

VILNIAUS UNIVERSITETAS
EKONOMIKOS IR VERSLO ADMINISTRAVIMO FAKULTETAS
FINANSŲ KATEDRA

Eimantas PALIONIS

Verslo ekonomika

MAGISTRO DARBAS

**FIZINIŲ ASMENŲ KREDITINGUMO VERTINIMO METODŲ
VERTINIMAS VIENOS IŠ LIETUVOJE VEIKIANČIŲ FINANSINĖS
PASLAUGAS TEIKIANČIŲ ĮMONIŲ ATVEJU**

**ASSESSMENT OF METHODS FOR CREDITWORTHINESS EVALUATION
OF CONSUMER CREDIT IN THE CASE OF ONE OF THE COMPANIES
PROVIDING FINANCIAL SERVICES IN LITHUANIA**

Leidžiama ginti _____
(parašas)

Katedros vedėjas prof. dr. A. Paškevičius

Magistrantas _____
(parašas)

Darbo vadovas: _____
(parašas)

prof. dr. A. Paškevičius

Darbo įteikimo data: _____

Registracijos Nr. _____

Vilnius, 2019

TURINYS

ĮVADAS	3
1. FIZINIŲ ASMENŲ KREDITO RIZIKOS VALDYMO SAMPRATA IR SVARBA BANKINIAME SEKTORIUJE, FIZINIŲ ASMENŲ KREDITINGUMO VERTINIMO METODAI 5	
1.1. Rizikos apibrėžimai, samprata ir svarba versle	5
1.2. Finansinės rizikos rūšys	9
1.3. Kredito rizika ir jos valdymas bankuose.....	14
1.4. Metodų apžvalga	16
1.5. Metodų apibendrinimas	24
2. FIZINIŲ ASMENŲ KREDITINGUMO VERTINIMO TYRIMO METODOLOGIJA	28
2.1. Tyrimo imtis.	29
2.2. Priklausomas kintamasis („geras“ ir „blogas“ klientas)	33
2.3. Nepriklausomi kintamieji ir duomenų paruošimas.	34
2.4. Modelių efektyvumo patikrinimo metodai.	40
3. FIZINIŲ ASMENŲ KREDITINGUMO VERTINIMO ANALIZĖ IR TYRIMO REZULTATAI	41
3.1. Kintamųjų svarba	41
3.2. Tyrimo rezultatai.....	42
IŠVADOS	46
SUMMARY	50
LITERATŪROS SĄRAŠAS	52
Priedai	57

ĮVADAS

Temos aktualumas. Kreditingumo vertinimo metodai pirmą kartą buvo pristatyti dar 1940m. ir bėgant metams jie tampa vis svarbesnis sėkmingų finansinių įstaigų įrankis. 1960m., kai buvo sukurtos kreditinės kortelės, kreditingumo vertinimo metodai tapo ypač svarbūs bankams. Šiais laikais kreditingumo vertinimo modeliai yra plačiai naudojami kaip įrankiai vartojimo kredito finansavimo ar nefinansavimo sprendimams priimti ar net prognozuoti vertinamų įmonių bankrotą. Pagrindinis kreditingumo vertinimo tikslas yra naudojant praeities vertinimų duomenis sukurti modelį, kuris nustato kliento kreditingumo lygtį (reitingą). Kreditingumo vertinimo modeliai reikšmingai tobulėja su vis augančiu surenkamos informacijos kiekiu. Efektyvi kreditingumo vertinimo sistema gali padėti apsaugoti finansų sektoriaus įmones nuo bankroto, padėti maksimizuoti pelną ir minimizuoti iš kredito teikimo kylančią klientų nemokumo riziką. Šiuo metu yra sukurta nemažai skirtingų kreditingumo vertinimo modelių, kurie duoda skirtingus rezultatus, skirtinguose klientų segmentuose.

Kiekviena išduota paskola banką įpareigoja prisiimti vieną svarbiausių savo veikloje rizikų – kredito riziką. Neatsakinga bankų veikla, skolinimas ir negebėjimas tinkamai įvertinti prisiimamos rizikos, sukėlė ir 2007 metais JAV prasidėjusią krizę. Krizės padariniai parodė, koks iš tiesų yra svarbus atsakingas kredito rizikos vertinimas ir valdymas. O ypač svarbus jis tapo komerciniams bankams, finansines paslaugas teikiančioms įmonėms ir bankų priežiūros institucijoms. Finansų sektorių prižiūrintys reguliatoriai vis krapščiau ir detaliau vykdo kredito rizikos valdymo procedūrų priežiūrą, norėdami užtikrinti sklandų ir saugų finansinio sektoriaus vystymąsi.

Skirtingi kreditingumo vertinimo metodai naudojami priklausomai nuo to, kas ir koku tikslu atlieka vertinimą. Vertinant galimybę suteikti vartojimo ar būsto kreditą privačiam asmeniui, taikomi modeliai įvertinantys konkretaus asmens ir jo šeimos finansinius išteklius, turimus įsipareigojimus ir socialinius duomenis (vaikų skaičius, gyvenamoji vieta ir pan.)

Šio *magistro baigiamojo darbo objektas* yra fizinių asmenų vartojimo kreditų rinka.

Mokslinė problema: Problema, su kuria susiduria bankai ir kitos finansinės institucijos Lietuvoje – kaip nustatyti, kuris klientas bus „geras“, o kuris „blogas“ dar prieš suteikiant jam finansines paslaugas, kiek laiko turi praeiti, kad būtų galima sakyti, jog klientas yra „blogas“, kokie kintamieji geriausiai apibūdina „blogą“ ir „gerą“ klientą ir kokį metodą naudoti, kad būtų pasiektas maksimaliai geras rezultatas. Juridinių asmenų vertinimas yra plačiai nagrinėjama tema Lietuvoje, o fizinių asmenų vertinimo analizė, nors ne ką mažiau svarbi, nėra taip plačiai nagrinėjama tema. Finansines paslaugas Lietuvoje teikiančios įmonės neretai turi pirkti fizinių asmenų kreditingumo

vertinimo modelius iš trečiųjų šalių, nes pačios turi labai menką suvokimą šiuo klausimu. Todėl **pagrindinė problema** yra nustatyti, kuris modelis yra geriausiai tinkamas Lietuvos rinkoje veikiančiai finansines paslaugas teikiančiai įmonei.

Darbo tikslas: įvertinti praktikoje naudojamus fizinių asmenų kreditingumo vertinimo metodus, suprasti rizikos vertinimo svarbą versle ir konkrečiai bankiniame sektoriuje, nustatyti, kuris metodas geriausiai tinka prognozuojant klientų nemokumą konkrečios Lietuvoje veikiančios finansines paslaugas teikiančios įmonės atveju.

Darbo uždaviniai:

1. Rizikos apibrėžimo, sampratos analizė, rizikos valdymo svarba versle, bankiniame sektoriuje;
2. Fizinių asmenų kreditingumo vertinimo modelių/metodų apžvalga, aprašomoji statistika;
3. Empirinė analizė - tiesinės diskriminantinės analizės (LDA), logistinės regresijos (LOG), neuroninių tinklų (NNET), apibendrinto tiesinio modelio (GLM), k-artimiausio kaimyno (KNN) ir sprendimų miško (RF) modelių pritaikymas konkrečios Lietuvoje veikiančios finansines paslaugas teikiančio įmonės atveju;
4. Išvados ir pasiūlymai.

Darbo metodika. Mokslinės literatūros šaltinių analizė, kredito rizikos vertinimo modelių, naudojamų fizinių asmenų vertinime, vertinimas ir modelių aprašomoji statistika. Empiriniame tyrime atliekama tiesinės diskriminantinės analizės (LDA), logistinės regresijos (LOG), neuroninių tinklų (NNET), apibendrinto tiesinio modelio (GLM), k-artimiausio kaimyno (KNN) ir sprendimų miško (RF) modelių analizė bei pateikiamos išvados ir pasiūlymai.

1. FIZINIŲ ASMENŲ KREDITO RIZIKOS VALDYMO SAMPRATA IR SVARBA BANKINIAME SEKTORIJE, FIZINIŲ ASMENŲ KREDITINGUMO VERTINIMO METODAI

1.1. Rizikos apibrėžimai, samprata ir svarba versle

Žodis rizika įvairiuose žodynuose yra kildinamas iš lotyniškų žodžių *resicum*, *risicum* ir *riscus*, kurie reiškė statų skardį ar rifą. Savo ruožtu, šie lotyniški žodžiai yra kilę iš graikiškų žodžių *rhizikon*, *rhiza*, kurie buvo metaforiškai naudojami apibūdinti sunkumams, kurių reikia vengti jūroje. Pasak Omero Ertekino (2010), nuo 16-o amžiaus šis terminas įgavo reikšmę susijusią su tam tikru naudos gavimu: 1507 metais vokiečių kalboje žodis *rysigo* įgavo prasmę versle, jis reiškė „išdrįsti, imtis, verslumas, tikėtis ekonominės sėkmės“. Kinai savo kabojė šiam žodžiui dar pridėjo „galimybės“ prasmę. O. Ertekinas taip pat teigia, kad labiausiai tikėtina, jog žodis rizika į arabų kalbą pateko per Viduržemio jūros regioną kaip *rizk*, kuris reiškė „viskas ką Dievas suteikė pragyvenimui“. Šiame kontekste rizika negali būti visiškai valdoma žmonijos ir žmogus gali tiesiog stengtis iš visų jėgų ir tikėtis sulaukti palankaus rezultato. Šiais laikais vienas tiksliausių rizikos apibrėžimų yra šis: „Pavojų, galinčių užkirsti kelią mūsų tikslams, nustatymas ir įvertinimas“ (Bernstein 1998).

Rizika yra neišvengiamas bet kokios ūkinės veiklos elementas. Skirtingose veiklos srityse (ekonomikoje, statistikoje, draudimo teorijoje, verslo aplinkoje) rizika yra apibrėžiama skirtingai, taip pat skirtingi autoriai, moksliniuose darbuose, skirtingai apibrėžia kas yra rizika. Vieni autoriai riziką apibūdina kaip emocinę būseną – ryžtą siekti tikslo, žinant, kad galima jo ir nepasiekti (Urnėžius 2001) arba ryžimąsi nepaisyti galimų neigiamų atsitiktinių aplinkybių padarinių (Vaitkevičiūtė 2001). Kiti riziką supranta kaip neapibrėžtumą, neapibrėžtą situaciją, galinčią turėti neigiamų pasekmių (Rescher 1983) ar įvykio vyksmo neapibrėžtį, kuri galėtų turėti įtakos siekiant tikslų (Hilson 2016). Dabartinės lietuvių kalbos žodynas riziką apibrėžia kaip su galimu pavojumi susijusį dalyką, pasiryžimą vykdyti kažką pavojingo. The Oxford dictionary riziką apibrėžia kaip situaciją, keliančią pavojų, tikimybę, kad atsitiks kažkas nepageidaujamo, galimybę nukentėti arba kažką prarasti, finansinės netekties galimybę. Tačiau autorės R. Martinkutė-Kaulienė ir V. Stasytytė (2018) teigia, kad rizikos ir neapibrėžties sąvokas reikalinga atskirti, jų manymu neapibrėžtimi (netikrumu) turėtų būti laikoma tokia rizika, kuriai labai sunku (ar neįmanoma) apskaičiuoti tikimybės. Rizika ir netikrumas pirmą kartą mokslinėje literatūroje buvo atskirti 1921m., kai F. H. Knight savo darbe „Rizika, neapibrėžtumas ir pelnas“ pareiškė, kad jeigu yra nežinoma kas atsitiks, bet žinomos atsitikimų galimybės, tai ir yra rizika, jeigu nežinomos net ir galimybės, tai – neapibrėžtumas (netikrumas). Netikrumas yra abstraktus dalykas, neapčiuopiamas procesas, jis nepriklauso nuo žmogaus valios ar noro ir jo buvimo pripažinimo ir supratimo konkrečioje situacijoje.

