

TALLINNA TEHNIKAÜLIKOOL

Majandusteaduskond

Isabel Silla

**ERAISIKU KREDIIDIRISKI HINDAVA MUDELI
KOOSTAMINE ÜHISRAHASTUSPORTAALILE MONEYZEN**

Bakalaureusetöö

Õppekava ärimatus, peeriala äriühendus

Juhendaja: Kristjan Liivamägi, PhD

Tallinn 2023

Deklareerin, et olen koostanud lõputöö iseseisvalt ja olen viidanud kõikidele selle koostamisel kasutatud teiste autorite töödele, olulistele seisukohtadele ja andmetele, ning ei ole esitanud sama tööd varasemalt ainepunktide saamiseks.

Töö pikkuseks on 6009 sõna sissejuhatusest kuni kokkuvõtte lõpuni.

Isabel Silla 20.12.2023

SISUKORD

LÜHIKOKKUVÕTE	4
SISSEJUHATUS	5
1. ÜHISRAHASTUS JA KREDIIDIRISK	8
1.1. Ülevaade ühisrahastusest	8
1.1.1. Ühisrahastusportaal MONEYZEN	10
1.2. Krediidiriski mõjutavad karakteristikud	11
1.3. Krediidiriski arvutamise võimalused	12
1.4. Kitsaskohad krediidiriski hindamisel	14
2. METOODIKA	16
2.1. Andmebaas ja muutujate valik	16
2.2. Valim	17
2.3. Uurimismeetod	19
3. ANALÜÜS JA JÄRELDUSED	21
3.1. Mudeli koostamine ja kitsendamine	21
3.2. Valmismudel	23
3.3. Tõenäosuse marginaalväärtused ja šansside suhe	24
3.4. Järeldused	25
KOKKUVÕTE	28
SUMMARY	30
KASUTATUD ALLIKATE LOETELU	32
LISAD	36
Lisa 1. Tunnused	36
Lisa 2. Peamised statistilised näitajad	38
Lisa 3. Korrelatsioonimaatriks	39
Lisa 4. Esialgne mudel	40
Lisa 5. VIF varieeruvusindeksi analüüs	41
Lisa 6. Kõik mudelid tabelina	42
Lisa 7. Lõplik mudel	43
Lisa 8. Šansside suhted	44
Lisa 9. Tõenäosuse marginaalväärtused	45
Lisa 10. Lihtlitsents	46

LÜHIKOKKUVÕTE

Peer-to-peer (P2P) laenuplatvormid on krediivaldkonda viimase 10 aasta jooksul oluliselt kujundanud, vahendades krediiti otse investorite ja laenuvõtjate vahel ning minnes sellega mööda traditsioonilistest vahendajatest. Ka jõudsal turuosa kasvatamisel seisavad P2P-platvormid silmitsi väljakutsega hallata krediidiriski, et tagada investorite püsiv usaldus. Ühisrahastusettevõtetele on olulisel kohal investorite usalduse teenimine, minimeerides laenuvõtja maksejõuetuse riski, tuginedes täpsele andmete kogumisele. Uurimistö, mille autor viib läbi MONEYZENi andmebaasi alusel, keskendub eraisiku krediidiriski mudeli väljatöötamisele. Selle eesmärk on arvutada laenuvõtja maksejõuetuse tõenäosus, analüüsides, millised laenu ja laenuvõtja tunnused viivad maksehäire tekkimiseni 91. päeval pärast võla registreerimist. Peale hüpoteeside püstitamist ning valimi koostamist analüüsiti laenuvõtjat ja laenu selgitavaid tunnuseid, millest laenuvõtja puhul osutusid statistiliselt oluliseks vanus, MONEYZENi krediidiskoor, sissetulek ning osaliselt ka elukoht ning laenu puhul osutusid oluliseks laenuperioodi pikkus ning intressimäär. Laenajat iseloomustavad tunnused sugu ja tema laste arv ei ole oluliselt seotud maksehäire tekkimisega, samuti pole laenu selgitav tunnus laenusumma statistiliselt oluline.

Võtmesõnad: krediidirisk, ühisrahastus, MONEYZEN, investorid, maksejõuetus

SISSEJUHATUS

Peer-to-peer (P2P) laenuplatvormid on ettevõtted, mille ülesanne on jaotada laenukapitali, määrates turul intressimäär, millega laenuvõtja saab investoritelt raha laenuks ilma traditsioonilist vahendajat, näiteks pankka, kaasamata. P2P laenuplatvormide kontseptsioon arenes välja 2000. aastate alguses osana laiapõhjalistest tehnoloogiliselt muudatustest, mis samaaegselt viisid ka teiste sarnaste veebipõhiste toote- ja teenuste oksjoniplatvormide tekkimiseni (nt eBay, Amazon, Uber, Lyft, Grubhub). (Duarte *et al.*, 2023)

Üks peamisi probleeme, millega ühisrahastusplatvormid silmitsi seisavad, on laenajate krediidiriski maandamine. Selleks, et platvorm oleks investoritele atraktiivne, on platvormi haldajatel vaja viia laenajate maksehäirete tekkimise või üldse laenu tagastamata jätmise risk miinimumini. Riski vähendamisele aitab kaasa eelkõige laenaja ja taotletava laenu kohta piisavate andmete kogumine ning analüüsimine. Krediidiriski analüüsi tegemisel on oluline läbi mõelda, milliseid andmeid analüüsitava kliendi puhul tasub üldse arvesse võtta ehk selgitada välja, millised on need karakteristikud, mis võiksid mõjutada kliendi makseraskustesse sattumist. Samuti on igal karakteristikul osatähtsus, kui palju ta mõjutab potentsiaalse maksehäireni jõudmist. Krediidiriski analüüsi mugavamaks läbiviimiseks on võimalik ühisrahastusplatvormil koostada olemasoleva kliendibaasi andmete põhjal krediidiriski hindav mudel.

Kuna ühisrahastuse populaarsus alternatiivina pankadele näib tänaseni olevat püsinud, otsustas töö autor lähemalt uurida krediidiriski mudeli loomise protsessi ja võimalusi ning omal käel läbi proovida ka üks taolise mudeli loomise meetod. Teema uurimisel otsustas autor aluseks võtta ühisrahastusplatvormi MONEYZEN. Tegemist on ettevõttega, kus autor ise lõputöö kirjutamise hetkel töötab ning igapäevaselt laenajate krediidiriski hindab. Kuna viimase krediidiriski mudeli loomisest on küllalt aega möödunud, oleks ettevõttele igati kasulik uue krediidiriski mudeli loomine, arvestades seekord juurde ka viimase aasta jooksul lisandunud klientide andmed.

Lõputöö eesmärk on luua ettevõttele MONEYZEN eraisiku krediidiriski mudel, mille rakendamisel on võimalik välja arvutada laenuvõtja maksehäire tekkimise tõenäosus. Maksehäire registreerimine toimub 91. päeval peale esimest maksekuupäeva.

Töös püstitatud uurimisülesanded:

1. Anda ülevaade krediidiriski arvutamise võimalustest.
2. Uurida ühisrahastuse ajalugu ning arengut.
3. Anda ülevaade MONEYZENist.
4. Testida eraisiku krediidiriski mõjutavate tunnuste statistilist olulisust MONEYZENi laenuandmebaasi põhjal.
5. Tuvastada analüüsi abil statistiliselt oluliste tunnuste mõju laenuvõtja maksehäire tekkimise tõenäosusele.
6. Luua eraisiku krediidiriski hindav mudel MONEYZENile.
7. Teha järeldused krediidiriski mõjutavate tunnuste osas.

Töös püstitatakse kaks eesmärgiga kooskõlas olevat hüpoteesi:

Hüpotees 1. Laenuvõtja karakteristikud nagu sugu, vanus, elukoht, laste arv, sissetulek ning MONEYZENi krediidiskoor mõjutavad maksehäire tekkimise tõenäosust (on statistiliselt olulised).

Hüpotees 2. Laenu karakteristikud nagu laenu eesmärk, laenusumma, intressimäär ning laenuperiood mõjutavad maksehäire tekkimise tõenäosust (on statistiliselt olulised).

Peale hüpoteeside testimist esitatakse lähtuvalt tulemusest kaks uurimisküsimust:

1. Kuidas mõjutavad laenuvõtjat iseloomustavad statistiliselt olulised tunnused maksehäire tekkimise tõenäosust?
2. Kuidas mõjutavad laenu iseloomustavad statistiliselt olulised tunnused maksehäire tekkimise tõenäosust?

Küsimustele vastamiseks ning hüpoteeside kontrollimiseks andis ühisrahastusplatvorm MONEYZEN töö autorile loa kasutada oma laenude andmebaasi 31.10.2023 seisuga. Andmebaasist jäid valimisse 2458 laenu. Uurimine toimus kirjeldava statistika abil ning andmete analüüsimiseks valiti binaarne logistiline regressioonanalüüs.

Töö koosneb neljast osast. Esimeses osas tutvutakse lähemalt krediidiriski ja selle analüüsimise meetoditega. Samuti on esile toodud põhilised kitsaskohad krediidiriski arvutamisel. Teine osa annab ülevaate ühisrahastusest ning selle arengust, sinna hulka kuulub ka ülevaade ühisrahastusportaalist MONEYZEN. Kolmandas osas kirjeldatakse lahti uurimismeetod, määratletakse valim ning muutujad ning neljandas osas viiakse läbi analüüs, luuakse täiustatud krediidiriski mudel ja tehakse järeldused.

Loodetavasti on lõputööst kasu ettevõttele MONEYZEN edaspidisel klientide krediitvõimekuse hindamisel, vähendamaks maksehäireteni jõudvate laenude hulka ning muutes sellega portaali ka investorite jaoks atraktiivsemaks investeerimisvõimaluseks. Lõputöö autor tänab oma töö juhendajat Kristjan Liivamägi asjakohaste nõuannete ning efektiivse juhendamise eest.

1. ÜHISRAHASTUS JA KREDIIRISK

1.1. Ülevaade ühisrahastusest

Mõiste "peer-to-peer" („inimeselt inimesele“) kirjeldab kahe osapoole vahelist suhtlust, ilma et oleks vaja kesket vahendajat. Mõiste pärineb arvutivõrkude valdkonnast, et kirjeldada võrku, kus üks arvuti võib toimida kas kliendi või serverina teistele võrgu arvutitele, ilma et peaks tsentraliseeritud serveriga ühendust looma. Internet on ise P2P-võrk. (Milne *et al.*, 2016) Bholat *et al.* (2016) seletavad ühisrahastusplatvormide toimimist järgnevalt: platvormid toimivad vahendajatena, kes ühendavad laenuvõtjaid laenuandjatega. Kui laenuvõtja taotleb laenu, taotleb ta kindlat rahasummat. Tavaliselt rahastavad taotletud laenusummat mitu investorit, mistõttu iga laen koosneb erinevatest laenulepingutest üksikute investoritega. Wangi (2016) sõnul on P2P köitmas mitte ainult finantsspetsialistide, vaid ka tavainimeste tähelepanu, kes otsivad alternatiivseid võimalusi laenu saamiseks, ning samuti investorite tähelepanu, kes otsivad madalate intressimäärade kõrvalt kõrgemat tulu.

