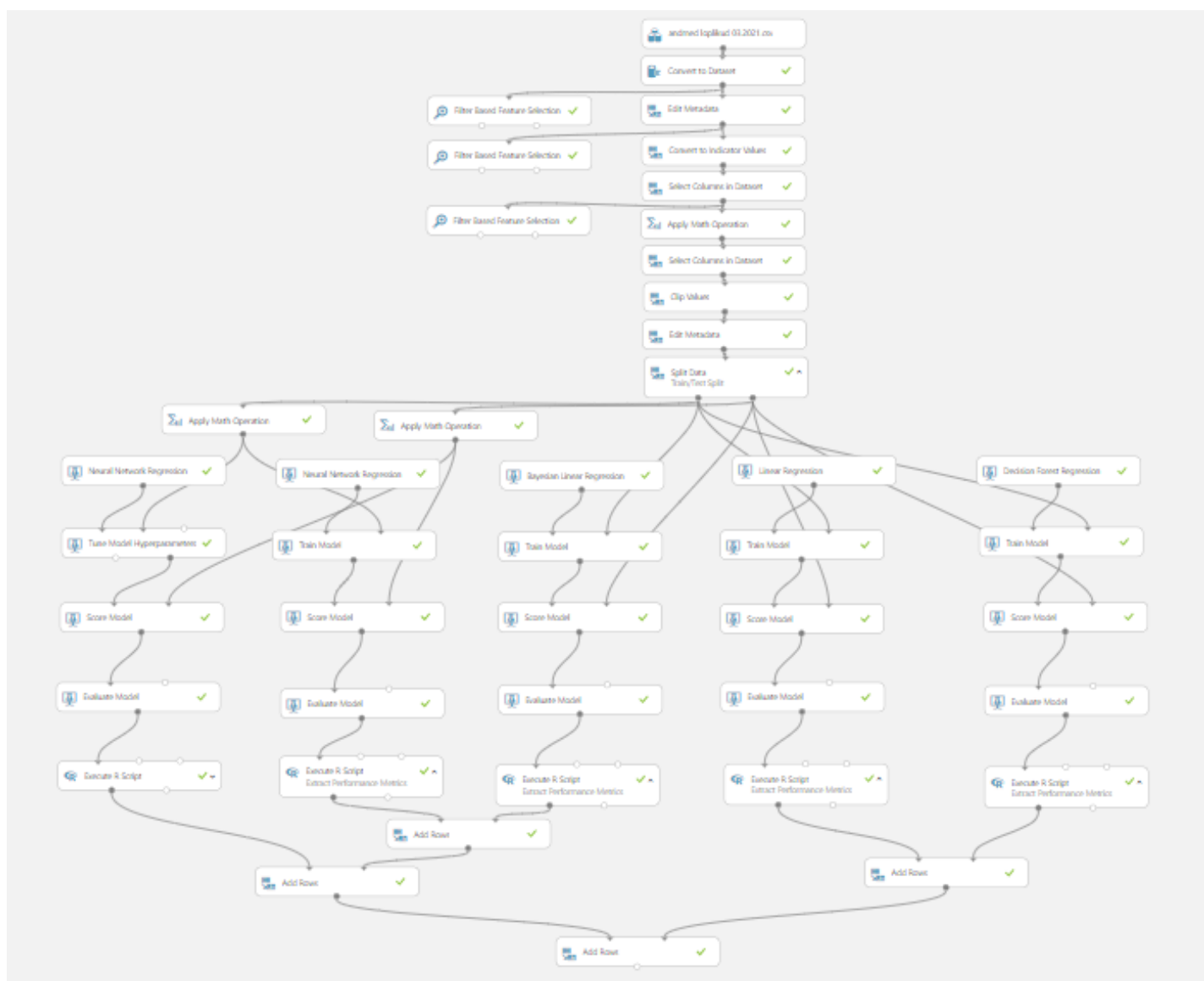
- Kok, N., Koponen, E. L., & Martínez-Barbosa, C. A. (2017). Big data in real estate? from manual appraisal to automated valuation. *Journal of Portfolio Management*, 43(6), 202–211. <https://doi.org/10.3905/jpm.2017.43.6.202>
- Mach, Ł. (2017). The Application of Classical and Neural Regression Models for the Valuation of Residential Real Estate. *Folia Oeconomica Stetinensia*, 17(1), 44–56. <https://doi.org/10.1515/fofi-2017-0004>
- Mackmin, D. (1999). Valuation of real estate in global markets. *Property Management*, 17(4), 353–367. <https://doi.org/10.1108/02637479910299651>
- Matysiak, G. (2017). *Report for TEGoVA The Accuracy of Automated Valuation Models (AVMs)*. https://www.tegova.org/data/bin/a591190c05b2c3_Geoge_Matysiak_Valuation_Report.pdf
- Matysiak, G. A., & TEGOVA. (2018). *Assessing the Accuracy of Individual Property Values Estimated by Automated Valuation Models* (Issue May).
- Moosavi, V. (2017). Urban Data Streams and Machine Learning: A Case of Swiss Real Estate Market. *ArXiv.Org*. <https://arxiv.org/abs/1704.04979>
- Oshodi, O. S., Thwala, W. D., Odubiyi, T. B., Abidoye, R. B., & Aigbavboa, C. O. (2019). Using neural network model to estimate the rental price of residential properties. *Journal of Financial Management of Property and Construction*, 24(2), 217–230. <https://doi.org/10.1108/JFMPC-06-2019-0047>
- Poursaeed, O., Matera, T., & Belongie, S. (2018). Vision-based real estate price estimation. *Machine Vision and Applications*, 29(4), 667–676. <https://doi.org/10.1007/s00138-018-0922-2>
- RICS Valuation Global Standards, (2020).
- Tehnilise normi ja standardi seadus*, 3 (1999) (testimony of Riigikogu). <https://www.riigiteataja.ee/akt/77317#>
- Robson, G., & Downie, M. Lou. (2007). Automated Valuation Models: an international perspective. In *The Council of Mortgage Lenders (CML) Research*. [http://nrl.northumbria.ac.uk/1019/1/Frequency conversion based on three-wave parametric solitons.pdf](http://nrl.northumbria.ac.uk/1019/1/Frequency%20conversion%20based%20on%20three-wave%20parametric%20solitons.pdf)
- Royal Institution of Chartered Surveyors. (2017). *the Future of Valuations* (Issue November). <https://www.rics.org/globalassets/rics-website/media/knowledge/research/insights/future-of-valuations-insights-paper-rics.pdf>
- Sauga, A. (2020). *Statistika õpik majanduseriala üliõpilastele Ako Sauga*.
- Sun, Y. (2019). Real estate evaluation model based on genetic algorithm optimized neural network. *Data Science Journal*, 18(1), 1–9. <https://doi.org/10.5334/dsj-2019-036>
- European Valuation Standards, Pub. L. No. 9/2020 (2020).