Tai subjektyvi sąvoka, todėl jo negalima tiksliai įvertinti. Tačiau yra ir mokslininkų, kurie prieštarauja rizikos ir neapibrėžtumo atskyrimui ir vartoja šiuos žodžius kaip sinonimus. Pavyzdžiui pasak Aleknevičienės (1997), sąvokų „rizika“ ir „neapibrėžtumas“ atskyrimas yra netikslingas, kadangi: „ir vienu, ir kitu atveju egzistuoja rezultatų nukrypimo galimybė, ir nuo to, ar turima tikimybinė informacija, ar ne, priklauso tik prognozuojamų rezultatų patikimumo laipsnis“. Plėtojant toliau rašoma, kad „dažniausiai patenkama į situaciją, kai faktiškų rezultatų nukrypimo nuo prognozuojamų tikimybė su pakankamu patikimumo laipsniu nežinoma, t.y. į situaciją tarp rizikos ir neapibrėžtumo“.

Verslo rizikai apibrėžti taip pat nėra vieningos nuomonės, tačiau dažniausiai rizika yra siejama su neigiamomis emocijomis ir nuostolio galimybe. Žvelgiant plačiąja prasme, verslas ir yra nuolatinis rizikos valdymas. Kiekviena įmonė kasdien susiduria su sprendimais, kurie vienaip ar kitaip daro įtaką prisiimamos rizikos lygio kitimui. Pasak Pasaulinės rizikos specialistų asociacijos (GARP 2012) vienos įmonės rizika prisiima pasyviai, kitos bando pasiekti pranašumą rinkoje prisiimdamos apskaičiuotą ir pasvertą riziką. Dalis autorių riziką suvokia kaip aiškiai paskaičiuojamą rodiklį. Autoriai Horcher, Karen A. (2005) teigia, kad rizika suteikia pagrindą galimybėms ir turėtų būti apibrėžiama, ne kaip galimybė, o kaip tikimybė patirti nuostolį.

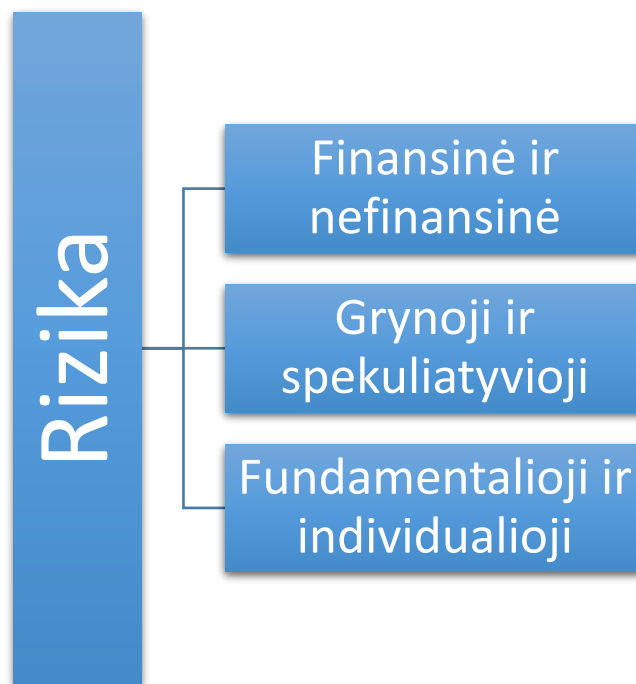
GARP (2012) finansinę riziką apibrėžia kaip netikėtą, nepastovų rezultatą, kuris gali būti išreikštas turto, nuosavybės ar pajamų verte. Pasak R. Urniežiaus (2001), pelną galima suvokti ir kaip atlygį už sėkmingą rizikos prisiėmimą. Išskiriamos dvi pagrindinės rizikos grupės: verslo rizika ir finansinę rizika.

Nepaisant to, kaip skirtingai yra suvokiama rizika, visi autoriai sutaria, kad riziką reikia kontroliuoti ir geriausi verslo rezultatai ir pranašumai rinkoje yra sėkmingo rizikos valdymo pasekmės. Nors pati rizikos sąvoka įprastai turi neigiamą prasmę, tačiau versle į riziką reiktų žiūrėti ne tik kaip į visą griauunančią grėsmę, tačiau ir kaip į galimą verslo partnerį, susiklosčius palankioms aplinkybėms rizika gali duoti daug didesnę naudą, negu neprisiimant jokios rizikos arba prisiimant tik labai mažą jos lygį.

Galima išskirti tris dažniausiai mokslinėje literatūroje analizuojamus rizikos klasifikavimo būdus: finansinę ir nefinansinę, grynoji ir spekuliatyvioji, fundamentalioji ir individualioji (1 pav.).

- Finansinė ir nefinansinė rizika:
 - Finansinė rizika – išmatuojama pinigine išraiška (pvz. Padaryta žala turtui);
 - Nefinansinė rizika – neišmatuojama pinigais, ją dažnai lydi neapibrėžtumas (pvz. Po fizinės traumos išliekantys moraliniai išgyvenimai, kurių neįmanoma išmatuoti pinigais).
- Grynoji ir spekuliatyvioji rizika:

- Grynoji rizika – rizika, kad įvykus įvykiui bus patirtas nuostolis arba bus gautas nulinis rezultatas (pvz. Stichinės nelaimės rizika);
- Spekuliatyvioji rizika – rizika, kad įvykus įvykiui bus patirtas nuostolis, nulinis rezultatas arba pelnas (pvz. Pinigų investavimas į akcijas ar obligacijas).
- Fundamentalioji ir individualioji rizika:
 - Fundamentalioji rizika – rizika, kurios neįmanoma kontroliuoti, o jos padarinius jaučia didelės žmonių grupės (pvz. Stichinės nelaimės, politinės intervencijos, karai);
 - Individualioji rizika – ši rizika kyla iš individualių priežasčių ir jų pasekmės veikia konkrečius individus (pvz. Vagystė, gaisras).



1 pav. Rizikos klasifikavimas

Visos išvardintos rizikos grupės yra vienodai svarbios ir aktualios, tačiau šiame darbe toliau daugiausiai dėmesio bus skiriama būtent finansinei rizikai, jos analizei ir iš jos valdymo išplaukiančioms teorijoms.

Rizikos valdymo veikla atsirado 20 amžiaus pradžioje ir buvo beveik išskirtinai siejama su draudimo veikla, norint apdrausti žmones nuo įvairių nelaimių. Finansinės rizikos vystymasis intensyviai prasidėjo tik 20 amžiaus pabaigoje (Dionne 2013). Rizikos valdymo veiklos atsiradimui ypatingą svarbą turėjo vadybos mokslininko - Henry Fayol darbai. Mokslininkas vienoje iš savo knygų išleistų 1916 m. pirmą kartą paminėjo rizikos valdymą tarp svarbiausių įmonės valdymo funkcijų. Nors tokių atsiradimai yra svarbus, rizikos valdymo fenomenas, toks koks praktikuojamas šiandien, greičiausiai atsirado po-II pasaulinio karo kai Jungtinėse Amerikos Valstijose buvo įkurta

„Rizikos ir draudimo valdymo organizacija“, kuri pradėjo tarpininkauti su Europos ir Azijos rizikos specialistais ir paskatino tarptautinės profesionalios bendruomenės formavimą.

Rizikos valdymas neabejotinai yra viena svarbiausių bet kokios įmonės valdymo funkcijų. Netinkamas potencialių rizikų valdymas įmonėje gali būti trumpalaikių verslo trikdžių, prarastų galimybių, o blogiausiu atveju net ir įmonės bankroto priežastimi. Rizikos valdymas versle yra ypač svarbus sėkmingos įmonės veiklai, jis saugo įmonę nuo pernelyg rizikingų sandorių ar nuostolių, padeda išlaikyti gerą reputaciją ir turėti stabilumą.

Visgi svarbiausia versle – sukurti vertę tiek ir klientams, kurie naudojami to verslo paslaugomis, tiek ir verslo savininkams. Protingas rizikos valdymas gali padidinti įmonės vertę. Pasak Nance, Smith ir Smithson (1993) rizikos valdymas apsidraudžiant sumažina konfliktų grėsmę bei laukiamus mokesčius, o taip pat padidina patrauklių investavimo progų pasinaudojimo galimybę. R. Martinkutė-Kaulienė ir V. Stasytė (2018) išskiria šiuos rizikos valdymo privalumus:

- Didėja tikslų pasiekimo tikimybė;
- Skatinamas aktyvus valdymas;
- Vadovai ir darbuotojai suvokia poreikį identifikuoti ir mažinti rizikas organizacijoje;
- Geriau nustatomos galimybės ir grėsmės;
- Organizacija atitinka svarbius teisinius reikalavimus ir tarptautinius standartus;
- Gerėja valdymo kokybė;
- Susiformuoja patikima bazė sprendimams priimti ir planuoti;
- Tobulėja kontrolė;
- Efektyviai paskirstomi ir naudojami ištekliai, skirti rizikai mažinti;
- Gerėja saugos ir sveikatos reikalavimų vykdymo rezultatyvumas, taip pat aplinkos apsauga;
- Gerėja nuostolių prevencija ir incidentų valdymas;
- Sumažinami nuostoliai;
- Gerėja organizacijos gebėjimas mokytis;
- Gerėja organizacijos reakcija į pokyčius.

Patį rizikos valdymo procesą galima išskirti į 5 pagrindines dalis (2 pav.):

1. Nepageidaujamo įvykio identifikavimas;
2. Nepageidaujamo įvykio priežasties nustatymas;
3. Nepageidaujamo įvykio tikimybės apskaičiavimas;
4. Nepageidaujamo įvykio vertės nustatymas;
5. Sprendimo dėl rizikingos situacijos priėmimas.



2 pav. **Rizikos valdymo schema** (sudaryta autoriaus)

1.2. Finansinės rizikos rūšys

Finansinės rizikos valdymas yra procesas, kurio metu siekiama susitvarkyti su neapibrėžtumais, kylančiais iš finansų rinkų (Horcher 2005). Šio proceso metu įvertinamos potencialios organizacijai kylančios finansinės rizikos ir sukuriami strategija tų rizikų valdymui. Nuolatinis finansinių rizikų stebėjimas ir valdymas gali užtikrinti įmonės pranašumą rinkoje. Pagrindinės įmonėms kylančios finansinės rizikos yra rinkos rizika, užsienio valiutų rizika, palūkanų normų rizika, kredito rizika, operacinė rizika, likvidumo rizika, sisteminė rizika (3 pav.).



3 pav. Finansinės rizikos rūšys (sudaryta autoriaus)

Rinkos rizika

EBA rinkos riziką apibrėžia kaip balansinių ir nebalansinių nuostolių, atsirandančių dėl nepalankių rinkos kainų pokyčių, riziką. Rinkos rizika gali būti dviejų rūšių:

- absoliučioji;
- santykinė.

Absoliučioji rizika yra matuojama potencialiais nuostoliais, kurie yra išreikšti pinigais ir apskaičiuojama pagal bendrojo pelno svyravimus. Santykinė rizika yra matuojama pagal kontrolinį indeksą ir matuojama pagal nuokrypį nuo šio indekso. Dar vienas rinkos rizikai įtaką darantis kriterijus yra akcijų rinkos nepastovumas. Akcijų rinkos nepastovumas yra ypatingai aktualus investicinėms kompanijoms. Pasak Mirkovic V., Dasic B., Siljkovic B. (2013) pagrindiniai rodikliai, kurie daro didžiausią įtaką rinkos rizikos lygiui yra nuosavo kapitalo kainos, palūkanų normos, užsienio valiutų kursai ir biržos prekių kainos rizika.

Nuosavo kapitalo (arba vertybinių popierių) rizika, fundamentaliai, yra nuosavybės turėjimas investuojamoje srityje, pavyzdžiui įsigyjant įmonės akcijų paketą. Dažniausiai ši rizika yra susijusi su akcijų pirkimu, nes investuotojai prieš pirkdami akcijas stengiasi įvertinti nuosavo kapitalo riziką, kad sumažintų potencialius nuostolius jų akcijų portfeliuose. Statman M. (1987) teigia, kad

diversifikuotame akcijų portfelyje turėtų būti ne mažiau 30 skirtingų akcijų. Kitas nuosavo kapitalo rizikos mažinimo būdas yra ekonominių sričių, į kurias investuojama, diversifikavimas. Investuojant į visiškai skirtingus ekonominius sektorius, tokius kaip mažmeninės prekybos, technologijų, žemės ūkio galima sumažinti riziką.