Havrylchuk *et al.* (2018) näevad ühisrahastuse funktsioonina pakkuda laenu interneti kaudu, kasutades krediidiriski hindamiseks ja laenuvõtjate jälgimiseks erinevat sorti andmeid ja algoritme. Selles rollis aitavad nad kaasa laiemale Fintech-liikumisele, mida iseloomustab tehnoloogia ja tehnoloogiapõhiste ärimudelite kasutamine finantsteenuste osutamisel. (*Ibid.*) Ühisrahastuslaenud on levinud laenuturu erinevatesse segmentidesse. Mõned P2P laenuettevõtted pakuvad konkreetseid laenutüüpe. Esimene näide on tarbimislaenud, mis hõlmavad võlgade konsolideerimist ja refinantseerimist, kodu parandamist, suuremaid oste ja viimasel ajal ka ravilaene. Mõned ühisrahastusettevõtted, nagu Funding Circle, OnDeck ja Raiseworks, keskenduvad eranditult väikeettevõtete laenudele. Teised laenuplatvormid, nagu CommonBond ja SoFi, on spetsialiseerunud õppelaenudele. Kinnisvaralaenuandjad (nt RealtyMogul, LendingHome, Fundrise) erinevad teatud määral ülejäänud P2P laenude pakkujatest. Sellised laenud on tavaliselt tagatud kinnisvaraga ja üksikisikud saavad laenuandjaks registreeruda ainult siis, kui nad kvalifitseeruvad „akrediteeritud“ investoriteks. (Mateescu, 2015)

P2P ajalugu rahanduses saab jälgida kahe ettevõtte, Ühendkuningriigis asuva Zopa (2005) ja USA-s asuva Prosperi rajamisega 2006. aastal. Mõlemad hõlbustasid inimeselt inimesele laenu andmist, mille käigus laenuvõtjad ja laenuandjad said pankadest ja muudest traditsioonilisematest laenuandjatest mööda minna ning ühisrahastusplatvormi kaudu omavahel otse suhelda. Prosperi kaasasutaja Chris Larsen kirjeldas oma ettevõtet kui "eBay krediidi jaoks". (Milne *et al.*, 2016) Bholat *et al.* (2016) uurimusest selgus, et ühisrahastusplatvormi Prosper asutaja David Nicholson, olles veel tööl oma eelmises töökohas internetipangas Egg, mõtiskles panga rolli ja põhieesmärgi üle. Kaaludes panga, eriti jaepanga põhifunktsioone, mis hõlmavad peamiselt hoiuste ja laenude sobitamist ning hoiustajate ja laenuvõtjate vahelise vahendajana tegutsemist, hakkas ta uurima alternatiivseid võimalusi, mis võiksid seda vahendaja rolli täita. See pani tema jaoks kahtluse alla vajaduse, et traditsiooniline pank peab olema ainuke vahendaja investorite ja laenuvõtjate vahel. (*Ibid.*) Erinevalt traditsioonilistest pankadest ei tegele P2P-laenuplatvormid raha loomisega ega riski ja tähtaegade ümberkujundamisega. Selle asemel hõlbustavad mõned ühisrahastusplatvormid järeלטurget laenulepingutega enne tähtaega kauplemiseks ja teatud platvormid püüavad pakkuda laenuandjatele stabiilset sissetulekut. Tõhusa ja jätkusuutliku finantsvahenduse tagamiseks on ülimalt oluline, et P2P-laenuplatvormid tegeleksid võimalike käsundusagendi probleemidega ja viiksid oma stiimulid kooskõlla laenuandjate omadega. (Havrylchuk *et al.*, 2018)

Lisaks P2P finantsteenustele arenesid samaaegselt välja mitmed muud alternatiivsed finantsteenused, mis tegutsevad väljaspool tavalisi pangandus- ja kapitaliturget. Nende hulka kuuluvad: 1) ühisrahastamine ehk *crowdfunding*, mille käigus kogutakse rahasumma konkreetse projekti jaoks (rahastamine - *fundng*) üksikisikute (rahvahulga - *crowd*) väiksemate panuste kaudu; 2) alternatiivsed valuutavahetusplatvormid, kus üksikisikud ja ettevõtted vahetavad välisvaluutat panku kasutamata; 3) pangaväliste arvete diskonteerimine, mille puhul väikesed ettevõtted saavad parandada oma rahavoogu, saades investoritelt ettemaksed tasumisele kuuluvate arvete eest; ja 4) krüptovaluutad, nagu Bitcoin ja Litecoin, mis toetavad digitaalsetes valuutades silmapilkseid veebimakseid ilma keskse vahendajata. Paljudel nendel alternatiivsetel rahastamisviisidel on ka võrdõiguslikud funktsioonid. (Milne *et al.*, 2016)

P2P laenuturul osalejad võib laias laastus jagada kahte rühma: laenuvõtjad ja laenuandjad. Laenuvõtja kohta olemasoleva teabe hulgast püüavad analüütikud leida laenu edukuse määravad. (Guo *et al.*, 2014) Enamikul P2P-laenuplatvormidel võtavad laenuandjad suurema osa riskidest. Finantseeriva poolena võtavad laenuandjad enda kanda laenuvõtjate ja platvormivahendajate

tekitatud maksehäirete riski, kuna need kaks osapoolt võivad otsustada platvormilt lahkuda. (Bao *et al.*, 2023) Maksehäireriski, mida nimetatakse ka krediidiriskiks, peetakse üheks kõige levinumaks riskiks nii traditsioonilisel laenuturul kui ka P2P laenusektoris. Infoasümmeetria ja teabega kokkupuute kitsaskohtade tõttu veebilaenuplatvormidel ei suuda laenuandjad tihti ette näha, kas laenuvõtjad suudavad neile raha tagasi maksta, nagu on kokku lepitud. (Bao *et al.*, 2023) P2P-laenude viivist on üha raskem kontrollida, mistõttu on praktiline tähtsus ettevõtete krediidiriski juhtimise alase uurimistöo tugevdamisel. See aitab vähendada laenuklentide isiklikest krediidiprobleemidest põhjustatud viivist ja pahatahtlikku maksejõuetust ning võib parandada ettevõtete tõhusust ja kaitsta efektiivselt investorite huve. Lisaks on sellel potentsiaal tõsta ettevõtete kindlustunnet. Krediidiskooride rakendamine võib samuti aidata parandada või isegi lahendada enamiku P2P-laenuplatvormide probleemidest. (Li *et al.*, 2021)

1.1.1. Ühisrahastusportaal MONEYZEN

MONEYZEN OÜ on 2013. aastal asutatud ettevõtte, mis pakub tarbijakrediiti Eesti residentidest füüsilistele isikutele ning investeerimisvõimalust nii eraisikutele kui ka ettevõtetele Euroopa Majanduspiirkonnas, kes on huvitatud kapitalitulu teenimisest. (MONEYZEN OÜ äriplaan 2019) Hetkeseisuga on MONEYZENi portaalis ligikaudu 1900 registreerunud investorit ning läbi ajaloo sõlmitud üle 3000 laenulepingu. (MONEYZEN andmebaas 2023) Laenude vahendamisel lähtutakse vastustustundliku laenamise põhimõtetest, tehes laenuvõtjale paremini arusaadavaks tema rahaline olukord ning kohustused, mis lisanduvad täiendava laamisega. Oluline on ära hoida kliendi makseraskustesse sattumine. (*Vastustustundlik laenamine MONEYZEN...* 2023)

MONEYZEN OÜ on ettevõtte Admirals Group AS tütar-ettevõtte, kuid on varasemalt kuulunud veel mitmetele ettevõtetele nagu Business Advisor OÜ (kustutatud), Nordvik Investments OÜ, Snegovik OÜ ning Certus Invest OÜ. MONEYZENile endale kuulub üks tütar-ettevõtte MONEYZEN Collateral Agent OÜ. 2023. aasta 3. kvartali seisuga töötab MONEYZEN OÜ ettevõttes 4 inimest. Eeldatav käive 2023. aastal on 74 665 eurot. (*MONEYZEN OÜ*, 2023) 2022. aasta seisuga oli väljastatud laene 4 369 458 eurot ning müügitulu 62 379 eurot. 2023. aasta plaan nägi ette jätkata kvaliteetse krediidivahendusteenuse pakkumist. Samuti planeeriti aastasse arendustöid, mis aitaksid lõpptulemusena lihtsustada laenuvõtja toiminguid alates taotluse esitamisest portaalis kuni laenulepingu lõppemiseni. Üks olulisematest eesmärkidest oli ka krediidiriski hindamise kvaliteedi tõstmine. (MONEYZEN OÜ majandusaasta aruanne 2022)

1.2. Krediidiriski mõjutavad karakteristikud

Uuringuid, mis konkreetselt püüavad tuvastada maksejõuetuse ja laenude tulemuslikkust määravaid tegureid P2P-laenu puhul, on vähe. P2P-laenuplatvormid koguvad erinevat laenu- ja laenuvõtjateavet, et vähendada teabe asümmeetriat. (Sharar, 2021) Sharar (2021) uuris USA ühisrahausettevõtte Lending Club andmete põhjal, millised on olulisemad tunnused maksejõuetuse tekkimise ennustamiseks ning tulemuste põhjal osutusid olulisemateks laenusumma, sissetulek, krediidi hinnang, võlgade ja sissetulekute suhe ning viimase 6 kuu jooksul peale võetud kohustuste arv. Tema sõnul nii võlgade ja sissetulekute suhte, viimase 6 kuu kohustuste arvu kui ka laenusumma suurenemisel makseraskuste tekkimise tõenäosus samuti suureneb. See-eest suurem sissetulek tõstab laenu edukuse tõenäosust. Ravina (2007) analüüsis Prospani, samuti ühe USA P2P-laenuplatvormi laenustatistikat, võttes arvesse näiteks laenuvõtja krediidi reitingu, laenusumma, võla ja sissetulekute suhte, olemasolevate laenukohustuste arvu ja viiviste koguarvu, mis on veebipõhiste P2P-laenude maksejõuetuse olulised tegurid. Tema poolt tehtud analüüsi tulemusel selgus, et ka "pehmed" andmed, nagu näiteks laenueesmärk või vanus, võivad mõjutada maksejõuetuse riski.

Näitena saab esile tuua ka Serrano-Cinca et al. (2015) poolt läbi viidud uurimuse 2015. aasta USA suurima ühisrahausplatvormi Lending Club poolt kasutatavate laenuvõtjate karakteristikute osas, millest olulisemaks osutusid laenuvõtja hinnang (A-G), laenu intressimäär, laenu eesmärk, sissetulek aastas, kinnisvara olemasolu, krediidi ajaloo pikkus, ametlike teadaannete arv, laenukohustuste arv ning kohustuste protsent sissetulekust. Sealjuures leiti näiteks, et laenusumma ning laenuvõtja tööstaaž ei mänginud olulist rolli maksevõimekuse määramisel. Kõige parema ennustusvõimega näitajaks osutus platvormi Lending Club poolt määratud hinnang laenuvõtjale (A-G). (*Ibid.*) Ka Santoso et al. (2020) viisid läbi regressioonanalüüsi uurimaks Indoneesia kolme ühisrahausettevõtte puhul, millised on peamised maksejõuetuseni viivad laenu ja laenuvõtjaga seotud tunnused. Muuhulgas osutusid kõigi kolme ettevõtte puhul statistiliselt olulisteks laenu intressimäär, laenusumma, laenu periood, laenuvõtja abielustaatus, sissetulek, sugu ning vanus. Santoso et al. (2020) rõhutasid eelkõige seost, mille kohaselt laenu intressimäär mõjutab oluliselt maksehäire tekkimist. Nende sõnul tasub ettevõtetele olla valvas, et mitte intressimäära liialt kõrgeks ajada, sest see on üks olulisemaid faktoreid maksejõuetuse tekkimisel. Järgmine tugevaim ja ühtlasi positiivne seos oli laenusumma ning maksehäire tekkimise vahel. Autor pakub siinkohal lahendusena välja ettevõtetele keskenduda pigem väiksemate laenude väljastamisele, et hajutada riski. (*Ibid.*)

Möllenkamp (2017) regressioonimudeli analüüsist platvormi Lending Club andmete põhjal selgus, et mida kõrgem on laenusumma, seda suurema tõenäosusega tekivad laenuvõtjal makseraskused. Samuti leidis ta, et aastase sissetuleku suurenemisel väheneb maksejõuetuse tõenäosus. Samas laenude protsendi sissetulekust kasvamisel suureneb ka makseraskuste tekkimise tõenäosus. (*Ibid.*) Kuna platvormide seatud kõrged intressimäärad pigem langetavad maksevõimekust, kajastab laenuintress otseselt maksejõuetuse riski: mida kõrgem on intressimäär, seda kõrgem ka maksejõuetuse risk. (Bao *et al.*, 2023) Sama järelduse tegid uuringu tulemuste põhjal ka Serrano-Cinca *et al.* (2015).

Rivera (2018) kasutas oma uurimistöös Mehhiko laenuettevõtete andmeid tuvastamiseks, millised tunnused mõjutavad maksejõuetuse tekkimist. Tema järeldustest selgus, et kõige enam viitavad makseraskuste tekkimisele näitajad laenumakse suhe sissetulekusse ning samal platvormil refinantseerimine. Teisalt laenueesmärk (sõltuvalt otstarbest) ning sugu (kui laenuvõtja on naine) võisid viidata madalamale makseraskuste tekkimise tõenäosusele. Sealjuures mainis Rivera (2018), et sugu kui näitaja, ei omanud eriti suurt tähtsust võrreldes teiste tunnustega maksejõuetuse ennustamisel.