<https://www.tegova.org/en/p5fa0280f9d296>

- Valier, A. (2020). Who performs better? AVMs vs hedonic models. *Journal of Property Investment and Finance*, 38(3), 213–225. <https://doi.org/10.1108/JPIF-12-2019-0157>
- Valier, A., & Micelli, E. (2020). Automated models for value prediction : a critical review of the debate. *Valori e Valutazioni*, 24, 151–162.
- Wang, D. (2019). Mass Appraisal Models of Real Estate in the 21st Century : A Systematic Literature Review. *Sustainability*, 11(7006), 1–14.
- World Economic Forum. (2020). *The future of Jobs Report* (Issue October). <https://www.weforum.org/reports/the-future-of-jobs-report-2020/digest>
- Wyman, D., Seldin, M., & Worzala, E. (2011). A new paradigm for real estate valuation? *Journal of Property Investment & Finance*, 29(4), 341–358. <https://doi.org/10.1108/14635781111150286>
- Xiao-sheng, L. I. U., Zhe, D., & Ting-li, W. (2011). Real estate appraisal system based on GIS and BP neural network. *Transactions of Nonferrous Metals Society of China*, 21, s626–s630. [https://doi.org/10.1016/S1003-6326\(12\)61652-5](https://doi.org/10.1016/S1003-6326(12)61652-5)
- You, Q., Pang, R., Cao, L., & Luo, J. (2017). Image-Based Appraisal of Real Estate Properties. *IEEE Transactions on Multimedia*, 19(12), 2751–2759. <https://doi.org/10.1109/TMM.2017.2710804>
- Zhou, G., Ji, Y., Chen, X., & Zhang, F. (2018). Artificial Neural Networks and the Mass Appraisal of Real Estate. *IJOE*, 14(3), 180–187.
- Zhou, X., Tong, W., & Li, D. (2019). Modeling Housing Rent in the Atlanta Metropolitan Area Using Textual Information and Deep Learning. *ISPRS International Journal of Geo-Information*, 8(8), 349. <https://doi.org/10.3390/ijgi8080349>

LISAD

Lisa 1. Loodud mudeli skeem



Lisa 2. Lihtlitsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina Raigo Petter

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Automaatse kinnisvara hindamismudeli kasutusvõimalused Eestis,

mille juhendajad on Ene Kolbre ja Ako Sauga,

1.1 reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2 üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.

3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

11.05.2021

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