Palūkanų normų rizika yra rizika, kad palūkanų normų pokyčiai gali neigiamai paveikti instrumento pozicijos ar portfelio vertę arba finansinės institucijos padėtį ir pelną apskritai. Tai rizika, kylanti iš obligacijų pelningumų lygio ir nepastovumo pokyčių, o taip pat iš obligacijų pelningumo skirtumų pokyčių. Ši rizika veikia visus finansinius instrumentus, tačiau labiausiai – skolos (fiksuojamų) instrumentus ir išvestinius instrumentus, sukurtus remiantis skolos instrumentais. Paprastai palūkanų normų pokyčiams jautresnė yra ilgesnio termino instrumentų vertė.

Užsienio valiutos kurso rizika, rinkos rizikos kontekste, yra siejama su prekių pirkimu užsienio valiuta. Jeigu sandoriai yra sudaromi su mokėjimo atidėjimu (atsiskaitoma ne iš karto, o tik po tam tikro laikotarpio), numatomos išlaidos ar įplaukos nacionaline valiuta tampa ne tikslios, o prognozuojamos. Nevaldoma ši rizika gali pridaryti įmonėms rimtų finansinių nuostolių, todėl ją būtina mažinti.

Biržos prekių kainos rizika tai rizika, kai dėl rinkos sąlygų pasikeitimo pasikeis turimų žaliavinių prekių vertė. Lietuvos bankas biržos prekių kainos riziką apibrėžia taip: „rizika, kad bankas, besiverčiantis prekyba biržos prekių rinkoje, gali patirti nuostolių dėl šių prekių atvirų pozicijų kainų pokyčių.“

Užsienio valiutų rizika

Šiuo metu vykstantis spartus verslo aplinkos ir ūkio subjektų veiklos kitimas yra stipriai susijęs su intensyviais pastaraisiais dešimtmečiais pasaulyje vykstančiais socialiniais, technologiniais ir ekonominiais procesais. Užsienio valiutos rizikos vertinimo svarba ypač išryškėja dėl vykstančios rinkos globalizacijos, o dėl nuolatinio valiutų kurso kitimo, svarbu įvertinti valiutos keitimo riziką, kuri yra susijusi su įmonės skolos tvarkymo išlaidomis.

Užsienio valiutų rizika, tai rizika, kad valiutos kurso pokyčiai gali neigiamai paveikti turimų užsienio valiuta denominuotų pozicijų vertę. Ši rizika pasireiškia laisvai svyruojant valiutai ir kai anksčiau buvusi fiksuoto kurso valiuta devaluojama. Lietuvos bankas užsienio valiutos kurso riziką apibrėžia taip: „tikimybė, kad bankas, turintis grynąją atvirąją užsienio valiutos (taip pat ir tauriųjų metalų) poziciją, susidariusią dėl prekybinių užsienio valiutos operacijų ir (ar) dėl jo turto ir įsipareigojimų struktūros, patirs nuostolių dėl tam tikros užsienio valiutos kurso arba jo svyravimo.“

Palūkanų normų rizika

Gali būti įvairių aplinkybių, kuomet kompanijos susiduria su palūkanų normos rizika, tačiau pagrindinė aplinkybė yra turto ir išpareigojimų terminų (instrumentai su fiksuota palūkanų norma) ir kainų pokyčiai (su kintama palūkanų norma). Šią riziką taip pat galima apibūdinti kaip grynujų palūkanų pajamų arba grynosios palūkanų normos (maržos) ir banko akcinio kapitalo rinkos vertės svyravimus dėl palūkanų normų pokyčių (Jasienė, 2002). Lietuvos bankas palūkanų normų riziką apibrėžia kaip riziką, kad bankas patirs nuostolių dėl palūkanų normos svyravimo. Palūkanų kitimo rizika, kartais dar vadinama kainų pokyčio rizika. Tai rizika, kad išlaidos, susijusios su skolos, už kurią mokamos palūkanos pagal kintamąją palūkanų normą, tvarkymu, išaugs dėl to, kad didės bazinė palūkanų norma, arba rizika, kad išlaidos, susijusios su skolos, už kurią mokamos palūkanos pagal fiksuotąją palūkanų normą, tvarkymu, taps nepagrįstai didelės lyginant su kintamosios palūkanų normos pokyčiais, susidariusiais dėl bazinės palūkanų normos mažėjimo.

Palūkanų normų rizikos laipsnis yra gyvybiškai svarbus komerciniams bankams, kurie turi žinoti savo palūkanų normų rizikos lygį ir laipsnį kad galėtų jį tinkamai valdyti. Nesugebėjimas tinkamai valdyti šios rizikos gali privesti prie konkurencingumo praradimo ar net bankroto.

Operacinė rizika

Operacinę riziką galima apibrėžti kaip tikimybę patirti nuostolių dėl žmonių, sistemų, netinkamų ar nepavykusių vidaus procesų arba dėl išorės įvykių poveikio, įskaitant teisinę riziką. Operacinė rizika yra plačiausia iš visų rizikos rūšių. Petria, Nicolae ir Petria, Licuța (2009) išskiria tokias pagrindines operacinės rizikos šakas:

- Atitikties rizika – sukianti nuostolių riziką dėl įstatymų, reguliavimų ir vidinių tvarkų nesilaikymo;
- Procesų neatlikimo rizika – sukianti nuostolių riziką dėl klaidų, neapdairumo, nelaimingų atsitikimų ar sukčiavimo atvejų tam tikrų procesų išpildyme;
- Sistemų gedimo rizika – sukianti nuostolių riziką dėl kompiuterinių ar komunikacinių sistemų gedimo;
- Materialiojo turto sugadinimo rizika – skelianti nuostolių riziką dėl materialiojo turto sunaikinimo ar sugadinimo dėl nelaimių ar nelaimingų atsitikimų;
- Žmogiškųjų išteklių rizika – sukianti nuostolių riziką dėl galimybės prarasti būtinus darbuotojus įmonės veiklai vykdyti arba nesugebėjimas išlaikyti pakankamo darbuotojų moralės lygio;
- Reguliavimo rizika – sukianti nuostolių riziką dėl verslo aplinkos reguliavimo pokyčių, įskaitant mokesčių ir apskaitos pokyčius;

- Nusikalstamumo rizika – sukianti nuostolių riziką dėl galimybės susidurti su nusikalstama veikla, tokia kaip vagystės, sukčiavimai, duomenų nutekėjimu, sistemų įsilaužimu ar pinigų plovimu;
- Stichinių nelaimių rizika – sukianti nuostolių riziką dėl stichinių nelaimių, tokių kaip potvyniai, miškų gaisrai, žemės drebėjimai ir pan.;
- Informacinių technologijų rizika – sukianti nuostolių riziką dėl IT sistemų gedimo;
- Klaidingų ataskaitų rizika – sukianti nuostolių riziką dėl klaidų atsiradusių ataskaitų formavimo procese;
- Apskaitos rizika – sukianti nuostolių riziką dėl klaidų padarytų apskaitos procese formuojant finansines ataskaitas;
- Nesusipratimų rizika – sukianti nuostolių riziką dėl galimo nesusišnekėjimo tarp kolegų produktų kūrimo procese, tokios rizikos pasekmė gali būti pagamintas produktas, kuris visiškai neatitinka pirkėjų lūkesčių;
- Modelio rizika – sukianti nuostolių riziką dėl netinkamo sprendimo priėmimo pasinaudojant neteisingai pasirinktais modeliais;
- Teisinė rizika – sukianti nuostolių riziką dėl netinkamo dokumentų paruošimo arba nesilaikymo teisės aktų reikalavimams.

Efektyvus šios rizikos valdymas gali užkirsti kelią krizėms, nesėkmėms, apsaugoti įmonės reputaciją, padidinti įmonės našumą (Olteanu 2012).

Likvidumo rizika

Likvidumo rizika gali būti dviejų tipų:

- rinkos/produkto likvidumo;
- pinigų srautų /finansavimo.

Rinkos/produkto likvidumo rizika atsiranda tada, kai rinkoje yra mažas aktyvumas ir sandorio neįmanoma įvykdyti esamomis rinkos kainomis. Šia riziką galima valdyti uždedant limitus rinkoms arba produktams ir diversifikuojant. Pinigų srautų/finansavimo rizika atsiranda tada, kai negalima vykdyti einamųjų įsipareigojimų. Ši rizika valdoma atsakingai planuojant pinigų srautų poreikius, nustatant limitus ir taikant diversifikaciją.

Likvidumo rizika bankiniame sektoriuje yra ypač svarbi, nes bankai turi ypač tiksliai įvertinti šią riziką siekdami maksimalaus pelningumo (Bareikaitė, Martinkutė-Kaulienė 2014). Jeigu likvidumo rizikos politika banke bus pernelyg griežta, bankas negalės investuoti savo lėšų ir pelningesnę likvidų turtą ir taip praras pajamas ir pelną. Jeigu bankas nepakankamai griežtai vertina

šią riziką, klientai gali norėti išsiimti iš banko daugiau pinigų negu bankas tikėjosi ir dėl to gali tekti pigiai parduoti dalį banko kapitalo likvidžiam turtui padidinti. Dėl šių priežasčių bankų vadovybė likvidumo rizikai skiria nemažą dalį dėmesio, kad užtikrintų pakankamą likvidžių pajamų dalį ateityje patenkinti klientų ir tiekėjų poreikius už bankui tinkamą kainą.

Sisteminė rizika

Sisteminė rizika – tai visos rinkos rizika, jos negalima panaikinti diversifikuojant, t.y. išskaidant, savo investicijas. Jeigu investuojama konkrečioje šalyje, tarkime, Lietuvoje, prisiimama šios šalies rinkos rizika.

Kredito rizika

Kredito rizika arba įsipareigojimų nevykdymo rizika, kyla kai viena iš šalių negali tinkamai vykdyti savo sutartinių įsipareigojimų, t.y. nesugeba laiku mokėti paskolos įmokų. Vertinant skolininko kredito riziką, atsižvelgiama į šiuos kriterijus:

- kredito gavėjo finansinė būklė, analizuojant jo pajamų kokybę, ankstesnius finansinius rezultatus (kredito istoriją), finansinį lankstumą;
- santykinė paskolos gavėjo finansinę padėtį rinkoje;
- demografinius duomenis.

1.3. Kredito rizika ir jos valdymas bankuose

Verslas neatsiejamas nuo rizikos, o viena iš pagrindinių ir labiausiai verslą veikiančių rizikos rūšių yra kredito rizika. Kreditų išdavimas išlieka svarbiausias visų pasaulio banko pajamų generavimo įrankis. Bazelio bankų rizikos komitetas (2001) apibūdina kredito riziką, kaip tikimybę visiškai ar dalinai prarasti skolinamą sumą dėl kliento nemokumo. Taigi, kuo labiau bankas pažeidžiamas kredito rizikai, tuo didesnė tikimybė, kad jam kils finansinių sunkumų.

Kredito rizika atsiranda visais atvejais, kai verslo proceso metu viena iš šalių pristato prekes, suteikia paslaugas ar atlieka darbus prieš gaudama už tai atlyginimą iš kitos sandorio šalies, t. y. atlieka šalių susitarimu numatytą savo įsipareigojimų dalį mainais už kitos sandorio šalies įsipareigojimą atlikti savąją prievolės dalį ateityje (Stasytytė, Aleksienė 2015). Kredito rizika – tai rizika, kad pasibaigus sutartam atsiskaitymų už pirktas prekes (ar suteiktas paslaugas) laikotarpiui, debitorius (pirkėjas, skolininkas) gali būti finansiškai nepajėgus atsiskaityti su kreditoriumi (pardavėju) arba tiesiog gali vengti sutartyje numatyto atsiskaitymo.

Pagrindinė finansinės rizikos valdymo esmė – siekimas minimizuoti rizikos kaštus ir maksimizuoti galimą pelną. Kadangi Lietuvoje veikiančių bankų pagrindinę turto dalį sudaro

paskolos (1 lentelė), tai galima daryti išvada, kad pagrindinė rizika, su kuria susiduria bankai yra kredito rizika ir šios rizikos valdymui bankai turi skirti daugiausiai dėmesio.