1.3. Krediidiriski arvutamise võimalused

Otsuse tegemine, kas rahastada teatud laene või mitte, on P2P-laenuturul laenuandjate keskne ülesanne. Krediidisüsteemi ebatäiuslikkuse ja veebiruumi piirangute tõttu seisavad ühisrahastusplatvormid laenude väljastamise käigus sageli silmitsi laenuvõtja krediidiriski korrektse hindamisega, sest suur osa laenuvõtjatest kipub sattuma makseraskustesse. (Ma *et al.*, 2021) Traditsioonilised laenude hindamismeetodid on samuti kasutatavad P2P-laenude puhul. Selleks, et eristada häid laenuaotlusi (see tähendab need, mille tagasimaksmist eeldatakse) halvadest (need, mille puhul ennustatakse maksejõuetust), on kirjanduses kasutatud palju erinevaid andmekaeve tehnikaid. Mõned näited nendest tehnikatest on logistiline regressioon, lineaarne diskriminantanalüüs, k-lähimad naabrid, otsustuspuud, Markovi ahelad, lineaarne ja mittelineaarne programmeerimine, närvivõrgustikud jpt. (Guo *et al.*, 2014) Lähemalt tutvutakse nendest viiega.

Logistiline regressioon on meetod, mida laenuturul kõige sagedamini kasutatakse krediidiskoori mudelite väljatöötamiseks. Logistilise regressiooni mudelites on sõltuv muutuja üldiselt binaarne muutuja (nominaalne või järguline) ja sõltumatud muutujad võivad olla kategoorilised (nii kaua, kui need on pärast teisendust dihhotoomilised) või pidevad. Logistilise regressiooni mudel on üldistatud lineaarsete mudelite erijuhtum. (Gouvêa *et al.*, 2007) Esialgu kaasatakse mudeli koostamisel kõik muutujad; lõplikus logistilises mudelis valitakse aga ainult mõned muutujad. Muutujate valik toimub meetodi *forward stepwise* abil, mis on logistilise regressiooni mudelites enim kasutatud. (*Ibid.*)

Lineaarse diskriminantanalüüsi (LDA) eesmärk on klassifitseerida heterogeenne kogum homogeenseteks alamhulkadeks ja edendada nende hulkade otsustusprotsessi. Võib eeldada, et iga hulga jaoks on saadaval teatud arv selgitavaid muutujaid. Idee on otsida sellist selgitavate muutujate lineaarset kombinatsiooni, mis eraldaks enamiku alamhulkadest üksteisest. Kahe alamhulga lihtsal juhul on eesmärk luua selgitavate muutujate lineaarne kombinatsioon, mis jätab kahe alamhulga keskmiste vahele maksimaalse kauguse. (Bolton, 2009)

Närvivõrkude idee pärineb inimese aju struktuurist. Inimese aju võib neuron saata (vastu võtta) signaali teistele neuronitele. Samamoodi koosnevad närvivõrgud elementidest, millest igaüks saab palju sisendeid ja toodab ühe väljundi. Närvivõrgud koosnevad suurest hulgast lihtsatest sõlmedest või neuronielementidest, mis on ühendatud kas ühe või mitme kihiga. Iga neuroni aktiivsuse oluline element on funktsioon, mis vastutab väljundsignaali väärtuse eest, mida nimetatakse aktiveerimisfunktsiooniks. Sellel võib olla üks kolmest vormist: binaarne samm, lineaarne ja mittelineaarne (nt sigmoidne). (Dzik-Walczak *et al.*, 2021)

Otsustuspuu on klassifitseerimistehnika, mis määrab andmestiku iga objekti prognoositavasse klassi (nt hea/halb maksejõuetuse risk), mis sõltub iga objekti atribuutidest. Algoritm kasutab teabe võimendust, et leida parim atribuut andmete klassifitseerimiseks, kus P ja n on i -nda objekti binaarse tulemuse väärtused 0 ja 1. Seejärel kordab algoritm iga parima atribuudi otsustusväärtuste jaoks määratud väärtuse puhul protsessi täiendavate atribuutidega. (Rudd *et al.*, 2017)

K lähimate naabrite (KNN) algoritm on teatud tüüpi juhendatud masinõppe algoritm, mida saab kasutada klassifitseerimise ja regressiooni ennustamise probleemide lahendamiseks. KNN on „laisk“ õppimisalgoritm, kuna sellel pole spetsiaalset koolitusfaasi ja see kasutab klassifitseerimisel kõiki andmeid treenimiseks. See on ka mitteparameetiline õppealgoritm, kuna

see ei eelda alusandmete kohta midagi. KNN-i algoritm kasutab funktsioonide sarnasust, et ennustada uusimate andmepunktide väärtusi, millele saab omistada väärtuse sõltuvalt sellest, kui täpselt see ühtib treeningkomplekti punktidega. See on lihtne algoritm, mis on mõistetav ja tõlgendatav ning mittelineaarsete andmete jaoks väga kasulik, kuna selles algoritmis pole andmete kohta eeldusi. (Dewani *et al.*, 2020)

1.4. Kitsaskohad krediidiriski hindamisel

Krediidiriski efektiivne hindamine sõltub mitte ainult täpse mudeli loomisest, vaid ka muudest aspektidest, millest olulisemad on selles peatükis esile toodud. Suryono *et al.* (2019) töid oma uuringu põhjal peamiste murekohtadena välja krediidiriski hindamisel informatsiooni asümmeetria probleemi, laenuvõtja krediidiskoori määramise, moraalariski, regulatsioonide ja ettevõtte eeskirjade klapitamise, samuti investeerimisotsuste tegemise.

Tupangiu (2017) sõnul juhul kui krediidiagentide teabe tase on erinev, tekib teabe asümmeetria probleem. Krediidisuhte kontekstis seisab finantsasutus silmitsi teabe asümmeetria väljakutsega, mis tuleneb eelkõige krediidiriski hindamise keerukusest. Tõhus krediidiriski juhtimine hõlmab finantsasutuse võimet koguda ja töödelda teavet laenuaotluse vastuvõtmise ajal, vähendades seeläbi krediidiriski. Finantsinstitutsioonil on oluline omada asjakohast teavet laenuvõtja kohta nii laenu taotlemise faasis kui ka pärast krediidi kinnitamist. Asjakohast teavet võimendades saab finantsasutus teostada kontrolli laenuvõtja tegevuse üle. Järelikult puututakse teabe otsimisel kokku teabe asümmeetria loomupärase väljakutsega (*Ibid.*).

Samuti mõjutavad krediidiriski kujunemist makroökonomilised faktorid ning majandustsüklid. Peale 2007. – 2008. aasta finantskriisi viisid Ali ja Daly (2010) läbi uurimuse selle kohta, milline oli makroökonomiliste tegurite mõju krediidiriskile ning leidsid, esile kerkis arvukalt makromajanduslikke ja poliitilisi takistusi. Väljakutseks olid krediidi- ja finantstsüklite kaasamine tavapärase makromajanduslikesse mudelitesse, makromajanduslikest seostest arusaamise parandamine ja makrotasandi usaldatavusnormatiivide reguleerimise vahendite väljatöötamine.

Kindlasti mõjutab riski hindamist inimlik aspekt. Näiteks teevad finantsettevõtetes vastutaval kohal töötavad inimesed oma loomujoontest mõjutatuna otsuse, milliseid krediidiriski hindamise vahendeid kasutada ja see võib kokkuvõttes ettevõtte käekäiku oluliselt mõjutada. Eriti

enesekindlad finantsettevõtete juhid otsustavad krediidiriski arvutamisel kvantitatiivsete meetodite kasuks, see-eest optimistlikud juhid usaldavad enda otsustusvõimet ja isiklikku võimekust krediidiriski hindamisel. (Azouzi, Bacha, 2023)

Grundke (2003) väidab, et paljude krediidimudelite puhul on levinud puuduseks stohhastilise modelleerimise puudumine riskitegurite puhul, nagu intressimäärad või krediidimarginaalid, ümberhindlusprotsessi ajal riskiperioodil. Näiteks fikseeritud tulumääraga instrumendid, nagu võlakirjad või laenud, hinnatakse riskiperioodil ümber, kasutades tulevaste rahavoogude diskonteerimiseks praeguseid forvardintresse ja (hinnanguklassides spetsiifiliselt) tähtpäevakrediidivahesid. Järelikult ei võeta arvesse instrumendi tulevase väärtuse stohhastilist olemust, mis tuleneb krediidikvaliteedi muutustest väljapoole jäävatest teguritest, mis toob kaasa krediidiportfelli riskantsuse alahindamise riskihorisondis (*Ibid.*)

Üks laialdaselt kasutatud laenuotsuste tegemise meetod hõlmab lävendikriteeriumite kehtestamist, võttes arvesse hajutatud teabe erinevaid aspekte. Kui laenuvõtja täidab need tingimused või on edukalt läbinud isiksuse hindamise, kiidetakse laen heaks. P2P internetilaenu puhul ei pruugi üksikisikud sageli teatud künniseid täita, kuid võivad olla mingites muudes tingimustes suurepärased, mis viib nende laenusoovi tagasilükkamiseni. Tegelikult puuduvad paljudel laenuplatvormidel piisavalt täpselt määratletud reeglid ning efektiivsed lävendikriteeriumid, mis aitaksid krediidiriski paremini hallata. (Li *et al.*, 2021)

Mõnes riigis võivad krediidiriski hindamise keeruliseks muuta ka kultuurieripärad ning käitumisharjumused. Näiteks kirjeldasid Zheng *et al.* (2018) oma uurimistöös lähemalt Bangladeshis krediidituru olukorda ning nende hinnangul on oluline krediidiriski mõjufaktor eetilise käitumise ebapopulaarsus ning üldine moraaltunde vähesus osana sotsiaalsest kultuurist riigis. Samuti võib krediidiriski olla keeruline hallata laenajate puhul, kes on pärit välisriigist. Näitena toob Vodová (2003) esile situatsiooni, kus teisest riigist pärit laenuvõtjale antakse krediiti, kuid riigis toimunud sotsiaalsete, majanduslike või poliitiliste muutuste tõttu ei saa laenuvõtja enam laenu tagasi maksta. Sel juhul pakub Vodová (2003) välja lahenduse sätestada erinevate riikide laenajatele vastavalt selle riigi riskitasemele erinevad limiidid, kui suurt summat maksimaalselt tohib selle riigi elanikule laenata. Ühtlasi mainib ta kitsaskohana ära ka teadmatuse selles osas, mis toopi maksehäirega on tegemist, kui klient maksehäireni jõuab: kas põhjus on selline, mis võimaldab kliendil hiljem laenu siiski tagastada või võib laenu lugeda lootusetuks. (*Ibid.*)

2. METOODIKA

2.1. Andmebaas ja muutujate valik

Lõputöö eesmärk on luua ühisrahastusportaalile MONEYZEN krediidiriski mudel, mistõttu kasutatakse andmete analüüsiks ning valemi loomiseks MONEYZENi laenude andmebaasi seisuga 31.10.23. Andmebaasis on informatsioon 2933 laenu kohta alates aastast 2014. Kuna autor uurib lõputöö raames milliste karakteristikutega klientidel on tekkinud maksehäired, siis võtab ta vaatluse alla nende laenuvõtjate lepingud, kelle laenu välja maksmisest on lõputöö kirjutamise ajal möödunud vähemalt 90 päeva ehk on olnud võimalik maksehäire tekkimine. Kuna andmebaasis on eraldi kategoriseeritud ka kohtulahendini jõudnud lepingud, siis loetakse ka need tekkinud maksehäirete kategooriasse, sest tegemist on sellegipoolest 90 päeva maksmata võlaga.

Otsustavaks teguriks valemis kasutatavate muutujate valikul osutus MONEYZENi poolt varasemalt klientide kohta kogutud info maht. Kuigi töö teoreetilises osas said ära mainitud mitmed tavapäraselt olulisemad karakteristikud krediidiriski hindamisel, pidi lõputöö autor arvestama ka sellega, milliste karakteristikute kohta on MONEYZEN pidanud vajalikuks laenuvõtjate ja laenude puhul andmeid koguda ning säilitada. Leidsid tunnuseid (näiteks kinnisvara arv ning perekonna staatus), mille puhul oli andmeid kogutud vaid teatud lühemal perioodil, mistõttu oleks mainitud karakteristikute lisamine analüüsi oluliselt vähendanud valimi mahtu ning mudeli statistilist täpsust. Töö autor valis analüüsimiseks ning püstitas hüpoteesid nende muutujate baasil (Lisa 1), mille kohta oli informatsioon olemas enamiku MONEYZENi andmebaasis olevate laenude puhul:

Hüpotees 1. Laenuvõtja karakteristikud nagu sugu, vanus, elukoht, laste arv ning sissetulek mõjutavad maksehäire tekkimise tõenäosust (on statistiliselt olulised).

Hüpotees 2. Laenu karakteristikud nagu laenu eesmärk, laenusumma, intressimäär, laenuperiood ning MONEYZENi krediidiskoor mõjutavad maksehäire tekkimise tõenäosust (on statistiliselt olulised).