1 lentelė. **Pagrindiniai bankų veiklos rodikliai** (sudaryta autoriaus remiantis Lietuvos banko pateiktais bankų veiklos rodikliais)

Balansinės ataskaitos straipsnis	Bankų sektorius		Viso turto dalis 2017 3Q	Viso turto dalis 2018 3Q
	2017 3Q	2018 3Q		
Skolos vertybiniai popieriai	1 301 606	1 420 216	4,91%	5,19%
Nuosavybės vertybiniai popieriai	337 513	328 506	1,27%	1,20%
Išvestinės finansinės priemonės	90 248	79 002	0,34%	0,29%
Grynieji pinigai	426 784	434 682	1,61%	1,59%
Lėšos centriniuose bankuose	2 908 249	4 008 610	10,98%	14,66%
Lėšos kredito įstaigose	2 540 958	1 076 347	9,59%	3,94%
Paskolos ir kiti išankstiniai mokėjimai (su lizingu)	18 661 275	19 773 389	70,44%	72,31%
Kitos turto pozicijos	224 629	223 267	0,85%	0,82%
Visas turtas	26 491 262	27 344 019		

Pagrindinė kredito rizikos vertinimo dalis yra klientų patikimumo įvertinimas. Prieš priimant kiekvieną reikšmingą sprendimą, paprastai nustatomas sandorio šalies reitingas. Rizikos reitingas nustatomas naudojant rizikos vertinimo modelius, sistemas, kurie parenkami pagal vertinamo sandorio dydį ir sudėtingumą. Sandorio šalies įvertinimas yra svarbus santykių su klientais valdymo srityje – kuo didesnė klientų rizika, tuo didesnis dėmesys kreipiamas į kliento mokumą.

Norint nustatyti, kuriuos klientus yra naudinga finansuoti, o kuriuos nuostolinga, klientai suklasifikuojami pagal rizikos laipsnį į tam tikras grupes. D. Ivaškevičius ir A. Sakalas (1997) pateikia penkių kategorijų bankų kreditus, priklausomai nuo kreditinės rizikos procento:

- 1) Įprastas kreditas – tai kreditas, kuris bus padengtas pagal grafiką ir kurio rizika yra leistino dydžio.
- 2) Ypatumų turintis kreditas – tai kreditas, kuris kelia tam tikrų abejonių.
- 3) Kreditas, kurio rizika didesnė nei įprastinė, įsipareigojimų neįvykdymo rizika yra didesnė nei 20%.
- 4) Abejotinas, kai nuostolių tikimybė apie 50%.
- 5) Aiškiai nuostolingas, kai yra mažai tikėtina, kad kreditas bus padengtas.

1.4. Metodų apžvalga

Kredito rizikos vertinimas

Kreditingumo vertinimas yra statistinio modeliavimo procedūra, kuri leidžia vertintojui nuspėti kliento ateities veiksmus naudojant praeities veiksmų analizę. Pirmas kreditingumo vertinimo žingsnis yra apibrėžti kas yra „blogas“ ir „geras“ klientai. Šis kriterijus yra nustatomas atsižvelgiant į turimus duomenis, dažniausiai praktikoje „blogas“ klientas yra suprantamas tas klientas, kuris vėluoja atlikti įmoką 60 dienų ir daugiau. Toliau yra sutvarkomi ir modeliui pritaikomi prediktoriai – kintamieji, kurie bus pagrindinis modelio variklis. Šie kintamieji gali būti ir labai paprastai suprantami, tokie kaip amžius ar lytis, tačiau gali būti ir kompleksiniai, kelių kintamųjų junginys. Galutinis kreditingumo vertinimo modelių rezultatas įprastai yra reitingas, kuris parodo kliento kreditingumo lygį (Hand & Henley, 1997).

Tiesinė diskriminantinė analizė

Tiesinę diskriminantinę analizę (LDA) sukūrė Fisher (1936), kuris pasiūlė, kad geriausias būdas atskirti dvi grupes (gerus klientus ir blogus klientus) yra surasti tiesinę nepriklausomųjų kintamųjų kombinaciją, kuri atspindėtų maksimalų atstumą tarp šių grupių vidurkių (1 pav.). Tiesinės diskriminantinės analizės modelis yra:

$$S(x) = w_1x_1 + w_2x_2 + \dots + w_nx_n,$$

čia: $S(x)$ – skolininko kredito rizikos įvertinimas (balų skaičius); x_n – veiksniai, turintys įtakos skolininko kredito rizikai; w_n – veiksmų svoriai.

Taikant diskriminantinę analizę vienas pagrindinių metodo trūkumų yra tai, kad modelio rezultatas (prognozė), gali kisti plačiame spektre, vadinasi rezultate gaunamas tik santykinis rodiklis, parodantis kliento kredito rizikos vertinimą, kito kliento atžvilgiu. Taip pat modelyje naudojami kintamieji yra susieti tik tiesiniais ryšiais, kurie realybėje pasitaiko gerai rečiau, nei netiesiniai.

Pagrindiniai LDA modelio privalumai yra jo paprastumas ir tai, kad jis sąlyginai lengvai apskaičiuojamas. Šis metodas yra pagrįstas prielaida, kad duomenys yra pasiskirstę pagal normalųjį skirstinį. Tačiau, Reichert ir kt. (1983) teigė, kad duomenys nebūtinai turi būti pasiskirstę pagal normalųjį skirstinį, kad modelis duotų gerus rezultatus. Wiginton (1980) teigė, kad nepasiskirstymo pagal normalųjį skirstinį problema galima įveikti naudojant logistinės regresijos modelį, kuris aprašytas šio dokumento 2.2. skirsnyje.

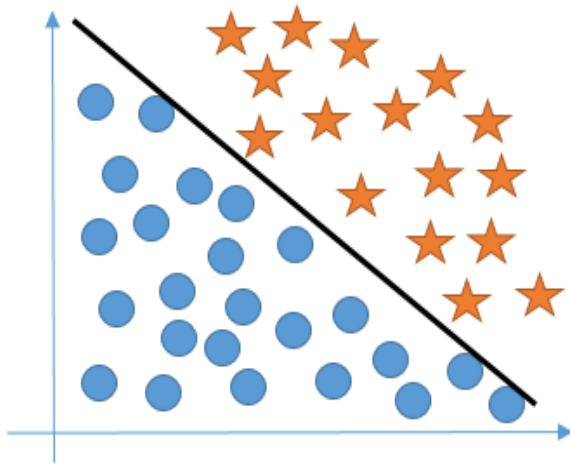
Taigi, išnagrinėjus mokslinius straipsnius, nagrinėjančius tiesinės diskriminantinės analizės modelius (LDA), galima teigti, kad metodo pranašumai yra šie:

- Efektyvumas kreditingumo vertinimo procesuose;

- Lengvas modelio pritaikymas ir interpretavimas;
- Efektyvumas turint didelį duomenų kiekį.

Tačiau šis metodas turi ir trūkumų:

- Taiko nemažai statistinių prielaidų;
- Reikia tvarkingai surūšiuotų kategorinių kintamųjų;
- Jautrumas išskirtims. (Doumpos & Zopounidis, 2014a).



4 pav. Duomenų grupės (sukurta autoriaus pagal Fisher (1936))

Logistinė regresija

Logistinė regresija (LR) yra daugiamatė analizės technika, kuria siekiama paaiškinti ryšį tarp atsitiktinio dvejetainio priklausomo kintamojo ir vieno ar kelių nepriklausomų kintamųjų (Hosmer, Lemeshow, 2000). Lessmann, Baesens, Seow ir Thomas (2015) atliko išsamų klasifikacijos metodų, naudojamų kredito rizikos vertinimo modeliams kurti, tyrimą ir nurodė logistinę regresiją kaip finansų sektoriaus standartą.

Logistinės regresijos tikslas - priskirti priklausomą kintamąjį dvi grupės (0; 1). Viena grupė yra „geri“ (mokūs), o kita „blogi“ (nemokūs) klientai (Genriha, Voronova 2012). Kadangi logistinis modelis įgyja reikšmes nuo 0 iki 1, jis leidžia prognozuoti tikimybes. Tikimybės apskaičiuojamos naudojant šią lygtį:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta x + \varepsilon \quad (1)$$

Šioje lygtyje p yra prognozuojamo įvykio tikimybė (nagrinėjamu atveju prognozuojamas įvykis yra „klientas yra nemokus“, toliau – PD (angl. *Probability of Default*)). Pertvarkius (1) lygtį, PD galime išreikšti taip:

$$PD = \frac{1}{1 + e^{-(\alpha + \beta x)}} \quad (2)$$

Apie logistinės regresijos tinkamumą kreditingumo vertinimo procese yra parašyta nemažai mokslinių straipsnių. Dauguma pritaria, kad logistinė regresija daugeliu atveju yra aiškiausias ir lengviausiai paaiškinamas būdas gana tiksliai įvertinti kliento mokumą (Zenzerović 2011), tačiau jei analizuojamas „neapdorotų“ duomenų rinkinys, kuriame duomenys yra su klaidingomis ar trūkstamomis reikšmėmis ar turinčius skaitinę ir kategorinę vertę, logistinė regresija negalės savarankiškai tokių duomenų apdoroti. Tam, kad logistinė regresija analizuotų tokius duomenis, analitikas turi iš anksto juos apdoroti. Toks duomenų sutvarkymas užima daug laiko ir išteklių, o rezultate gali likti klaidų dėl žmogiškosios klaidos (angl. *Human error*) faktoriaus. Pasak Priya Ranganathan, C. S. Pramesh ir Rakesh Aggarwal naudojant logistinę regresiją svarbiausia nepamiršti šių punktų:

- **Pasirinkti tinkamus priklausomus kintamuosius.** Vienas svarbiausių tinkamo logistinio modelio kriterijų yra tikslus priklausomų kintamųjų, kurie bus įtraukti į modelį, parinkimas. Natūraliai, į modelį norisi įtraukti kiek įmanoma daugiau kintamųjų, tačiau per didelis kintamųjų skaičius gali sukelti klasifikatorių persimokymo (angl. *overfitting*) efektą.
- **Vengti stipriai koreliuojančių kintamųjų.** Jei modelio kintamieji turi stiprią koreliaciją vienas su kitu (multikolineariškumas), tuomet kiekvieno kintamojo poveikis regresijos modeliui tampa ne toks tikslus ir modelis, vietoj to, kad vieną kriterijų (kintamąjį) padarytų labai reikšmingu, padalins reikšmingumo lygį visiems koreliuojantiems kintamiesiems.
- **Tinkamai paruošti daugiareikšmius modelio kintamuosius.** Naudojant logistinę regresiją, daugiareikšmius duomenis, tokius kaip amžius, mėnesinis atlyginimas ir pan. reikia suskirstyti į reikšmingus intervalus. Suskaidžius į per daug intervalų, jie gali tapti nereikšmingi, o naudojant per mažai intervalų galima prarasti dalį informacijos.
- **Atsižvelgti į įvesties (angl. *Input*) ir išvesties (angl. *Output*) kintamųjų ryšius.** Regresijos modeliai daro prielaidą, kad ryšys tarp prediktorių ir priklausomo kintamojo yra vienakryptis, t.y. seka viena kryptimi – teigiama arba neigiama. Pavyzdžiui, jeigu jauni klientai (iki 25 metų) yra tokie pat rizikingi kaip ir pensijinio amžiaus klientai, o tarpinio amžiaus klientai mažiau rizikingi, toks kintamasis modelyje neduos teisingų rezultatų.

Pagrindinis šio metodo privalumas yra tas, kad jis gali sukurti paprastą tikimybinę klasifikavimo formulę. Trūkumai yra tai, kad LR negali tinkamai susidoroti netiesiniais aiškinamųjų kintamųjų (prediktorių) ryšiais.