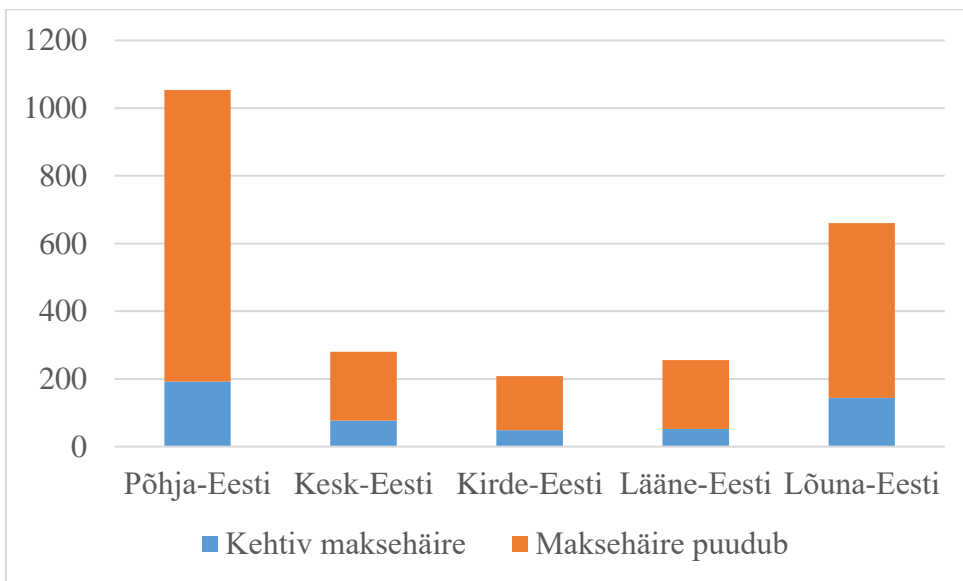
Sõltuvaks muutujaks valiti ettevõtte MONEYZEN soovil maksehäire tekkimine, mis tähendab, et laenajal on tekkinud MONEYZENi ees maksevõlg, mis on tasumata vähemalt 90 päeva ning 91. päeval registreerib ettevõtte maksehäire. Kvalitatiivsetele ja järjestustunnustele (laenu eesmärk, elukoht) lisas autor vastavad fiktiivsed tunnused. Siinkohal soovib autor täpsustada, et jaotas tunnuse elukoht riigivõimu poolt määratud haldusjaotuse alusel 5 erinevaks kategooriaks (Eesti Entsüklopeedia, 2013):

- 1) Põhja-Eesti: Harjumaa;
- 2) Lõuna-Eesti: Jõgevamaa, Põlvamaa, Tartumaa, Valgamaa, Viljandimaa, Võrumaa;
- 3) Kesk-Eesti: Järvamaa, Lääne-Virumaa, Raplamaa;
- 4) Lääne-Eesti: Hiiumaa, Läänemaa, Pärnumaa, Saaremaa;
- 5) Kirde-Eesti: Ida-Virumaa.

Suuremateks kategooriateks jaotamine tundus mõistlik eelkõige seetõttu, et osades maakondades oli laenuvõtjate esindatus väga madal, näiteks olid 1055 inimest elukohaks märkinud Harjumaa, kuid Hiiumaal elavaid laenuvõtjaid oli vaid 11.

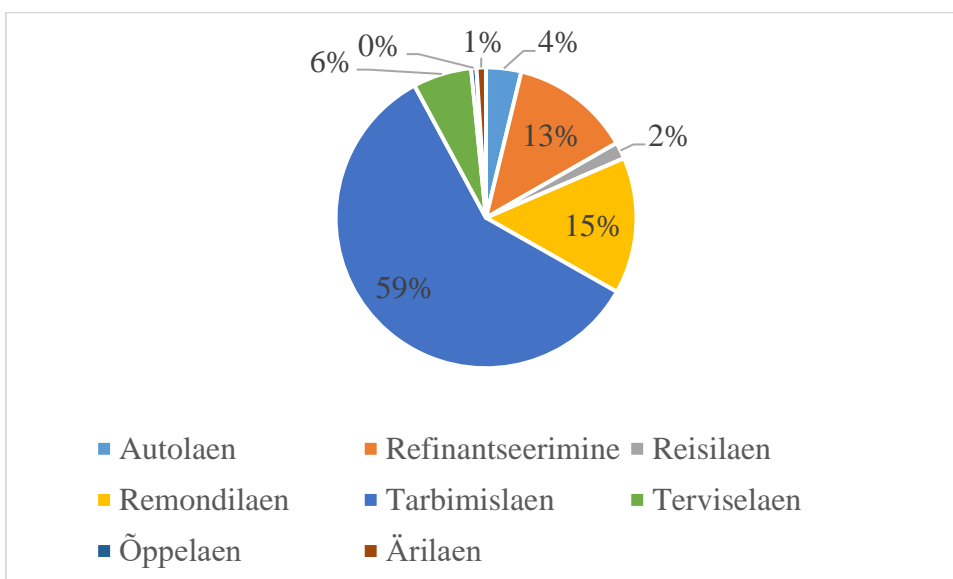
2.2. Valim

Valim koosnes 2458 lepingust, mille baasil autor selgitas välja edaspidi kirjeldatud statistilised näitajad. Lähtudes käesoleva töö eesmärgist hindas autor, et laenuvõtjatest 79,1% puudus hindamise hetkel maksehäire ning 20,1% esines maksehäire. 54,3% laenajatest olid mehed ning 45,7% naised. Keskmiselt oli igal laenajal 1 laps ning leibkonna igakuine keskmine sissetulek oli valimi baasil 1213 eurot. Keskmine laenuvõtja on 38-aastane, kuid kõige suurem hulk laenuvõtjaid leidis 33-aastaste hulgas. Laenuvõtjate keskmiseks MONEYZENi krediidiskooriks osutus 824, samas kõige sagedamini esinev krediidiskoor oli 850. Kõige rohkem laenuvõtjaid elab Põhja-Eestis (43%) ning kõige väiksem oli esindatus Kirde-Eestis (9%). Kõige madalam maksehäiresse sattunud klientide protsent oli Põhja-Eestis (18,2%) ning kõige suurem protsent maksehäiretega kliente oli Kesk-Eestis (27,5%) (Joonis 1). Arvutades protsentuaalselt kumma soo puhul esines rohkem maksehäired, selgus, et 22,1% meestest oli tekkinud maksehäire, võrreldes 19,5% naistest.



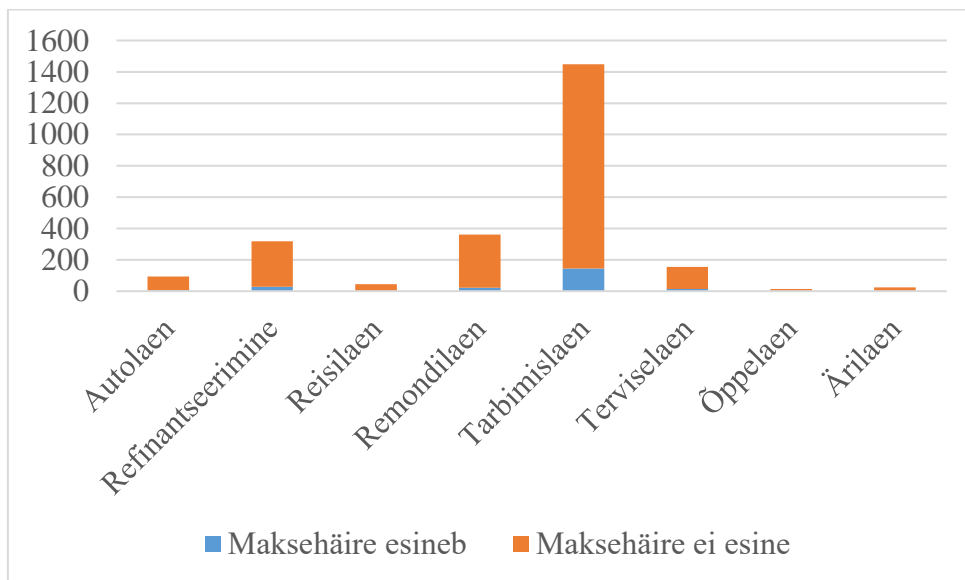
Joonis 1. Tekkinud maksehäirete osakaal elukohtade kaupa
Allikas: autori poolt koostatud

Keskmine laenatav summa oli 1910 eurot (Lisa 2) intressimääraga 23,2%. Kõige madalam laenusumma oli 100 eurot ning madalaim intressimäär 14,16% ja kõige kõrgem laenusumma on 10 000 eurot ning kõrgeim intressimäär 36,3%. Valimi hulgas leidub enim tarbimise (igapäevaste kulutuste) eesmärgil võetud laene, mille osakaal kõikide laenude hulgas oli koguni 59%. Kõige vähem kasutatakse laenu hariduslikul (0,57%), äritegevuse (1,02%) ning reisimise (1,79%) eesmärgil (Joonis 2). Keskmine laenuperiood oli 33 kuud, samas kõige enam laene võeti 24 kuuks.



Joonis 2. Laenu eesmärkide protsentuaalne jaotus
Allikas: autori poolt koostatud

Laenuesmärkide kaupa analüüsid selgus, et kõige suurem maksehäire esinemise protsent (9,9%) oli tarbimislaenu valinud laenuvõtjate hulgas (Joonis 3). Ligi sama palju protsentuaalselt oli maksehäireni jõudnud laenuvõtjaid nende hulgas, kes valisid tervise-laenu (9,7%). Kõige vähem maksehäireni jõudnud laenajaid leidis äri-laenu võtnute hulgas (4%).



Joonis 3. Tekkinud maksehäirete osakaal laenuesmärkide kaupa
Allikas: autori poolt koostatud

2.3. Uurimismeetod

Mudeli koostamisel kasutatakse binaarset logistilist regressiooni, mis võimaldab maksehäire tekkimist hinnata kaheväärtuselise tunnusega. Selleks luuakse mudel, kus sõltuv muutuja väljendab, kas klient jõuab maksehäire registreerimiseni (väärtus 1) või ei jõua (väärtus 0) ning ülejäänud muutujad on regressorid. Siinkohal tutvutakse lähemalt binaarse logistilise regressiooni teoreetiliste alustega Sauga (2020), Grimes *et al.* (2008) Gujarati (2009, 554, 558-560) ning Kleinbaum *et al.* (2010) allikate baasil.

Selleks, et uurida sõltuva binaarse tunnuse (Y ehk MAKSEHÄIRE) väärtuse 1 esinemise tõenäosust, kasutatakse järgnevalt toodud valemit.

$$P(Y = 1|X) = \frac{1}{1+e^{-\lambda}}$$

Valemist nähtuva logiti Λ väärtus omakorda väljendub mudelina:

$$\Lambda = \ln\left(\frac{p_i}{1-p_i}\right) = i_0 + i_1x_1 + i_2x_2 + i_3x_3 + \dots + i_mx_m$$

Mudeli esimeses pooles esindab p_i väärtus maksehäire tekkimist ning vastavalt $1 - p_i$ maksehäire puudumist. Teises pooles kuulub iga regressori (x_m) juurde vastav parameeter (i_m). Kaotades mudelist logaritmi, annab see võimaluse arvutada välja šansi, mida on lihtsam tõlgendada kui tõenäosust. Selleks võetakse mõlemad mudeli pooled astmesse e .

Järgnevalt kirjeldatakse šansside teooriat. Erinevalt tõenäosusest ei pruugi šansid olla koheselt intuiitiivsed. Šansid võimaldavad teistsugusel viisil väljendada tõenäosust, täpsemalt kujul sündmuse tõenäosus jagatud sündmuse mittetoimumise tõenäosusega:

$$Odds = \frac{p_i}{1-p_i} = e^{i_0 + i_1x_1 + i_2x_2 + i_3x_3 + \dots + i_mx_m}$$

Šansside suhe, esindab ühe arvude grupi tulemise tõenäosust jagatuna teise arvude grupi tulemise tõenäosusega. See on suhe, mis ulatub nullist lõpmatuseni. Šansside suhtarvude valem on mõnevõrra keerulisem, kuid seda saab lihtsustada lühendite kasutamisega. Antud kontekstis tähistab p_K tulemise tõenäosust rühmas K, samas kui p_L tähistab tulemise tõenäosust rühmas L.

$$Odds\ Ratio = \frac{\frac{p_K}{1-p_K}}{\frac{p_L}{1-p_L}}$$

Eelnevate valemite abil saame teada, kui palju suureneb või väheneb protsentuaalselt maksehäire tekkimise šans valitud tunnuse suurendamisel 1 ühiku võrra. Juhul kui šans on võrdne ühega, ei mõjuta tunnus maksehäire tekkimist.

3. ANALÜÜS JA JÄRELDUSED

3.1. Mudeli koostamine ja kitsendamine

Enne logistilise regressioonimudeli koostamist, oli vaja läbi viia korrelatsioonanalüüs, testimaks, kas muutujate vahel esineb multikollineaarsust. Paasi (1995, 203) sõnul võib multikollineaarsuse esinemine tekitada olukorra, kus regressioonimudel osutub ebaloogiliseks ning selle kasutamine majandussituatsioonide analüüsimisel riskantseks. Korrelatsiooni, mis on üle 0,9 loetakse juba väga kõrgeks ning seda tuleks vältida. (Calkins 2005)

Korrelatsioonimaatriksisse (Lisa 3) lisati kõik pidevad tunnused, peale sõltuva muutuja (MAKSEHÄIRE). Enamus tunnuste vahel kõrget korrelatsiooni ei esinenud. Kõige kõrgem korrelatsioon, mis tabelist ilmnes, oli LAENUSUMMA ning LAENUPERIOODI vahel (0,7843). Kuigi korrelatsiooninäitaja ei ületa eelnevalt mainitud piirmäära 0,9, on ta siiski piisavalt kõrge ning autor peab seda edasisel analüüsil silmas. Kuna aga antud juhul multikollineaarsuse probleemi otseselt ei tekkinud, oli võimalik liikuda edasi regressioonimudeli koostamise ning hiljem ka fiktiivsete tunnuste multikollineaarsuse testimise juurde (VIF indeksi analüüsi abil).

Esialgmesse mudelisse (Lisa 4) lisati sõltuva tunnuseks MAKSEHÄIRE, kõik pidevad tunnused (VANUS, SUGU, LASTEARV, KREDIIDISKOOR, LAENUPERIOOD, INTRESSIMÄÄR, LAENUSUMMA, SISSETULEK) ja fiktiivsed tunnused, välja arvatud DELUKOHT_1 ning DLAENUEESMÄRK_1, mis on baaskategooriad ning seega oleksid mudelis üleliigsed. Analüüsist selgus, et mudeli kirjeldusvõime on 79,5%, mis tähendab, et mudel kirjeldab õigesti 1957 juhtumit. Mudelis kirjeldati juhtumitest, kus maksehäiret ei tekkinud (väärtsus 0), õigesti 1939 juhtumit ning juhtumitest, kus maksehäire tekkis, kirjeldati õigesti 14 juhtumit. Statistiliselt ebaoluliseks nivool 0,05 osutusid päris mitu tunnust, sealhulgas SUGU ($p = 0,0757$), LASTEARV ($p = 0,2876$), LAENUSUMMA ($p = 0,6884$), fiktiivsed tunnused DELUKOHT_2 ($p = 0,1234$), DELUKOHT_4 ($p = 0,5673$), DELUKOHT_5 ($p = 0,1591$) ning kõik tunnuse LAENUEESMÄRK fiktiivsed tunnused. Ülejäänud mudelis olevad tunnused osutusid oluliseks nivool 0,01. Tõepärasuhte testi järgi osutus mudel oluliseks ($p = 0,0000 < 0,05$). Seega oli vaja

samm-sammulise tagurpidi kõrvaldamise meetodi abil hakata eemaldama ebaolulisi tunnuseid mudelist.

Esmalt võeti ette kõige ebaolulisem muutuja DLAENUEESMÄRK_7 ($p = 0,9235$). Kuna fiktiivsete tunnuste puhul ei ole võimalik neid ükshaaval eemaldada, siis oli vaja kõrvaldada terve LAENUEESMÄRK tunnusega seotud fiktiivsete tunnuste grupp. Gretlis saab F-testi abil uurida, kuidas mõjutab tunnuse eemaldamine mudeli kvaliteeti. Sellega seoses püstitatakse hüpoteesid, kus 0-hüpotees väidab, et kitsenduse võib peale panna, sest kitsendatud mudel ei ole oluliselt halvem originaalmudelist ning 1-hüpotees väidab, et kitsendusi peale panna ei tohi, sest mudeli kvaliteet halveneb oluliselt. F-testi sooritamisel, eemaldades tunnuse LAENUEESMÄRK fiktiivsed tunnused, võis vastu võtta 0-hüpoteesi, sest testi p-väärtus ($0,314647 > 0,05$) ei osutunud oluliseks nivool 0,05. Küll langesid veidi mudeli kirjeldusvõime (79,3%) ning muud näitajad McFadden R-Squared (0,053272), Akaike kriteerium (2412,534), Hannan-Quinn (2439,965). See-eest tõusis kohandatud R-squared näitaja (0,042958). Seega oli võimalik edukalt eemaldada LAENUEESMÄRK fiktiivsete tunnuste grupp mudelist.

Järgmiseks statistiliselt kõige ebaolulisemaks tunnuseks osutus LAENUSUMMA ($p = 0,9435$). Ka sel juhul sooritati F-test kitsenduse tegemiseks ning tulemusena võeti taaskord vastu 0-hüpotees, sest F-testi p-väärtus ($0,943522 > 0,05$) ei osutunud oluliseks nivool 0,05. Analüüsidest suhtarve McFadden R-squared (0,051575), Akaike kriteerium (2410,811), Hannan-Quinn (2431,912), selgus, et need olid taas veidi langenud, samas nii kohandatud R^2 (0,043641) kui ka mudeli kirjeldusvõime (79,5%) kasvasid järjekordselt.

Edasi uuriti tunnuse ELUKOHT fiktiivsete tunnuste gruppi, mille tarbeks võeti kasutusele VIF varieeruvusindeksi analüüs (Lisa 5). Eeldus on, et multikollineaarsusega probleemi ei esine, kui tunnuse indeksi väärtus jääb alla 10. Antud juhul jäid kõigi tunnuste indeksid 1-2 vahele, ehk probleeme ei esinenud. Seega prooviti edasi eemaldada läbi F-testi mudelist tunnuse ELUKOHT fiktiivsed tunnused, kuid nüüd osutus p-väärtus ($0,0234365 < 0,05$) oluliseks nivool 0,05, mistõttu tuli vastu võtta 1-hüpotees ning mainitud fiktiivseid tunnuseid mudelist eemaldada ei tohtinud.

Järgnevalt eemaldati mudelist tunnus LASTEARV ($p = 0,2359$), mille F-testi tulemusel võeti vastu 0-hüpotees, sest testi p-väärtus (0,236049) oli suurem kui 0,05. Sel korral suhtarvud McFadden R-squared (0,052719) ning Hannan-Quinn (2433,137) olid hoopis kasvanud, samas Akaike kriteerium (2409,926) endiselt vähenenud. Mudeli kirjeldusvõime vähenes samuti 79,4% peale.

Viimane statistiliselt ebaoluline tunnus, mis mudelist eemaldati, oli SUGU ($p = 0,0901$). Sel korral osutus F-testi tulemus ($p = 0,0902245$) taas ebaoluliseks nivool 0,05 ning võeti uuesti vastu 0-hüpotees. McFadden R-Squared (0,051575) ning Hannan-Quinn (2410,811) näitajad langesid taas, see-eest Akaike kriteerium (2431,912) kasvas. Samuti kasvas mudeli kirjeldusvõime uuesti 79,5% peale.

Autor otsustas lisakontrolli jaoks veel statistiliselt ebaolulised tunnused eraldiseisvalt mudelisse lisada, et näha, kas nende statistiline olulisus muutub märgatavalt. Tunnuse SUGU puhul puudus ka eraldiseisvana statistiline olulisus nivool 0,05 ($p = 0,1192 > 0,05$). Tunnuse LAENUEESMÄRK fiktiivsete tunnuste puhul osutus statistiliselt oluliseks nivool 0,05 vaid DLAENUEESMÄRK_6 ($p = 0,0302$), ülejäänud fiktiivsed tunnused olid endiselt ebaolulised mudeli kontekstis. LASTEARV oli taaskord eraldiseisvana statistiliselt ebaoluline ($p = 0,6563 > 0,05$), kuid LAENUSUMMA osutus eraldi mudelis statistiliselt oluliseks nivool 0,05 ($p = 0,0274 < 0,05$). Autor pakub, et LAENUSUMMA olulisus eraldiseisvalt, kuid ebaolulisus koondmudelis võib olla seotud eelnevalt leitud teistest tunnustest veidi kõrgema korrelatsiooniga LAENUSUMMA ning LAENUPERIOODI tunnuste vahel (korrelatsiooninäitaja 0,7843).

3.2. Valmismudel

Lõplikust mudelist (Lisa 7) jäid välja tunnused LAENUSUMMA, LASTEARV, SUGU ning tunnuse LAENUEESMÄRK fiktiivsed tunnused. Kõik tunnused, mis mudelisse jäid, välja arvatud DELUKOHT_2 ($p = 0,0681$), DELUKOHT_4 ($p = 0,4576$) ja DELUKOHT_5 ($p = 0,1146$) on statistiliselt olulised nivool 0,05. Lõpliku mudeli kirjeldusvõime on 79,5% ning see kirjeldab korrekselt ära 1957 juhtumit. Sealhulgas juhtumitest, kus maksehäiret ei tekkinud (väärtus 0), kirjeldatakse õigesti 1945 juhtumit ning juhtumitest, kus maksehäire tekkis, kirjeldati õigesti 12 juhtumit. Mudeli Hannan-Quinn näitaja on 2432,890, Akaike kriteerium 2411,786 ning Schwarzzi kriteerium 2469,873. McFaddeni R Squared näitaja on 0,051894 ning kohandatud R Squared 0,043966. Lõplik mudel on järgnev:

$$\text{MAKSEHÄIRE} = -5,736 - 0,034 \cdot \text{VANUS} + 0,004 \cdot \text{KREDIIDISKOOR} +$$
$$0,020 \cdot \text{LAENUPERIOOD} + 7,450 \cdot \text{INTRESSIMÄÄR} - 0,0002 \cdot \text{SISSETULEK} +$$

$$0,232*DELUKOHT_2 + 0,549*DELUKOHT_3 + 0,133*DELUKOHT_4 +$$

$$0,296*DELUKOHT_5 + u$$

$$(0,187)$$

$$n = 2458$$

$$R^2 = 0,051894$$

u märgib juhuslikku komponenti.

Mudeli põhjal selgub, et vanuse kasvades laenuvõtja maksehäire tekkimise tõenäosus väheneb. Töö autori jaoks veidi üllatav tulemus oli see, et mida kõrgem on laenaja krediidiskoor, seda suurema tõenäosusega tekivad tal makseraskused. See viitab vajadusele kujundada ümber või parandada ettevõtte krediidiskoori hindamise mudelit. Laenuperioodi pikenedes suureneb laenuvõtja maksehäire tekkimise tõenäosus. Mida kõrgem on laenu intressimäär, seda kõrgem on ka tõenäosus, et laenuvõtja sattub makseraskustesse. Laenaja sissetuleku kasvades väheneb tõenäosus sattuda makseraskustesse. Lõuna-, Lääne-, Kirde- ja Kesk-Eestis elavad laenuvõtjad sattuvad suurema tõenäosusega makseraskustesse kui Põhja-Eestis elavad laenuvõtjad.

3.3. Tõenäosuse marginaalväärtused ja šansside suhe

Järgmine samm töös oli tõenäosuse marginaalväärtuste analüüs (Lisa 9), mille läbiviimiseks lisati Gretl programmis mudelisse tulba *p-value* asemele tulp *slope*, mille väärtused näitavad, kuidas iga regressori suurenemine ühe ühiku võrra mõjutab sündmuse (maksehäire tekkimine) esinemise tõenäosust. (Sauga 2023) Marginaalväärtuste analüüsi kohaselt keskmiste näitajatega laenuvõtja vanuse suurenedes aasta võrra väheneb maksehäire tekkimise tõenäosus 0,0052 võrra. Kui keskmiste näitajatega laenaja krediidiskoor suureneb ühe punkti võrra, siis suureneb ka maksehäire tekkimise võimalus 0,0006 võrra. Laenuperioodi pikenedes ühe kuu võrra, sattub laenaja makseraskustesse 0,0031 võrra suurema tõenäosusega. Kui intressimäär tõuseb ühe protsendi võrra, siis suureneb ka maksehäire tekkimise tõenäosus 1,1594 võrra. Tõstes keskmiste parameetritega laenuvõtja sissetulekut 1 euro võrra, väheneb tema makseraskustesse sattumise tõenäosus $-3,1577^{-5}$ võrra.