Neuroninių tinklų modelis

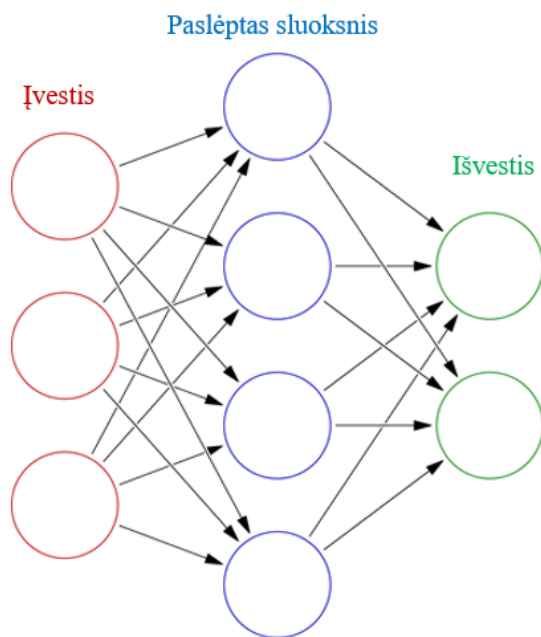
Kaip alternatyva LR modeliui gali būti neuroninių tinklų modelis (ANN – artificial neural network). Šis metodas laikomas labai daug žadančiu kreditingumo vertinime, tačiau yra stipriai kritikuojamas dėl ilgo modelio mokymosi proceso siekiant užtikrinti optimalų neuronų tinklą, taip pat šis metodas sunkiai paaiškina prediktorių svarbą modelyje, todėl jį pakankamai sunku interpretuoti. Dėl šių priežasčių praktinis modelio panaudojimas yra limituotas (Piramuthu, 1999). Nemažai mokslininkų, tyrusių šį metodą, nustatė, kad ANN yra tikslesnis nei LDA ar LR ar kiti tradiciniai kreditingumo vertinimo metodai (Jensen, 1992; Coats & Fant, 1993; Desai et al., 1996; Huang et al., 2004; Chuang & Lin, 2009).

ANN išsivystė iš neurobiologinių modelių ir jį sudaro stipriai tarpusavyje susiję ir sąveikaujantys procesiniai vienetai. Neuronų tinklai apdoroja informaciją per didžiulį kiekį tarpusavyje sąveikaujančių procesinių vienetų, vadinamų neuronais ir jų junginiais su išoriniais kintamaisiais. Pagrindinė ir pati svarbiausia neuroninių tinklų savybė – jų gebėjimas mokytis. Visai kaip žmogaus smegenys, neuroninių tinklų modelis gali mokytis iš praeities pavyzdžių ir lanksčiai keisti savo parametrus, kad prisitaikytų prie pateikto duomenų rinkinio.

Neuroninių tinklų modeliai gali būti skaidomi į dvi grupes:

- Feedforward tinklas;
- Feedback tinklas.

Pasak Vellido et al. (1999), beveik 80% visų verslų naudodami ANN naudoja feedforward tinklus. Paprastą feedforward tinklą sudaro trys sluoksniai: įvesties sluoksnis, paslėptas sluoksnis ir išvesties sluoksnis (5 pav). Įvesties sluoksnis apdoroja įvesties kintamuosius ir pateikia apdorotas reikšmes paslėptam sluoksniui. Tada paslėptas sluoksnis apdoroja tarpines vertes ir perduoda apdorotas reikšmes išvesties sluoksniui. Išvesties sluoksnis pateikia rezultatus. Siekiant veiksmingai taikyti ANN, buvo parengtos kelios gairės, kaip nustatyti mokymosi parametrus, įskaitant mokymosi greitį, tinklo sluoksnių skaičių ir pan. (Freeman & Skapura, 1992). Mokymosi tempas yra labai svarbus, nes mažesnis tempas sukels lėtą konvergavimą, o aukštesnis gali sukelti tinklo virpesius ir konvergavimą apsunkinti.



5 pav. Neuroninių tinklų modelio schema

Apibendrintas tiesinis modelis

Apibendrintas tiesinis modelis (GLM) gali būti išreikštas tokia formule:

$$\hat{Y}_i = \beta_0 + \beta_1 X$$

Čia \hat{Y} – priklausomas kintamasis, β_0 – regresijos kreivės atkarpa \hat{Y} ašyje, β_1 – koeficientas arba svoris, nulemiantis nepriklausomo kintamojo X svorį modelyje, X – nepriklausomas kintamasis.

Ši lygtis yra analogiška tiesinės regresijos modeliui, vienintelis skirtumas yra tas, kad GLM modelis yra pritaikytas naudoti ne tik skaitinius, bet ir kategorinius kintamuosius. Tiksliau, GLM sudaro ne vienas modelis, o modelių šeima (2 lentelė). GLM modelio generavimo metu, kategoriniai kintamieji yra pakečiami skaitiniais ir tada atliekami veiksmai.

2 lentelė. GLM modelio šeimos modeliai.

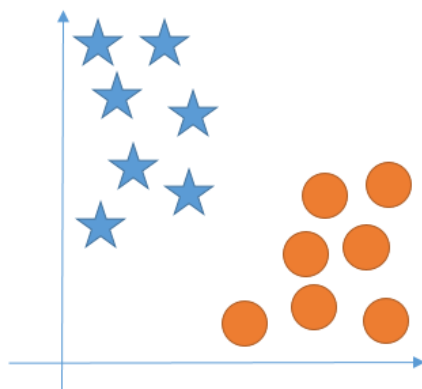
Modelis	Pasiskirstymas	Kintamųjų tipai
Tiesinės regresijos modelis	Normalusis	Skaitiniai
ANOVA	Normalusis	Kategoriniai
ANCOVA	Normalusis	Skaitiniai ir kategoriniai
Logistinės regresijos modelis	Binominis	Skaitiniai ir kategoriniai
Logtiesinis modelis	Puasono	Kategoriniai
Puasono regresijos modelis	Puasono	Skaitiniai ir kategoriniai
Multinominis modelis	Multinominis	Skaitiniai ir kategoriniai

Pasak Agresti (2007), Agresti (2013), ir McCullagh & Nelder (1989). Modelyje naudojamos prielaidos:

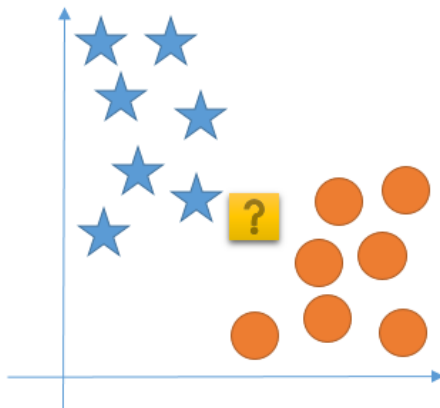
- Kintamieji Y_1, Y_2, \dots, Y_n yra nepriklausomai pasiskirstę, t.y. nepriklausomi.
- Kintamasis \hat{Y}_i neprivalo būti pasiskirstęs pagal normalųjį skirstinį, tačiau jo modelyje daroma prielaida, kad jis yra pasiskirstęs pagal eksponentinių skirstinių šeimos skirstinį (pvz. binominį, Puasono, multinominį, normalųjį ir t.t.).
- Modelis nedaro prielaidos, kad tarp priklausomo ir nepriklausomų kintamųjų yra tiesinis ryšys.
- Nepriklausomi kintamieji gali būti ir pradinių nepriklausomų kintamųjų transformacijos.
- Modelio paklaidos privalo būti nepriklausomos, tačiau jos neprivalo būti pasiskirsčiusios pagal normalųjį skirstinį.
- Priešingai nei įprasti regresijos modeliai, GLM naudoja ne mažiausių kvadratų metodą (angl. Ordinary least squares), o didžiausios tikimybės metodą (angl. Maximum likelihood estimation) modelio parametrams nustatyti.
- Modelio tinkamumo nustatymo kriterijams reikalinga gana didelė imtis.

K-artimiausio kaimyno modelis

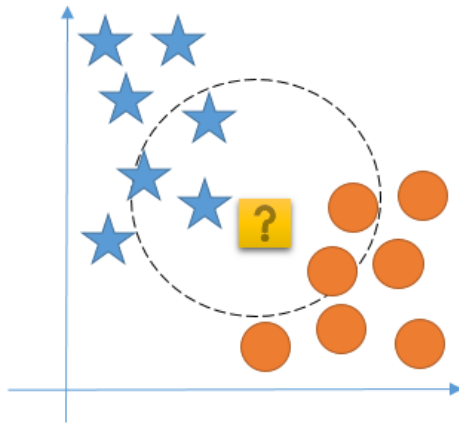
K-artimiausio kaimyno modelis yra neparаметrinis modelis, kuris mokosi iš panašumų. Kiekvienam naujam stebėjimui modelis iširia struktūrinį atstumą iki K artimiausių kaimynų (Karamizadeh et al 2013). Modelio principą galima lengviausiai paaiškinti grafiškai. Sakykime, kad turime dvi skirtingas kintamojo X grupes:



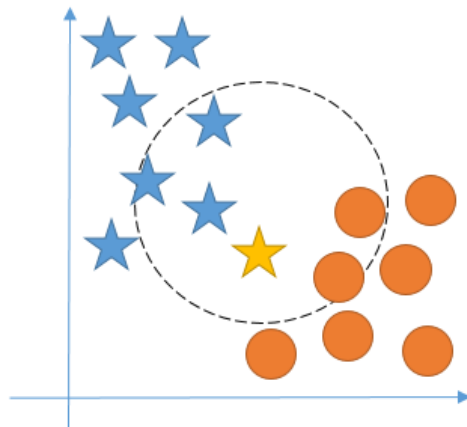
Ir turime vieną kintamojo X reikšmę, kuri dėl kažkokios klaidos liko neužpildyta:



K-artimiausio kaimyno modelis, naudodamas nustatytą kaimynų (kaimyninių reikšmių) skaičių k , nustatys, kuriai iš šių grupių priklauso nežinoma reikšmė. Sakykime, kad modeliui pateikta k reikšmė yra 5.



Palyginęs artimiausias 5 kaimynines reikšmes, modelis nuspręs, kad „?“ pažymėta reikšmė yra labiau panaši į žvaigždute pažymėtas, nei į apskritimu pažymėtas reikšmes, nes iš 5 artimiausių kaimynų 3 yra žvaigždutės.



Šiame modelyje nedaromos jokios prielaidos dėl naudojamų duomenų ir jis yra pakankamai lengvai intuityviai suprantamas jį taikantiems analitikams, tačiau modelio rezultatas nėra

tikimybinė išraiška, todėl negali būti tiksliai interpretuojamas. Taip pat modelis yra reiklus kompiuterio galiai, ypač atliekant veiksmus su didelėmis duomenų bazėmis.

Sprendimų medis (miškas)

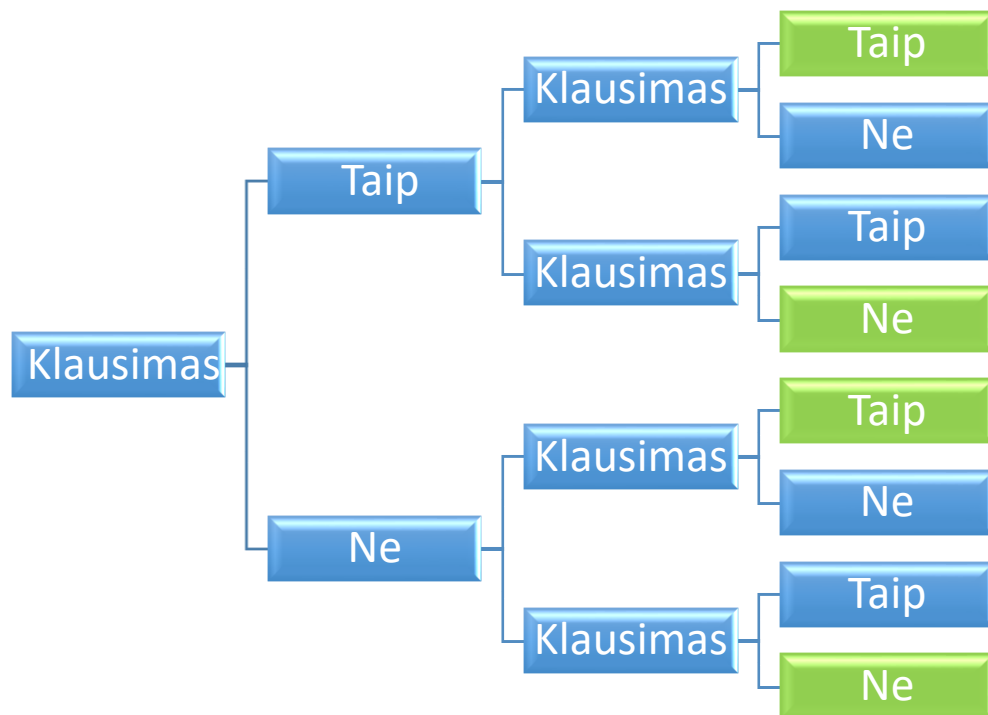
Sprendimų medžio modelis gali būti apibūdintas kaip medžio formos grafikas, kurio mazgai yra galimi įvykių rezultatai. Galutinis modelio sprendimas priimamas paskutiniame mazge, vadinamame lapiniu mazgu. Kiekviename mazge yra atliekamas testas įvesties kintamiesiems, kiekviena šaka veda mazgo rezultatą prie kito mazgo.