Edasi uuris autor šansside suhete analüüsi (Lisa 8) tulemusi, kust selgus, et laenaja krediiskoori kasvades ühe punkti võrra suureneb tema maksehäire tekkimise šanss 0,4%. Kui laenuvõtja saab aasta vanemaks, siis väheneb tema makseraskustesse sattumise šanss 3,29%. Juhul kui laenuperioodi pikendatakse ühe kuu võrra, tõuseb laenaja maksehäire tekkimise šanss 2,04%. Laenaja sissetuleku suurenemisel 1 euro võrra, väheneb tema makseraskustesse sattumise tõenäosus 0,02%. Kui laenuvõtja elab Lõuna-Eestis, on tema maksehäire tekkimise šanss 1,2 korda suurem kui Põhja-Eestis elaval laenuvõtjal. Nii samuti on võrreldes Põhja-Eesti laenajaga Kesk-Eesti laenaja maksehäire tekkimise šanss 1,7 korda, Lääne-Eesti laenaja šanss 1,1 korda ning Kirde-Eesti laenaja šanss 1,3 korda suurem.

3.4. Järeldused

Mudeli koostamisel ning analüüsimisel selgus, et kaks laenuvõtjat kirjeldavat muutujat: sugu ning laste arv ja samuti kaks laenu kirjeldavat muutujat: laenueesmärk ning laenusumma osutusid krediidiriski määramisel statistiliselt ebaolulisteks. Seega saab mõlemad töös püstitud hüpoteesid osaliselt vastu võtta. Laste arvu lisamine mudelisse oli autori idee, mis ei olnud otseselt läbi töötatud kirjanduses esile toodud kui oluline krediidiriski mõjutav faktor, kuid autor soovis siiski uurida tunnuse mõju makseraskuste tekkimisele. Analüüsist selgus, et laste arv polnud ka MONEYZENi laenajate maksehäire tekkimise määramisel piisavalt oluline, et seda mudelisse lisada. Samas laenuga seotud käesolevas töös ebaoluliseks osutunud muutujaid laenusummat ning -eesmärki mainiti Serrano-Cinca *et al.* (2015) poolt kui olulisi mõjufaktoreid krediidiriski hindamisel Lending Club ühisrahastusplatvormil.

Oluliseks osutusid laenuvõtjat kirjeldavatest karakteristikutest vanus, ettevõtte poolt määratud krediiskoor, sissetulek ning elukoht. Analüüsist selgus, et vanemad inimesed kipuvad vähem sattuma makseraskustesse võrreldes noorematega. Selle põhjus võib seisneda finantsilise olukorra stabiliseerumisega ning kogemuse tekkimisega valdkonnas aastate jooksul. Samuti võib vanuse kasvades väheneda riskiisu ning impulsiivsus rahaliste otsuste tegemisel ning pigem on seoses kogemusega kindlustatud ka mingit sorti stabiilsem sissetulek. Eriti pensionäride puhul on kindel igakuine sissetulek üldjuhul garanteeritud. Vanus osutus laenuvõtja maksehäire tekkimisel oluliseks faktoriks ka Ravina (2007) analüüsis.

Autorit üllatas regressioonanalüüsi käigus selgunud krediidiskoori ning maksehäire tekkimise vaheline positiivne seos. Lähemal uurimisel ning MONEYZENi äriprotsesside juhiga vesteldes selgus, et ettevõttes on laenuvõtja krediidiskoori arvutust erinevate juhtide poolt vastavalt nende visioonile ajas mõnevõrra muudetud. Seega võib järeldada, et mingil perioodil ettevõtte ajaloos on krediidiskoori arvutamise mudel osutunud ebaefektiivseks, kuid just läbi järjepideva analüüsi ning katse-eksituse meetodil saabki täpsema mudeli krediidiskoori arvutamiseks välja töötada ning seda ajas parandada. See on ühtlasi ka põhjus, miks autor otsustas käesoleva uurimise läbi viia.

Laenaja sissetuleku negatiivne suhe maksehäire tekkimise osas tundus loogiline, sest suurema sissetuleku saajal jagub tõenäoliselt rohkem rahalisi vahendeid ka laenude tasumiseks. Võib eeldada, et kõrgema sissetuleku teenijal on olnud rohkem distsipliini, et sellise sissetuleku teenimiseni jõuda ning võimalus panna suurem hulk raha kõrvale finantsiliste kriisidega toimetulekuks. Makseraskuste esinemise ning sissetuleku vahelise negatiivse seose tõi esile ka Sharar (2021) oma analüüsi tulemusena. Elukoha puhul sattus kõige suurema tõenäosusega makseraskustesse Kesk-Eesti elanik ning kõige väiksema tõenäosusega Põhja-Eesti elanik. Autor pakub, et kuna Põhja-Eesti piirkonda jääb pealinn Tallinn, võivad selles piirkonnas olla paremad tingimused töökoha leidmiseks ning kõrgema sissetuleku teenimiseks. Samuti võib lugeda eelisteks paremat infrastruktuuri ning hariduse kättesaadavust.

Laenu iseloomustavatest tunnustest osutusid statistiliselt oluliseks laenuperiood ning intressimäär. Selgus, et mida pikem on laenuperiood, seda suurema tõenäosusega jõuab laenuvõtja maksehäire tekkimiseni. Loogiliselt järeldades jõuab pikema laenuperioodi jooksul ette tulla rohkem ootamatuid olukordi nii näiteks majanduslikul kui ka tervislikul tasandil, mis võivad raskendada laenuvõtja rahalist seisu. Pikema aja jooksul on tõenäolisem, et esineb mingit sorti muutus laenuvõtja käitumismustrites. Intressimääral on kõige olulisem mõju maksehäire tekkimisele. Laenu intressimäära oluline mõju maksejõuetuse tekkimisele ilmnes ka Serrano-Cinca *et al.* (2015), Santoso *et al.* (2020) ning Bao *et al.* (2023) analüüside tulemustes. Mida kõrgem on intressimäär, seda suurema tõenäosusega sattub laenuvõtja makseraskustesse. Siinkohal järeldab autor, et kuna kõrgem intressimäär tõstab igakuise laenumakse summat ning suuremaid igakuiseid kohustusi ongi keerulisem tasuda, siis on ka võlgnevused tõenäolisemad tekkima, sest suurem osa igakuisest sissetulekust läheb laenumakse alla.

Üldjoontes märkas autor analüüsi käigus, et tõenäoliselt oleks võimalik saada veelgi täpsemad ning asjakohasemad tulemused, kui viia sarnane regressioonanalüüs uuesti läbi näiteks kahe või

kolme aasta pärast, mil eelduste kohaselt on ettevõtte andmebaasi kogunenud veel suurema hulga laenude andmed. Siis oleks võimalik valimisse jätta vaid näiteks viimase 5 aasta laenud ning valim oleks mahult ikkagi piisavalt suur, et teha põhjalikumat analüüsi. Antud juhul lisati valimisse kõikide lepingute info aastatest 2014-2023, kuid vahepeal on muutunud mõnevõrra informatsioon, mida laenuvõtjate kohta kogutakse, samuti on periooditi muudatusi tehtud MONEYZENi krediidiskoori mudelis. Need asjaolud võisid mõjutada mõnevõrra mudeli ennustusvõimet. Seega eeldades, et tunnused, mille kohta infot kogutakse, jäävad samaks ning valides analüüsiks kahekolme aasta pärast värskemad andmed, on võimalik luua ka täpsem ning kaasaegsem krediidiriski mudel.

KOKKUVÕTE

Käesolevas lõputöös keskenduti krediidiriski mõjutavate tegurite uurimisele ühisrahastusettevõtte MONEYZEN OÜ andmete baasil. Töö autor seadis eesmärgiks koostada MONEYZENile eraisiku krediidiriski hindav mudel. Enne põhjalikumat analüüsi püstitati töö teoreetilises osas kirjeldatud uuringute baasil ning MONEYZENi soovidest lähtuvalt kaks hüpoteesi:

Hüpotees 1. Laenuvõtja karakteristikud nagu sugu, vanus, elukoht, laste arv, sissetulek ning MONEYZENi krediidiskoor mõjutavad maksehäire tekkimise tõenäosust (on statistiliselt olulised).

Hüpotees 2. Laenu karakteristikud nagu laenu eesmärk, laenusumma, intressimäär ning laenuperiood mõjutavad maksehäire tekkimise tõenäosust (on statistiliselt olulised).

Hüpoteeside testimiseks sai autor ettevõttelt loa kasutada MONEYZENi andmebaasi, kus esialgselt leidsid 2933 laenu kohta andmed alates aastast 2014. Valimisse see-eest peale osaliselt puudulike andmetega lepingute eemaldamist ning andmete korrastamist jäid 2458 lepingu andmed, mille kohta autor tegi statistilise analüüsi, et valimit kirjeldada. Hüpoteesides mainitud tunnuste statistilise olulisuse testimiseks viis autor läbi binaarse logistilise regressioonianalüüsi. Analüüsi esimeses osas selgus, et multikollineaarsust ei esinenud, mistõttu oli võimalik liikuda edasi esialgse mudeli loomise ning edaspidi ebaoluliste tunnuste eemaldamise juurde samm-sammulise tagurpidi kõrvaldamise meetodi abil. Lõpliku mudeli loomise käigus selgus, et statistiliselt ebaolulisteks tunnusteks osutusid sugu, laste arv, laenusumma ning laenueesmärk. Seega sai autor mõlemad hüpoteesid vastu võtta vaid osaliselt.

Tulemused näisid olevat enamjaolt autori silmis loogilised, peale ühe seose, mis puudutas MONEYZENi krediidiskoori. Järelduste põhjal selgus, et mida vanem on laenuvõtja, seda väiksema tõenäosusega tekivad tal makseraskused, mis võib olla seotud elukogemuse ning vastutustunde suurenemisega ajas. MONEYZENi krediidiskoori kasvades suurenes ühtlasi ka maksehäire tekkimise tõenäosus, mis viitab asjaolule, et krediidiskoori hindamismeetodit

ettevõttes saaks uute andmete põhjal tõenäoliselt parandada. Laenuvõtja sissetuleku suurenemisel 1 euro võrra, vähenes tema maksehäire tekkimise tõenäosus 0,02%. Lisaks selgus, et kõige suurema tõenäosusega sattub makseraskustesse Kesk-Eesti elanik ehk laenuvõtja, kes elab kas Järvamaal, Lääne-Virumaal või Raplamaal. Üheks kõige olulisemaks näitajaks makseraskuste tekkimise hindamisel osutus laenu intressimäär, mille kasvamisel maksehäire tekkimise tõenäosus suurenes.

Põhilised takistuskohad, mis autori hinnangul võisid mõjutada lõputöö tulemuste täpsust, olid esiteks pikk ajavahemik, mille jooksul andmed olid kogutud ning muudatused ettevõtte korralduses, mis toimusid selle ajavahemiku jooksul. Eri perioodidel koguti laenuvõtjate kohta mõnevõrra erinevat informatsiooni, samuti on aja möödudes tehtud väiksemaid muudatusi krediitdiskoori hindamise protsessis. Seetõttu tekkis ka lõputöö analüüsi teostamisel osaliselt informatsiooni asümmeetria probleem. Teiseks oleks ilmselt saanud veelgi täpsema mudeli luua, kui tunnuseid, mida laenuvõtjate kohta koguti, oleks olnud rohkem ning uuritud järjepidevamalt. Sealhulgas oleks autorile huvi pakkunud uurida ka näiteks kinnisvara olemasolu, suhtestaatuse, ning ameti mõju maksehäire tekkimisele.

SUMMARY

DEVELOPMENT OF AN INDIVIDUAL CREDIT RISK ASSESSMENT MODEL FOR THE CROWDFUNDING PLATFORM MONEYZEN

Isabel Silla

Peer-to-peer (P2P) lending platforms have revolutionized the financial landscape by facilitating direct lending between investors and borrowers, circumventing traditional intermediaries like banks. Originating in the early 2000s, P2P lending platforms, including MONEYZEN, leverage technological advancements akin to other online auction platforms (e.g., eBay, Amazon, Uber). However, these platforms grapple with the inherent challenge of managing credit risk to ensure investor confidence and platform sustainability.

Crowdfunding platforms like MONEYZEN aim to be attractive to investors by minimizing the risk of borrower default. Achieving this goal hinges on the effective collection and analysis of extensive borrower and loan data. To streamline this process, credit risk assessment models based on existing customer data become instrumental. The research, conducted by an author intimately connected with MONEYZEN, focuses on developing a private credit risk model for the company. The objective is to calculate the probability of a borrower defaulting, with a specific emphasis on payment defaults registered on the 91st day after the first payment.

The research tasks outlined in the work include providing an overview of credit risk calculation options, studying the history and current state of crowdfunding, presenting an overview of MONEYZEN, testing the statistical significance of borrower and loan characteristics, identifying the impact of significant characteristics on default probability, and ultimately creating a credit risk assessment model for private individuals. Two hypotheses guide the analysis: the significance of borrower characteristics (Hypothesis 1) and loan characteristics (Hypothesis 2) in predicting default probability.

Upon testing the hypotheses using the MONEYZEN loan database, the research presents intriguing findings. Notably, borrower age inversely correlates with default probability, suggesting a potential link between increased age, life experience, and financial responsibility. Surprisingly, MONEYZEN's credit score exhibits a positive relationship with default probability, suggesting the need for a reevaluation of the credit scoring method. Other significant factors include borrower income, borrower location (Central Estonia residents being more prone to payment difficulties), and the interest rate of the loan.

Despite providing valuable insights, the research acknowledges potential obstacles, such as changes in data collection processes over time and the need for more comprehensive models. The thesis concludes with the hope that the credit risk model developed will enhance MONEYZEN's ability to evaluate customer creditworthiness, reducing defaults and making the platform more appealing to investors.

The research not only contributes to the understanding of credit risk in P2P lending but also offers practical implications for MONEYZEN's risk management strategy. By addressing identified factors influencing default probability, the platform can refine its credit assessment processes, fostering a more secure and attractive investment environment. The study, conducted internally by an author intimately familiar with MONEYZEN's operations, adds a layer of practicality and relevance to the findings, making them directly applicable to the company's risk management practices.

KASUTATUD ALLIKATE LOETELU

- Ali, A., Daly, K. (2010). *Macroeconomic determinants of credit risk: Recent evidence from a cross country study*. Kättesaadav: <https://www.sciencedirect.com/science/article/abs/pii/S1057521910000268>
- Azouzi, M. A., Bacha, S. (2023) *Do behavioral biases affect Credit risk assessment methods?* Kättesaadav: https://ijms.ut.ac.ir/article_88480_0454cd5a929f091bda7122d843e714c7.pdf
- Bao, T., Ding, Y., Gopal, R., Möhlmann, M. (2023). *Throwing Good Money After Bad: Risk Mitigation Strategies in the P2P Lending Platforms*. Kättesaadav: <https://link.springer.com/article/10.1007/s10796-023-10423-4>
- Bholat, D., Ulrich, A. (2016). *Peer-to-Peer Lending and Financial Innovation in the United Kingdom*. Kättesaadav: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2774297
- Bolton, C. (2009). *Logistic regression and its application in credit scoring*. Kättesaadav: <https://repository.up.ac.za/bitstream/handle/2263/27333/dissertation.pdf?sequence=1&is>
- Calkins, K. G. *Correlation Coefficients*. (2005) Kättesaadav: <https://www.andrews.edu/~calkins/math/edrm611/edrm05.htm>
- Dewani, P., Sippy, M., Punjabi, G., Hatekar, A. (2020). *Credit Scoring : A Comparison between Random Forest Classifier and K- Nearest Neighbours for Credit Defaulters Prediction*. Kättesaadav: <https://d1wqtxts1xzle7.cloudfront.net/64967206/>
- Duarte, J., Siegel, S., Young, L. A. (2023). *The Evolution of P2P Lending*. Kättesaadav: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4476944
- Dzik-Walczak, A., Heba, M. (2021). An implementation of ensemble methods, logistic regression, and neural network for default prediction *in Peer-to-Peer lending*. Kättesaadav: <https://hrcak.srce.hr/file/377565>
- Eesti haldusjaotus*. (2013). *Eesti Entsüklopeedia*. Kättesaadav: http://entsyklopeedia.ee/artikkel/eesti_haldusjaotus
- Gouvêa, M. A., Gonçalves, E. B. (2007). *Credit Risk Analysis Applying Logistic Regression, Neural Networks and Genetic Algorithms Models*. Kättesaadav: <https://www.pomsmeetings.org/confpapers/007/007-0210.pdf>

- Grimes, A. D., Schulz, K. F. (2008). *Making Sense of Odds and Odds Ratios*. Kättesaadav: https://journals.lww.com/greenjournal/abstract/2008/02000/making_sense_of_odds_and_odds_ratios.24.aspx
- Grundke, P. (2003) *Integrating interest rate risk in Credit portfolio models*. Kättesaadav: <https://www.cofar.uni-mainz.de/dgf2003/paper/paper56.pdf>
- Gujarati, D. N., Porter, D. C. (2009) *Basic Econometrics: Fifth edition*. USA: McGraw-Hill Irwin.
- Guo, Y., Whou, W., Luo, C., Liu, C., Xiong, H. (2014). *Instance-based credit risk assessment for investment decisions in P2P lending*. Kättesaadav: https://www.researchgate.net/profile/Yanhong-Guo-3/publication/277981453_Instance-Based_Credit_Risk_Assessment_for_Investment_Decisions_in_P2P_Lending/links/5bd1b90192851cabf266f3f2/Instance-Based-Credit-Risk-Assessment-for-Investment-Decisions-in-P2P-Lending.pdf
- Havrylchyck, O., Verdier, M. (2018). *The Financial Intermediation Role of the P2P Lending Platforms*. Kättesaadav: <https://link.springer.com/article/10.1057/s41294-017-0045-1>
- Kleinbaum, D. G., Klein, M. (2010) *Logistic Regression: A Self-Learning Text: Third edition*. USA: Springer.
- Li, D., Na, S., Ding, T., Liu, C. (2021). *Credit Risk Management of P2P Network Lending*. Kättesaadav: <https://hrcak.srce.hr/clanak/379404%3F>
- Ma, W., Hou, W., Zhang, D. (2021). *A credit risk assessment model of borrowers in P2P lending based on BP neural network*. Kättesaadav: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0255216>
- Mateescu, A. (2015). *Peer-to-Peer Lending*. Kättesaadav: <https://www.datasociety.net/pubs/dcr/PeertoPeerLending.pdf>
- Milne, A., Parboteeah, P. (2016). *The Business Models and Economics of Peer-to-Peer Lending*. Kättesaadav: http://aei.pitt.edu/76108/1/ECRI_RR17_P2P_Lending.pdf
- MONEYZEN OÜ lepingute andmebaas 2023.
- MONEYZEN OÜ majandusaasta aruanne 2022.
- MONEYZEN OÜ äriplaan 2019.
- MONEYZEN OÜ. Inforegister. Kättesaadav: <https://www.inforegister.ee/12541882-MONEYZEN-OU>, 14. november 2023.
- Möllenkamp, N. (2017) *Determinants of Loan Performance in P2P Lending*. Kättesaadav: https://essay.utwente.nl/72876/1/M%C3%B6llenkamp_BA_BMS.pdf
- Paas, T. (1995). *Sissejuhatus Ökonomeetriasse*. Kättesaadav: <https://core.ac.uk/download/154745737.pdf>

- Ravina, E. (2007) *Beauty, Personal Characteristics and Trust in Credit Markets*. Kättesaadav: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=972801
- Rivera, C. E. C. (2018). *Determinants of Default in P2P Lending: The Mexican Case*. Kättesaadav: <https://dialnet.unirioja.es/servlet/articulo?codigo=6325924>
- Rudd, J. M., Priestley, J. L. (2017). *A Comparison of Decision Tree with Logistic Regression Model for Prediction of Worst NonFinancial Payment Status in Commercial Credit*. Kättesaadav: https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1004&context=datap_hdgreylit
- Santoso, W., Trinugroho, I., Risfandy, T. (2020). *What Determine Loan Rate and Default Status in Financial Technology Online Direct Lending? Evidence from Indonesia*. Kättesaadav: <https://repository.feb.uns.ac.id/dok/publikasi/1153.pdf>
- Sauga, A. *Binaarne logit mudel*. (2023) Kättesaadav: https://www.sauga.pri.ee/gretl/logit_binary.html, 29.11.2023
- Serrano-Cinca, C., Gutiérrez-Nieto, B., López-Palacios, L. (2015). *Determinants of Default in P2P Lending*. Kättesaadav: <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0139427&type=printable>
- Sharar, M. L. (2021). *Online Peer-to-Peer Lending: Determinants of Loan Performance*. Kättesaadav: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3785323
- Silla, I. *Programmi Gretl toorväljavõtted*. (2023) Kättesaadav: <https://docs.google.com/document/d/1PmFXovojTgSUHwsXxGQnplnCAyDxR-0XfZKOF46IwR4/edit?usp=sharing>, 29.11.2023
- Suryono, R. R., Purwandari, B., Budi, I. (2019). *Peer to Peer (P2P) Lending Problems and Potential Solutions: A Systematic Literature Review*. Kättesaadav: <https://www.sciencedirect.com/science/article/pii/S1877050919318265>
- Tupangiu, L. (2017). *Information Asymmetry and Credit risk*. Kättesaadav: https://web.archive.org/web/20180421093912id_/http://www.financejournal.ro/fisiere/revista/76554199516_Tupangiu_en.pdf
- Vastutustundlik laenamaine. (2023). MONEYZEN. Kättesaadav: <https://moneyzen.eu/meist/>, 14. november, 2023.
- Vodová, P. (2003). *Credit Risk as a Cause of Banking Crises*. Kättesaadav: https://www.researchgate.net/profile/Pavla-Klepkova-Vodova/publication/228420636_Credit_Risk_as_a_Cause_of_Banking_Crises/links/02bfe5110df238b65e000000/Credit-Risk-as-a-Cause-of-Banking-Crises.pdf
- Wang, R. (2015). *Peer-to-Peer Lending Industry and Risk Control Measures*. Kättesaadav: <https://digitalworks.union.edu/cgi/viewcontent.cgi?article=1399&context=theses>

Zhang, W., Wang, C., Zhang, Y., Wang, J. (2020). Credit risk evaluation model with textual features from loan descriptions for P2P lending. Kättesaadav: <https://www.sciencedirect.com/science/article/pii/S1567422320300661>

Zheng, C., Sarker, N., Nahar, S. (2018). Factors affecting bank credit risk: An empirical insight. Kättesaadav: https://d1wqtxts1xzle7.cloudfront.net/77215513/Vol_208_2_3-libre.pdf?1640309822=&response-content-disposition=inline%3B+filename%3DFactors_affecting_bank_credit_risk_An_em.pdf&Expires=1702588018&Signature=K6wg6YsLl9bBLVqcGFP15~AJ0Vz~XLQ3apd0oI g0SdpBtVPIhVwPEB4Gd5FhvGSG6oCDaW0pfdlW2QYEqF4tCGKi536al8BGQ8hnu LoyG4tS4LSZyUkpr7YUHWMRHUeLQv4NKOSu30xwpcN~DYKYWY8Rbo00Hcz0 uEz3HIA-HKv0jfcYshDJxUsOYdiYZyo~eZAChnFx0MFXuktINsFCh7uVOw3zyPiphfYiCkP4b0kNBQTqwF57ggfJlZqXWLK wHKf7EoKag-CC8mE9lOZ9aLO67PWP-khg1-PlSmDluhMcyxjG7mGYXRHQL2wLYJ9AYx5-N5OWIVl3M53QtEVA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

LISAD

Lisa 1. Tunnused

Tunnuse selgitus	Lühend
0 – laen on viivises alla 91 päeva, 1 – laen on viivises 91 või rohkem päeva	MAKSEHÄIRE
Laenuvõtja vanus (aastates)	VANUS
0 – laenuvõtja on mees, 1 – laenuvõtja on naine	SUGU
Laenuvõtja elukoht: 1 – Põhja-Eesti, 2 – Lõuna-Eesti, 3 – Kesk-Eesti, 4 – Lääne-Eesti, 5 – Kirde-Eesti	ELUKOHT
1, kui laenuvõtja elab Põhja-eesis, 0, kui elab mujal	DELUKOHT_1
1, kui laenuvõtja elab Lõuna-eesis, 0, kui elab mujal	DELUKOHT_2
1, kui laenuvõtja elab Kesk-eesis, 0, kui elab mujal	DELUKOHT_3
1, kui laenuvõtja elab Lääne-eesis, 0, kui elab mujal	DELUKOHT_4
1, kui laenuvõtja elab Kirde-eesis, 0, kui elab mujal	DELUKOHT_5
Laenuvõtja laste arv	LASTEARV