Modelis sudarytas iš daug tokių medžių vadinamas sprendimų mišku. Kiekvienas sprendimų miško medis mokosi iš atsitiktinių mokymų stebėjimo pavyzdžių. Imtis imama atsitiktinai, o tai reiškia, kad kai kurie imties elementai bus naudojami kelis kartus viename medyje. Idėja yra ta, kad mokant kiekvieną medį pagal skirtingus pavyzdžius, visas miškas turės mažesnę dispersiją, bet ne sumažėjusį tikslumą.

Pagrindinė atsitiktinio miško idėja yra sujungti daugelio sprendimų medžių prognozes į vieną modelį. Atskirai sprendimų medžių prognozės gali būti netikslios, tačiau sujungtos kartu, prognozės bus artimesnės tikrajai vidutinei vertei.

Pagrindiniai modelio privalumai yra:

- Įvertinus keletą medžių sumažėja permokymo tikimybė;
- Prognozuojant lengva įvertinti kiekvienos savybės santykinę svarbą;
- Jame nėra daug statistinių prielaidų, tokių kaip duomenų normalusis pasiskirstymas, tiesinės kintamųjų priklausomybės egzistavimas ar skaičiavimų iteracijų skaičius;
- Lengvas algoritmo panaudojimas;
- Gali būti naudojamas tiek klasifikavimo tiek regresijos tikslais;
- Loginiai sprendimų išvedimai;
- Lankstumas.



6 pav. Sprendimų medžio metodo vizualizacija

Atmetimo taisyklėmis pagrįsta mokumo vertinimo sistema

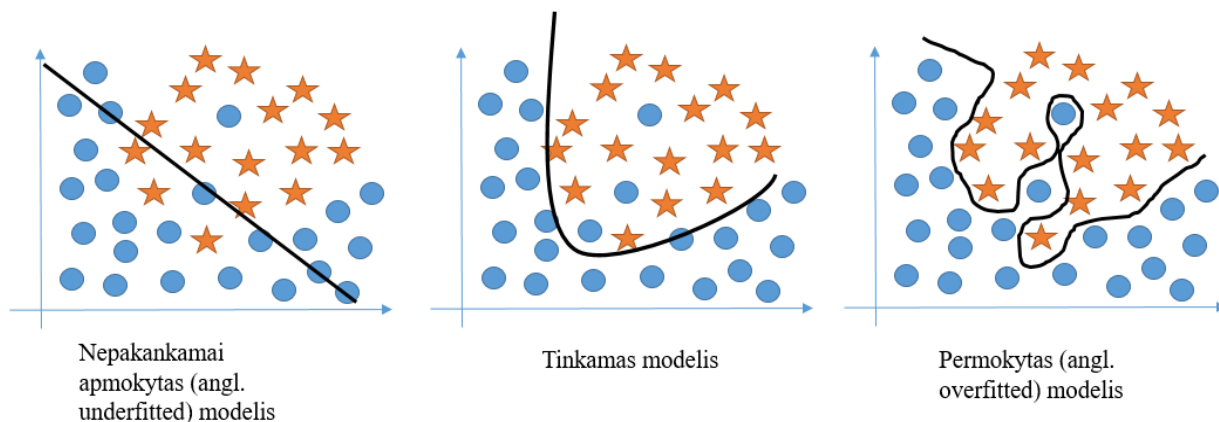
Atmetimo taisyklėmis pagrįsta mokumo vertinimo sistema yra skirta mokumo vertinimui, kai nėra turima pakankamai duomenų modelių sudarymui. Tokiu atveju tikrinamos klientų grupės pagal ekspertinio vertinimo kriterijus ir sudaromos vadinamos kietosios taisyklės (angl. Hard rules), kurias turi patenkinti mokumo vertinimą atliekantis klientas. Šio metodo įgyvendinimui dažniausiai naudojami tokie intuityvūs kintamieji kaip mokėjimų istorija, kredito istorija, pajamos. Nustatant, kuriuos veiksnius naudoti, dažniausiai naudojama vadovų patirtis, jų nuomonė, taip pat atsižvelgiama į teikiamų paslaugų ir produktų specifiką, rinkos patirtį. Modelis gali būti nuolatos pildomas atsiradus naujoms įžvalgoms. Tam, kad tokia sistema tinkamai funkcionuotų, reikalinga nuolatinė duomenų bazės pokyčių analizė.

Šis metodas yra lengvai pritaikomas, taisyklės nesunkiai pritaikomos sistemose ir sistema gali būti lengvai interpretuojama, tačiau palieka daug erdvės žmogiškajai klaidai, reikalauja daug žmogiškųjų išteklių nuolatinėms analizėms, yra imlus laikui. Taip pat metodas yra labai nelankstus, nes negeba kompleksiskai vertinti kliento mokumo atsižvelgiant į keletą savybių kombinacijų.

1.5. Metodų apibendrinimas

Viena dažniausiai mokslinėje literatūroje nagrinėjamų problemų, kai analizuojami matematiniai modeliai yra persimokymas (angl. overfit) ir nepakankamas mokymasis (angl. underfit)

(7 pav.). Šie veiksniai yra labai svarbūs modelio kūrimo procese ir gali padaryti didžiulę įtaką galutiniam rezultatui (prognozei).



7 pav. Nepakankamas apmokymas, subalansuotas apmokymas ir persimokymas

Dažna neapankamo apmokymo priežastis – pernelyg paprastas modelis, .t.y. modelyje visai nėra arba yra nepakankamai kintamųjų galinčių paaiškinti priklausomo kintamojo variaciją. Dar viena galima priežastis – netiesiniams ryšiams modeliuoti pasirenkamas tiesinis modelis (pvz. tiesinė regresija). Tuo tarpu persimokymas atsiranda tuomet, kai modelis yra per daug sudėtingas, .t.y. modelyje naudojama per daug nepriklausomų kintamųjų, kurie apibūdina priklausomą kintamąjį. Iš pirmo žvilgsnio, toks modelis atrodo labai tikslus ir puikiai atspindi apmokymo imties duomenis, tačiau jis išmoksta ne tik esminius kintamųjų ryšius, bet ir triukšmą (angl. noise). Norint išvengti persimokymo, tyrimo imtis yra paskirstoma į 2 dalis: apmokymo ir testavimo. Testavimo imtyje atliekami modelio patvirtinimo (angl. validation) darbai, siekiant įsitikinti kaip jis veikia naudojant duomenis, kuriuos „mato“ pirmą kartą. Visų modelių ir metodų apibendrinimas pateikiamas 2 lentelėje.

2 lentelė. Kreditingumo vertinimo modelių privalumai ir trūkumai

Modelis	Privalumai	Trūkumai
Tiesinė diskriminantinė analizė	Gali būti naudojamas turint nedidelį duomenų kiekį. Jeigu modelio prielaidos yra tenkinamos – modelis yra labai efektyvus.	Turi labai daug prielaidų, kurias praktikoje labai sunku patenkinti.

2 lentelės tęsinys. **Kreditingumo vertinimo modelių privalumai ir trūkumai**

Logistinė regresija	Pagrindinis modelio privalumas – nėra daromos prediktorių priklausomybės matematiniam skirstiniams prielaidos. Tai modeliui suteikia daugiau tikslumo ir padaro kintamųjų tvarkymo procesą gerokai paprastesnį.	Pagrindiniai modelio trūkumai yra tai, kad modelis neatsižvelgia į prediktorių tarpusavio koreliaciją ir naudoja tik tiesines kintamųjų priklausomybes.
Neuroninių tinklų modelis	Pagrindinis modelio privalumas yra tas, kad analitikui nereikia nurodyti modelio kintamųjų tarpusavio ryšių, o tiesiog įdėti kintamuosius į modelį.	Modelio sudėtingumas neleidžia modeliui būti lengvai interpretuojamam, modelio struktūra dažnai yra nematoma analitikui ir nėra tiksliai žinoma kodėl klientas gavo vienokį ar kitokį reitingą.
Apibendrintas tiesinis modelis	Priklausomo kintamojo nereikia transformuoti į normalųjį skirstinį. Modeliai naudoja maksimalios tikimybės metodą modelio parametrus nustatyti, taigi modelis bus optimalus. Lengvai pritaikomas.	Modelio kokybei labai daug įtakos turi naudojami kintamieji. Pridėjus netinkamai pasiskirsčiusį ar perteklinį kintamąjį, modelio tikslumas gali reikšmingai sumažėti.
K-artimiausio kaimyno modelis	Nedaromos jokios prielaidos dėl naudojamų duomenų. Intuityviai nesunkiai suprantami modelio rezultatai. Labai efektyvus turint didelį duomenų kiekį.	Nėra aiškaus tikimybinio rezultato (modelis negali būti tiksliai paaiškintas). Reikalauja labai galingos technikos apdorojant dideliu duomenų kiekius.
Sprendimų medis (miškas)	Lengvai suprantamas ir paaiškinamas (matoma kokie duomenys naudojami). Nesunkiai kombinuojamas su kitais vertinimo modeliais. Jame nėra naudojama daug statistinių prielaidų. Maža permokymo tikimybė.	Nestabilus – net ir maži pokyčiai kintamųjų struktūroje gali padaryti ženklų įtaką galutiniam modelio rezultatui. Reikalauja brangių resursų, nes pasiekti tikslų modelį reikia sugeneruoti pakankamai daug medžių, o turint daug duomenų tai gali reikalauti brangios įrangos. Modelis gana lėtas, daugelio medžių generacija užima daug laiko. Nepaisant to, kad žinoma kokie duomenys naudojami modelyje jį yra gana sunku interpretuoti, nes nėra iki galo aišku kodėl modelis priėmė vieną ar kitą sprendimą.

2 lentelės tęsinys. **Kreditingumo vertinimo modelių privalumai ir trūkumai**

<p>Atmetimo taisyklėmis pagrįsta mokumo vertinimo sistema</p>	<p>Lengvai interpretuojamas, paprastas praktinis pritaikymas, nereikia didelio duomenų kiekio modelio taikymui.</p>	<p>Neefektyvus, maža prognozuojamoji galia, nelankstus, didelė žmogiškosios klaidos tikimybė. Modeliui sukurti dažnai naudojama ne statistinė informacija, o vadovų patirtis, nuomonė.</p>
--	---	--

2. FIZINIŲ ASMENŲ KREDITINGUMO VERTINIMO TYRIMO METODOLOGIJA

Lietuvoje dažna praktika, vertinant įmonių kreditingumą, yra ekspertinis vertinimas. Tačiau ekspertinis vertinimas turi labai daug ženklų trūkumų. Visų pirma norint atlikti ekspertinį vertinimą, reikia turėti patyrusių žmonių komandą. Taip pat ekspertinis vertinimas yra imlus laiko atžvilgiu, nes reikia tikrinti kiekvieną klientą individualiai. Taigi toks kreditingumo vertinimo būdas reikalauja brangių žmogiškųjų resursų, laiko ir palieka „žmogiškosios klaidos“ tikimybę. Vienas iš šių problemų sprendimo būdų yra sukurti automatinį kreditingumo vertinimo algoritmą, kuris padarytų ekspertinį vertinimą gerokai efektyvesnį ir sumažintų klaidos tikimybę.