Laenuvõtja MONEYZENi krediidiskoor vahemikus 505-1000	KREDIIDISKOOR
0 – remont, 1 – tarbimine (väikelaen), 2 – tervis, 3 – refinantseerimine, 4 – auto, 5 – reisimine, 6 – haridus, 7 - ettevõtlus	LAENUEESMÄRK
1, kui on remondilaen, 0 kui on muu otstarve	DLAENUEESMÄRK_1
1, kui on tarbimisolaen, 0 kui on muu otstarve	DLAENUEESMÄRK_2
1, kui on terviselaen, 0 kui on muu otstarve	DLAENUEESMÄRK_3
1, kui on refinantseerimislaen, 0 kui on muu otstarve	DLAENUEESMÄRK_4
1, kui on autolaen, 0 kui on muu otstarve	DLAENUEESMÄRK_5
1, kui on reisilaen, 0 kui on muu otstarve	DLAENUEESMÄRK_6
1, kui on õppelaen, 0 kui on muu otstarve	DLAENUEESMÄRK_7
1, kui on ärilaen, 0 kui on muu otstarve	DLAENUEESMÄRK_8
Laenu kestvus kuudes	LAENUPERIOOD
Laenult makstav intress protsentides	INTRESSIMÄÄR
Laenuks saadud summa eurodes	LAENUSUMMA
Laenuvõtja igakuine keskmine sissetulek eurodes	SISSETULEK

Allikas: autori poolt koostatud toorväljavõtete põhjal

Lisa 2. Peamised statistilised näitajad

	Min	Max	Keskmine	Mediaan	Standardhälve	Mood	Variats.
Vanus	19	73	38	36	12	33	0,31
Krediidiskoor	505	1000	824	825	100	850	0,12
Sissetulek	342	8451	1213	1060	655,17	1000	0,54
Laenusumma	100	10000	1910	1000	2137,81	1000	1,12
Intressimäär	14,16%	36,30%	23,21%	23,17%	3,78%	20,80%	0,16
Laenuperiood	3	108	33	24	20	24	0,61

Allikas: autori poolt koostatud

Lisa 3. Korrelatsioonimaatriks

Correlation Coefficients, using the observations 1 - 2458							
5% critical value (two-tailed) = 0,0395 for n = 2458							
VANUS	LASTEARV	KREDIIDI SKOOR	LAENU PERIOOD	INTRESSI MÄÄR	LAENU SUMMA	SISSE TULEK	
1	0,0745	0,0288	0,1570	-0,0791	0,1841	0,0893	VANUS
	1	-0,0596	0,0996	-0,0256	0,1241	0,1031	LASTEARV
		1	-0,2516	-0,6057	-0,2162	0,0604	KREDIIDI SKOOR
			1	-0,0109	0,7843	-0,0290	LAENU PERIOOD
				1	-0,1310	-0,0110	INTRESSI MÄÄR
					1	0,0883	LAENU SUMMA
						1	SISSETULEK

Allikas: autori poolt koostatud

Lisa 4. Esialgne mudel

Logit, using observations 1-2458 (n = 2458)					
Dependent variable: MAKSEHÄIRE					
Standard errors based on Hessian					
	coefficient	Std. error	z	p-value	
const	-5,48957	0,999937	-5,490	4,02 ⁻⁸	***
VANUS	-0,0325834	0,00495505	-6,576	4,84 ⁻¹¹	***
SUGU	-0,189709	0,106819	-1,776	0,0757	*
LASTEARV	0,0569690	0,0535764	1,063	0,2876	
KREDIIDISKOOR	0,00377626	0,000745645	5,064	4,10 ⁻⁷	***
LAENUPERIOOD	0,0211905	0,00418032	5,069	4,00 ⁻⁷	***
INTRESSIMÄÄR	6,88577	1,89567	3,632	0,0003	***
LAENUSUMMA	1,66231 ⁻⁵	4,14498 ⁻⁵	0,4010	0,6884	
SISSETULEK	- 0,000251831	9,69734 ⁻⁵	-2,597	0,0094	***
DELUKOHT_2	0,198646	0,128925	1,541	0,1234	
DELUKOHT_3	0,508123	0,162701	3,123	0,0018	***
DELUKOHT_4	0,103268	0,180550	0,5720	0,5673	
DELUKOHT_5	0,267432	0,189906	1,408	0,1591	
DLAENUEESMÄRK_2	0,158144	0,167991	0,9414	0,3465	
DLAENUEESMÄRK_3	0,116396	0,261849	0,4445	0,6567	
DLAENUEESMÄRK_4	-0,219766	0,214228	-1,026	0,3050	
DLAENUEESMÄRK_5	-0,109298	0,306704	-0,3564	0,7216	
DLAENUEESMÄRK_6	-1,40763	0,746169	-1,886	0,0592	*
DLAENUEESMÄRK_7	0,0667087	0,694373	0,09607	0,9235	
DLAENUEESMÄRK_8	0,150262	0,540551	0,2780	0,7810	
Mean dependent var	0,209113	S.D. dependent var		0,406758	
McFadden R-squared	0,057536	Adjusted R-squared		0,041668	
Log-likelihood	-1187,892	Akaike criterion		2415,784	
Schwarz criterion	2531,926	Hannan-Quinn		2457,986	
Number of cases correctly 'predicted'			1953 (79,5%)		
f(beta'x) at mean of independent vars			0,154		
Likelihood ratio test: Chi square ()			145,038 [0,0000]		
			Predicted		
			0	1	
Actual	0	1939		5	
	1	500		14	

Allikas: autori poolt koostatud toorväljavõtete põhjal

Lisa 5. VIF varieeruvusindeksi analüüs

Variance Inflation Factors	
Minimum possible value = 1.0	
Values > 10.0 may indicate a collinearity problem	
VANUS	1,057
SUGU	1,064
LASTEARV	1,057
KREDIIDISKOOR	1,817
LAENUPERIOOD	1,154
INTRESSIMÄÄR	1,681
SISSETULEK	1,083
DELUKOHT_2	1,217
DELUKOHT_3	1,135
DELUKOHT_4	1,130
DELUKOHT_5	1,116

Allikas: autori poolt koostatud toorväljavõtete põhjal

Lisa 6. Kõik mudelid tabelina

Logit estimates					
Dependent variable: MAKSEHÄIRE					
	(1)	(2)	(3)	(4)	(5)
const	-5,490*** (0,9999)	-5,820*** (0,9726)	-5,691*** (0,9513)	-5,784*** (0,9569)	-5,736*** (0,9546)
VANUS	-0,03258*** (0,004955)	-0,03310*** (0,004879)	-0,03301*** (0,004883)	-0,03253*** (0,004796)	-0,03348*** (0,004759)
SUGU	-0,1897*** (0,1068)	-0,1932* (0,1060)	-0,2101** (0,1054)	-0,1781* (0,1051)	
LASTEARV	0,05697 (0,05358)	0,06326 (0,05329)	0,08248 (0,05266)		
KREDIIDISKOOR	0,003776*** (0,0007456)	0,004110*** (0,0007168)	0,004143*** (0,0007013)	0,004071*** (0,0007031)	0,003961*** (0,0006977)
LAENUPERIOOD	0,02119*** (0,004180)	0,02016*** (0,004107)	0,01952*** (0,002671)	0,02015*** (0,002683)	0,02019*** (0,0022682)
INTRESSIMÄÄR	6,886*** (1,896)	7,677*** (1,846)	7,884*** (1,776)	7,609*** (1,783)	7,449*** (1,776)
LAENUSUMMA	1,662 ⁻⁵ (4,145 ⁻⁵)	-2,775 ⁻⁵ (3,917 ⁻⁵)			
SISSETULEK	-0,0002518*** (9,697 ⁻⁵)	-0,0002440** (9,621 ⁻⁵)	-0,0002757*** (9,544 ⁻⁵)	-0,0002282** (9,348 ⁻⁵)	- 0,0002029** (9,166 ⁻⁵)
DELUKOHT_2	0,1986 (0,1289)	0,2125 (0,1280)		0,2271* (0,1273)	0,2321* (0,1272)
DELUKOHT_3	0,5081*** (0,1627)	0,5259*** (0,1618)		0,5440*** (0,1609)	0,5493*** (0,1607)
DELUKOHT_4	0,1033 (0,1806)	0,1246 (0,1797)		0,1397 (0,1793)	0,1330 (0,1791)
DELUKOHT_5	0,2674 (0,1899)	0,2641 (0,1882)		0,2804 (0,1876)	0,2955 (0,1873)
DLAENUEESMÄRK_2	0,1581 (0,1680)				
DLAENUEESMÄRK_3	0,1164 (0,2618)				
DLAENUEESMÄRK_4	-0,2198 (0,2142)				
DLAENUEESMÄRK_5	-0,1093 (0,3067)				
DLAENUEESMÄRK_6	-1,408* (0,7462)				
DLAENUEESMÄRK_7	0,06671 (0,6944)				
DLAENUEESMÄRK_8	0,1503 (0,5406)				

Allikas: autori poolt koostatud toorväljavõtete põhjal

Lisa 7. Lõplik mudel

Logit, using observations 1-2458 (n = 2458)					
Dependent variable: MAKSEHÄIRE					
Standard errors based on Hessian					
	coefficient	Std. error	z	p-value	
const	-5,73607	0,954629	-6,009	$1,87^{-9}$	***
VANUS	-0,0334761	0,00475925	-7,034	$2,01^{-12}$	***
KREDIIDISKOOR	0,00396142	0,00069767 1	5,678	$1,36^{-8}$	***
LAENUPERIOOD	0,0201887	0,00268206	7,527	$5,18^{-14}$	***
INTRESSIMÄÄR	7,44938	1,77633	4,194	$2,74^{-5}$	***
SISSETULEK	-0,000202881	$9,16642^{-5}$	-2,213	0,0269	**
DELUKOHT_2	0,232063	0,127219	1,824	0,0681	*
DELUKOHT_3	0,549349	0,160690	3,419	0,0006	***
DELUKOHT_4	0,133006	0,179073	0,7427	0,4576	
DELUKOHT_5	0,295505	0,187264	1,578	0,1146	
Mean dependent var	0,208773	S.D. dependent var	0,406515		
McFadden R-squared	0,051894	Adjusted R-squared	0,043966		
Log-likelihood	-1195,893	Akaike criterion	2411,786		
Schwarz criterion	2469,873	Hannan- Quinn	2432,890		
Number of cases correctly 'predicted'	1957 (79,5%)				
f(beta'x) at mean of independent vars	0,156				
Likelihood ratio test: Chi square ()	130,912 [0,0000]				
		Predicted			
		0		1	
Actual	0	1945		3	
	1	502		12	

Allikas: autori poolt koostatud toorväljavõtete põhjal

Lisa 8. Šansside suhted

Odds-ratios for MAKSEHÄIRE		
Variable	Odds-ratio	95% conf interval
VANUS	0,9671	[0,958, 0,976]
KREDIIDISKOOR	1,0040	[1,003, 1,005]
LAENUPERIOOD	1,0204	[1,015, 1,026]
INTRESSIMÄÄR	1718,7984	[52,870, 55878,154]
SISSETULEK	0,9998	[1,000, 1,000]
DELUKOHT_2	1,2612	[0,983, 1,618]
DELUKOHT_3	1,7321	[1,264, 2,373]
DELUKOHT_4	1,1423	[0,804, 1,623]
DELUKOHT_5	1,3438	[0,931, 1,940]

Allikas: autori poolt koostatud toorväljavõtete põhjal

Lisa 9. Tõenäosuse marginaalväärtused

Logit, using observations 1-2458				
Dependent variable: MAKSEHÄIRE				
Standard errors based on Hessian				
	coefficient	Std. error	z	slopes
const	-5,73607	0,954629	-6,009	
VANUS	-0,0334761	0,00475925	-7,034	-0,00521031
KREDIIDISKOOR	0,00396142	0,000697671	5,678	0,000616567
LAENUPERIOOD	0,0201887	0,00268206	7,527	0,00314223
INTRESSIMÄÄR	7,44938	1,77633	4,194	1,15944
SISSETULEK	-0,000202881	9,16642 ⁻⁵	-2,213	-3,15769 ⁻⁵
DELUKOHT_2	0,232063	0,127219	1,824	0,0373169
DELUKOHT_3	0,549349	0,160690	3,419	0,0966156
DELUKOHT_4	0,133006	0,179073	0,7427	0,0213729
DELUKOHT_5	0,295505	0,187264	1,578	0,0494768

Allikas: autori poolt koostatud toorväljavõtete põhjal

Lisa 10. Lihtlitsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Isabel Silla

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose Erasiku krediidiriski hindava mudeli koostamine ühisrahastusportaalile MONEYZEN,

mille juhendaja on Kristjan Liivamägi,

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.

20.12.2023

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. jq 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.