Kredito rizikos modeliai yra dažnai analizuojama tema, nuolat atnaujinami ir kuriami vis nauji modeliai, kurie yra plačiai naudojami bankų praktikoje. Pasaulyje vis garsiau kalbama apie mašininio mokymosi modelių svarbą ir begalinį potencialą. Huseyin Ince ir Bora Aktan (2009) atlikta banko klientų analizė parodė, kad geriausiai klientus klasifikuoja sprendimų medžio modelis, o neuroninių tinklų modelis turi geriausias savybes apskaičiuoti klientų kreditingumą. Sustersic, Maja ir Mramor, Dusan ir Zupan, Jure, Consume (2007) savo tyrime taip pat padarė išvadą, kad neuroninių tinklų modelis rodo geriausius rezultatus klientų kreditingumo vertinime. Tuo tarpu Gouvêa, Maria ir Bacconi, Eric. (2007) atliko Brazilijos finansinių institucijų klientų vertinimo modelių tyrimą ir padarė išvadą, kad tiksliausias yra logistinės regresijos modelis. Iš nagrinėtos literatūros galima daryti išvadą, kad paprastas logistinės regresijos modelis vis dar išlieka populiariausias pasirinkimas bankų fizinių asmenų rizikos vertinimo ekspertų tarpe. Pagrindinė to priežastis – modelio paprastumas ir lengva interpretacija.

Įvertinus tai, kad bankai dažniausiai renkasi lengviau paaiškinamą ir paprastesnį būdą klientų mokumui vertinti, iškeltas tyrimo klausimas:

„ar sunkiau paaiškinami mašininio mokymosi modeliai yra reikšmingai tikslesni už klasikinių logistinės regresijos modelių?“

Darbo tikslas. Naudojant statistinės analizės programą R-project nustatyti, kuris iš dažniausiai mokslinėje literatūroje analizuojamų mašininio mokymosi metodų yra tiksliausias siekiant nustatyti fizinių asmenų mokumą vienoje iš Lietuvoje finansines paslaugas teikiančių įmonių.

Išsikelti tyrimo uždaviniai:

- 1) Matematinių modelių pasirinkimas;
- 2) Modelio kūrimo imties nustatymas;

- 3) Nepriklausomų kintamųjų sudarymas ir pasirinkimas;
- 4) Modelių kūrimas;
- 5) Sukurtų modelių testavimas;
- 6) Rezultatų analizė.

2.1. Tyrimo imtis.

Kertinis kiekvieno statistinio modelio kūrimo elementas – imtis. Tik kokybiški ir gerai klientų profilį atspindintys duomenys gali būti naudojami modelių sudarymui. Tyrimo imtis - vienos Lietuvos finansinių paslaugų įmonės duomenys skirti fizinių asmenų kreditingumo vertinimui. Imtį sudaro 14 kintamųjų ir 3286 stebėjimai (klientų, sudariusių vartojimo kredito sandorius nuo 2017m., profiliai). Naudojamų duomenų aprašymas pateiktas 4 lentelėje, duomenų struktūra pateikta 8 pav. panaudojus Rstudio funkciją *str*, kintamųjų rodikliai (Rstudio funkcija *summary*) – 9 pav.

```
Classes 'tbl_df', 'tbl' and 'data.frame':      3286 obs. of  14 variables:
 $ HR_12                : Factor w/ 2 levels "BAD","GOOD": 1 1 1 1 1 1 1 1 1 1 ...
 $ SODRA_MONTHLY_INCOME : num  131 134 140 160 176 ...
 $ SAIS_FIN_SUM_2Y      : num   93 0 0 0 0 ...
 $ KB_MONTHLY_INST_PRDB : num   0 0 0 0 67 ...
 $ NUMBER_OF_CHILDREN   : num   0 0 0 0 0 0 0 1 0 0 ...
 $ GRT_MARITAL_STATUS   : Factor w/ 4 levels "DIVORCED","MARRIED",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ X_GRT_COUNTY_CODE     : Factor w/ 2 levels "a/ Registered",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ SODRA_PROF_STATUS    : Factor w/ 4 levels "PENSIONER","SALARIED",...: 1 1 1 2 2 1 1 1 1 1 ...
 $ SEX                  : Factor w/ 2 levels "F","M": 1 2 2 2 2 2 1 1 2 2 ...
 $ GF_NBDEAL_HISTO      : num   0 1 0 1 0 0 0 0 1 1 ...
 $ SODRA_SENIORITY_OLD_SAL_JOB: num  NA NA NA 71 3 NA NA NA NA ...
 $ LOAN_AMOUNT_FINANCED : num  400 480 50 2200 120 ...
 $ GF_MAX_DPD_HISTO     : num   0 0 0 76 0 0 0 0 0 30 ...
 $ DTI                  : num   0 0 0 0 0.38 ...
```

8 pav. Duomenų struktūra

HR_12	SODRA_MONTHLY_INCOME	SAIS_FIN_SUM_2Y	KB_MONTHLY_INST_PRDB	NUMBER_OF_CHILDREN	GRT_MARITAL_STATUS
BAD : 494	Min. : 106.5	Min. : 0.00	Min. : 0.00	Min. : 0.0000	DIVORCED: 104
GOOD:2792	1st Qu.: 445.5	1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.:0.0000	MARRIED : 66
	Median : 577.1	Median : 0.00	Median : 26.43	Median :0.0000	SINGLE :2586
	Mean : 639.8	Mean : 22.06	Mean : 65.86	Mean :0.2249	WIDOWED : 530
	3rd Qu.: 760.6	3rd Qu.: 0.00	3rd Qu.: 104.69	3rd Qu.:0.0000	
	Max. :5222.2	Max. :895.11	Max. :1192.13	Max. :5.0000	

	X_GRT_COUNTY_CODE	SODRA_PROF_STATUS	SEX	GF_NBDEAL_HISTO	SODRA_SENIORITY_OLD_SAL_JOB	LOAN_AMOUNT_FINANCED	
a/ Registered	:3259	PENSIONER	: 265	F:1467	Min. : 0.00	Min. : 0.00	Min. : 50.0
b/ Non Registered:	27	SALARIED	:2982	M:1819	1st Qu.: 0.000	1st Qu.: 8.00	1st Qu.: 300.0
		SELF_EMPLOYED	: 3		Median : 1.000	Median : 21.00	Median : 500.0
		SPECIAL_OCCUPATION:	36		Mean : 1.471	Mean : 42.67	Mean : 920.2
					3rd Qu.: 2.000	3rd Qu.: 52.00	3rd Qu.: 922.3
					Max. :41.000	Max. :302.00	Max. :15000.0
						NA's :275	

GF_MAX_DPD_HISTO	DTI
Min. : 0.00	Min. :0.00000
1st Qu.: 0.00	1st Qu.:0.00000
Median : 0.00	Median :0.04542
Mean : 10.48	Mean :0.09435
3rd Qu.: 2.00	3rd Qu.:0.17723
Max. :951.00	Max. :0.39364

9 pav. Kintamųjų rodikliai

9 pav. matoma visa duomenų bazės struktūra. Vidutinės klientų analizuojamų klientų pajamos – 639,8 eurų per mėnesį, vidutiniai klientų turimi finansiniai įsipareigojimai – 65,86 eurų per mėnesį. Analizuojami klientai turi visai gerą paskutinių 2 metų kredito istoriją – vidutiniškai turėjo vos 22,06 eurų skolas. Dauguma klientų neturi nepilnamečių vaikų, o daugiausiai vaikų turintis klientas turi 5 vaikus. Duomenų bazę sudaro 104 išsiskyrusių klientų, 66 susituokusių klientų, 2586 vienišų klientų ir 530 našlių. Registruotą gyvenamąją vietą turi absoliuti dauguma 3259 (99%) klientų. Pensininkų duomenų bazėje yra 265, 2982 klientai gauna pastovias su darbo santykiais susijusias pajamas, 3 klientai dirba pagal verslo liudijimą arba vykdo individualią veiklą, 36 klientai yra statutiniai valstybės tarnautojai. Žvelgiant pagal lytį, duomenų bazė yra gana tolygiai pasiskirsčiusi, tačiau didesnę dalį sudaro vyrai – 55,4%, o moterys sudaro 44,6% imties duomenų. Vidutiniškai analizuojami klientai turėjo 1 sandorį analizuojamoje finansines pasaugas teikiančioje įmonėje, klientai vidutiniškai norėjo skolintis 920 eurų (maksimali skolinama suma – 15000 eurų, minimali – 50 eurų), o jų vidutinis darbo stažas 42,67 mėnesiai. Klientai vidutiniškai turėjo 10,48 vėlavimo mokėti įmokas dienų.

4 lentelė. Naudojamų kintamųjų aprašymas.

Kintamasis	Aprašymas
HR_12	Priklausomas binarinis kintamasis, kuris gali įgyti 2 reikšmes: „GOOD“ – klientas po 12 mėn. nuo sandorio sudarymo buvo mokus ¹ ir „BAD“ – klientas po 12 mėn. nuo sandorio sudarymo buvo nemokus.
SODRA_MONTHLY_INCOME	Mėnesinės kliento gaunamos NETO pajamos sandorio sudarymo metu.
SAIS_FIN_SUM_2Y	Kliento skolų, turėtų per paskutinius 2 metus iki sandorio sudarymo datos, finansų institucijoms suma.
KB_MONTHLY_INST_PRDB	Kliento turimų mėnesinių finansinių įsipareigojimų suma sandorio sudarymo metu.
NUMBER_OF_CHILDREN	Kliento turimų nepilnamečių vaikų skaičius sandorio sudarymo metu.
GRT_MARITAL_STATUS	Kliento šeimyninis statusas sandorio sudarymo metu.
X_GRT_COUNTY_CODE	Indikatorius rodantis ar klientas sandorio sudarymo metu turėjo registruotą gyvenamąją vietą.
SODRA_PROF_STATUS	Kliento profesinis statusas (gaunantis pastovias, su darbo santykiais susijusias, pajamas, pensininkas, dirbantis pagal verslo liudijimą/vykdantis individualią veiklą, statutinis valstybės tarnautojas).
SEX	Kliento lytis.
GF_NBDEAL_HISTO	Kliento pasibaigusių sandorių skaičius įmonėje.
SODRA_SENIORITY_OLD_SAL_JOB	Kliento darbo stažas.
LOAN_AMOUNT_FINANCED	Klientui paskolinta suma.
GF_MAX_DPD_HISTO	Didžiausias buvęs kliento vėlavimo dienų skaičius.

¹ Kas yra mokus ir kas yra nemokus klientas detaliau aprašyta 2.2 punkte.

4 lentelės tęsinys. Naudojamų kintamųjų aprašymas.

DTI	Kliento turimų mėnesinių įsipareigojimų ir gaunamų mėnesinių pajamų santykis.
------------	---

Tam, kad sugeneruoti kreditingumo vertinimo modeliai galėtų būti racionaliai palyginami, tyrimo imtis yra skaidoma į dvi viena nuo kitos nepriklausomas dalis: apmokymo ir testavimo. Tokiu būdu modeliai bus apmokomi atpažinti „gerus“ ir „blogus“ klientus vienoje duomenų grupėje, o jų tikslumas matuojamas naudojant kitos grupės duomenis, kurių modeliai „nematė“ savo mokymosi procese. Tam, kad abi imtys būtų visiškai nepriklausomos, jos yra sudaromos atsitiktine tvarka paskirstant 80 proc. (2570 stebėjimai) duomenų į apmokymo ir 20 proc. (656 stebėjimai) duomenų į testavimo grupes. Duomenys po suskirstymo pateikiami 10 ir 11 pav., panaudojus Rstudio funkciją *summary*.

```

HR_12      SODRA_MONTHLY_INCOME SAIS_FIN_SUM_2Y  KB_MONTHLY_INST_PRDB  NUMBER_OF_CHILDREN  GRT_MARITAL_STATUS
BAD : 396   Min. : 106.5      Min. : 0.00   Min. : 0.00          Min. :0.000         DIVORCED: 77
GOOD:2234  1st Qu.: 449.3    1st Qu.: 0.00  1st Qu.: 0.00        1st Qu.:0.000       MARRIED : 54
           Median : 583.8  Median : 0.00  Median : 27.41      Median :0.000       SINGLE  :2079
           Mean : 647.8    Mean : 21.15   Mean : 67.14       Mean :0.227        WIDOWED : 420
           3rd Qu.: 770.5  3rd Qu.: 0.00  3rd Qu.: 106.87    3rd Qu.:0.000
           Max. : 5222.2    Max. : 895.11  Max. : 1192.13     Max. : 5.000

      X_GRT_COUNTY_CODE      SODRA_PROF_STATUS SEX      GF_NBDEAL_HISTO  SODRA_SENIORITY_OLD_SAL_JOB  LOAN_AMOUNT_FINANCED
a/ Registered :2610  PENSIONER      : 218  F:1182  Min. : 0.00  Min. : 0.00          Min. : 50.0
b/ Non Registered: 20  SALARIED      :2380  M:1448  1st Qu.: 0.00  1st Qu.: 8.00        1st Qu.: 300.0
           SELF_EMPLOYED      : 3      Median : 1.00  Median : 22.00       Median : 500.0
           SPECIAL_OCCUPATION: 29      Mean : 1.44  Mean : 42.93         Mean : 930.5
           3rd Qu.: 2.00  3rd Qu.: 52.00       3rd Qu.: 947.4
           Max. : 41.00  Max. : 301.00        Max. : 15000.0
           NA's : 221

GF_MAX_DPD_HISTO      DTI
Min. : 0.00  Min. : 0.00000
1st Qu.: 0.00  1st Qu.: 0.00000
Median : 0.00  Median : 0.04720
Mean : 10.66  Mean : 0.09535
3rd Qu.: 2.00  3rd Qu.: 0.17817
Max. : 951.00  Max. : 0.39364

```

10 pav. Apmokymo grupė.

Iš 10 pav. matomų duomenų matoma, kad vidutinės klientų esančių apmokymo grupėje pajamos yra 647,8 eurų per mėnesį, vidutiniai klientų turimi finansiniai įsipareigojimai – 67,14 eurų per mėnesį. Apmokymų grupę sudaro 1182 vyrai ir 1448 moterys, 218 pensininkai, 2380 klientai gaunantys pastovias pajamas, 3 klientai dirbantys pagal verslo liudijimą arba vykdantys individualią veiklą ir 29 statutiniai valstybės tarnautojai.

HR_12	SODRA_MONTHLY_INCOME	SAIS_FIN_SUM_2Y	KB_MONTHLY_INST_PRDB	NUMBER_OF_CHILDREN	GRT_MARITAL_STATUS
BAD: 98	Min. : 127.8	Min. : 0.00	Min. : 0.00	Min. : 0.0000	DIVORCED: 27
GOOD: 558	1st Qu.: 436.7	1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.: 0.0000	MARRIED : 12
	Median : 553.9	Median : 0.00	Median : 21.23	Median : 0.0000	SINGLE : 507
	Mean : 607.5	Mean : 25.72	Mean : 60.74	Mean : 0.2165	WIDOWED : 110
	3rd Qu.: 723.5	3rd Qu.: 0.00	3rd Qu.: 98.51	3rd Qu.: 0.0000	
	Max. : 2052.0	Max. : 845.07	Max. : 449.22	Max. : 5.0000	

	X_GRT_COUNTY_CODE	SODRA_PROF_STATUS	SEX	GF_NBDEAL_HISTO	SODRA_SENIORITY_OLD_SAL_JOB	LOAN_AMOUNT_FINANCED
a/ Registered	:649	PENSIONER	: 47	F:285	Min. : 0.000	Min. : 0.00
b/ Non Registered:	7	SALARIED	:602	M:371	1st Qu.: 0.000	1st Qu.: 9.00
		SELF_EMPLOYED	: 0		Median : 1.000	Median : 21.00
		SPECIAL_OCCUPATION:	7		Mean : 1.596	Mean : 41.64
					3rd Qu.: 2.000	3rd Qu.: 50.75
					Max. : 32.000	Max. : 302.00
					NA's : 54	Max. : 13000.0

GF_MAX_DPD_HISTO	DTI
Min. : 0.000	Min. : 0.00000
1st Qu.: 0.000	1st Qu.: 0.00000
Median : 0.000	Median : 0.03597
Mean : 9.756	Mean : 0.09035
3rd Qu.: 3.000	3rd Qu.: 0.16717
Max. : 205.000	Max. : 0.39033

11 pav. Testavimo grupė

11 pav. atvaizduotuose duomenyse matoma, kad vidutinės klientų esančių testavimo grupėje pajamos yra šiek tiek mažesnės, nei esančių apmokymo grupėje – 607,5 eurų per mėnesį, vidutiniai klientų turimi finansiniai įsipareigojimai taip pat mažesni – 60,74 eurų per mėnesį. Testavimo grupę sudaro 371 vyras ir 285 moterys, 47 pensininkai, 602 klientai gaunantys pastovias pajamas, 0 klientų dirbančių pagal verslo liudijimą arba vykdančių individualią veiklą ir 7 statutiniai valstybės tarnautojai.

2.2. Priklausomas kintamasis („geras“ ir „blogas“ klientas)

Priklausomas kintamasis tyrime yra binarinis kategorinis kintamasis (HR_12), kuris gali įgyti 2 reikšmes:

1. GOOD – klientas po 12 mėn. nuo sandorio sudarymo nėra vėluojantis atlikti įmokos 90 ar daugiau dienų.
 2. BAD – klientas po 12 mėn. nuo sandorio sudarymo vėluoja atlikti įmoką 90 ar daugiau dienų.
- „Gero“ ir „blogo“ kliento apibrėžimas yra naudojamas pagal Europos bankininkystės institucijos (EBI) pateiktą nemokaus kliento išaiškinimą (EBA, 2016). Taip pat, Baselio II kapitalo susitarimo nuostatuose pateiktą kredito rizikos apibrėžimą, skolininkas nevykdo įsipareigojimų, jeigu tenkinama bent viena iš šių sąlygų:

- Bankas yra įsitikinęs, kad skolininkas neįvykdys savo turimų įsipareigojimų ir dėl šios priežasties bankui gali tekti pasinaudoti įsipareigojimų įvykdymo užtikrinimo priemonėmis;
- Skolininkas vėluoja gražinti paskolą, didesnę paskolos dalį ar palūkanas, daugiau nei 90 dienų.

2.3.Nepriklausomi kintamieji ir duomenų paruošimas.

Nepriklausomi kintamieji, naudojami modelio sudarymui yra pagrindiniai standartiniai kliento mokumo profilį nusakantys rodikliai. Kaip pastebi Kazi Rashedul Hasan (2016), dauguma kredito rizikos vertinimo modelių naudoja tokius sociogeografinius kintamuosius kaip pajamos, amžius, gyvenamoji vieta, su darbo santykiais susiję kintamieji bei kintamieji susiję su kliento kredito istorija, kreditų biuro teikiami duomenys. Būtent tokio pobūdžio kintamieji naudojami ir šiame tyrime.

Modelių gebėjimas tiksliai prognozuoti labai priklauso nuo jiems pateikiamų duomenų kokybės. Taip pat skirtingi modeliai ir programinės įrangos, skirtingai interpretuoja tam tikrus duomenų tipus. Tyrimui naudojami modeliai (Rstudio CARET paketas) nėra pritaikyti apdoroti tuščias reikšmes bei kategorinius kintamuosius. Paveiksle XY matoma, kad turimi duomenys sudaryti tiek iš skaitinių, tiek iš kategorinių kintamųjų, o kintamasis SENIORITY_OLD_SAL_JOB kartais įgyja tuščias reikšmes. Vadinasi duomenims reikalinga transformacija.

Kadangi tyrime naudojami mašininio mokymosi modeliai negali apdoroti tuščių reikšmių, galimos 2 išeitys:

1. Ištrinti visus stebėjimus, kurie įgyja kintamojo SENIORITY_OLD_SAL_JOB tuščią reikšmę.
2. Pabandyti statistiškai atkurti prarastus duomenis.

Kadangi tyrime turime sąlyginai nedaug duomenų, o kintamasis SENIORITY_OLD_SAL_JOB įgyja tuščias reikšmes 275-iuose stebėjimuose, ištrinus duomenis būtų prarastas reikšmingas kiekis labai svarbios informacijos. Tačiau kiekvienai kategorijai aklaai surašyti nuo N iki N-1 sveikųjų skaičių, pakeičiant jais kategorinių kintamųjų reikšmes, ne visada duos norimą rezultatą. Kategoriniams kintamiesiems, kai jų reikšmėse nėra akivaizdžių skirtumų, pakeitus reikšmes skaičiais modelio rezultatai gali atspindėti nelogiškas išvadas (prognozė per vidurį tarp dviejų kategorijų). Norint išvengti informacijos praradimo buvo pasirinktas antras variantas – duomenų atstatymas.

Duomenų atstatymui pasirinktas k-artimiausio kaimyno metodas, kurio pagalba tuščios kintamojo reikšmės užpildomos artimiausios tokio profilio klientams reikšmėmis. Pan, Ruilin & Yang, Tingsheng & Cao, Jianhua & Lu, Ke & Zhang, Zhanchao. (2015) savo tyrime įrodė, kad šiuo metodu gaunamos prognozės yra labai artimos realiems duomenims. Taip pat Beretta, L., &

Santaniello, A. (2016) teigia, kad šis metodas yra puikiai tinkamas norint išlaikyti duomenų vientisumą ir jų neiškraipyti. Tam, kad metodas galėtų tinkamai užpildyti reikšmes, jis automatiškai atlieka tokią duomenų transformaciją:

1. Centruoja (atima vidurkį) 9 kintamuosius;
2. Padalina 9 kintamuosius iš standartinio nuokrypio;
3. Naudojant 9 kintamuosius įvertino 5 artimiausius kaimynus užpildant tuščias reikšmes.

Kadangi naudojami mašininio mokymosi algoritmai taip pat negali susidoroti su kategoriniais kintamaisiais, juos reikia pakeisti skaitiniais. Tačiau aklaai surašyti 0 ir 1 ne visada duos norimą rezultatą. Tam, kad nebūtų prarasta reikšminga informacija, kiekvienas kategorinio kintamojo elementas yra atskiriamas (perkoduojamas) kaip atskiras skaitinis elementas (pavyzdys pateiktas 5 lentelėje).

5 lentelė. Kintamojo GRT_MARITAL_STATUS duomenų perkodavimas

GRT_MARITAL_STATUS	GRT_MARITAL_STATUS.DIVORCED	GRT_MARITAL_STATUS.MARRIED	GRT_MARITAL_STATUS.SINGLE	GRT_MARITAL_STATUS.WIDOWED
DIVORCED	1	0	0	0
MARRIED	0	1	0	0
SINGLE	0	0	1	0
WIDOWED	0	0	0	1

Kita problema yra tai, kad tyrimui naudojami modeliai subalansuoti veikti ant duomenų, kurie įgyja reikšmes nuo 0 iki 1. Taigi naudojamiems duomenims reikalingas normalizavimas (Rstudio funkcija `preProcess`). Normalizavimas atliekamas pasinaudojant Rstudio funkciją `dummyVars`, kuri atlieka duomenų transformacijas pagal šią formulę:

$$Normalizuotas(a_i) = \frac{a_i + A_{min}}{A_{max} + A_{min}}$$

čia a_i – i-toji kintamojo A reikšmė,

A_{min} – mažiausia kintamojo A reikšmė,

A_{max} – didžiausia kintamojo A reikšmė,

Jei A_{max} lygus A_{min} , tada $Normalizuotas(a_i) = 0,5$.

Po visų pakeitimų duomenų bazėje turimi 22 kintamieji (21 nepriklausomas kintamasis ir 1 priklausomas). Mokymo duomenų bazė, pasinaudojus k-artimiausio kaimyno metodu, užpildžius trūkstamas kintamojo SENIORITY_OLD_SAL_JOB reikšmes, sutvarkius kategorines reikšmes,

atlikus duomenų transformacijas ir normalizavimą pateikiama 12 pav. naudojant Rstudio funkciją *summary*.

```

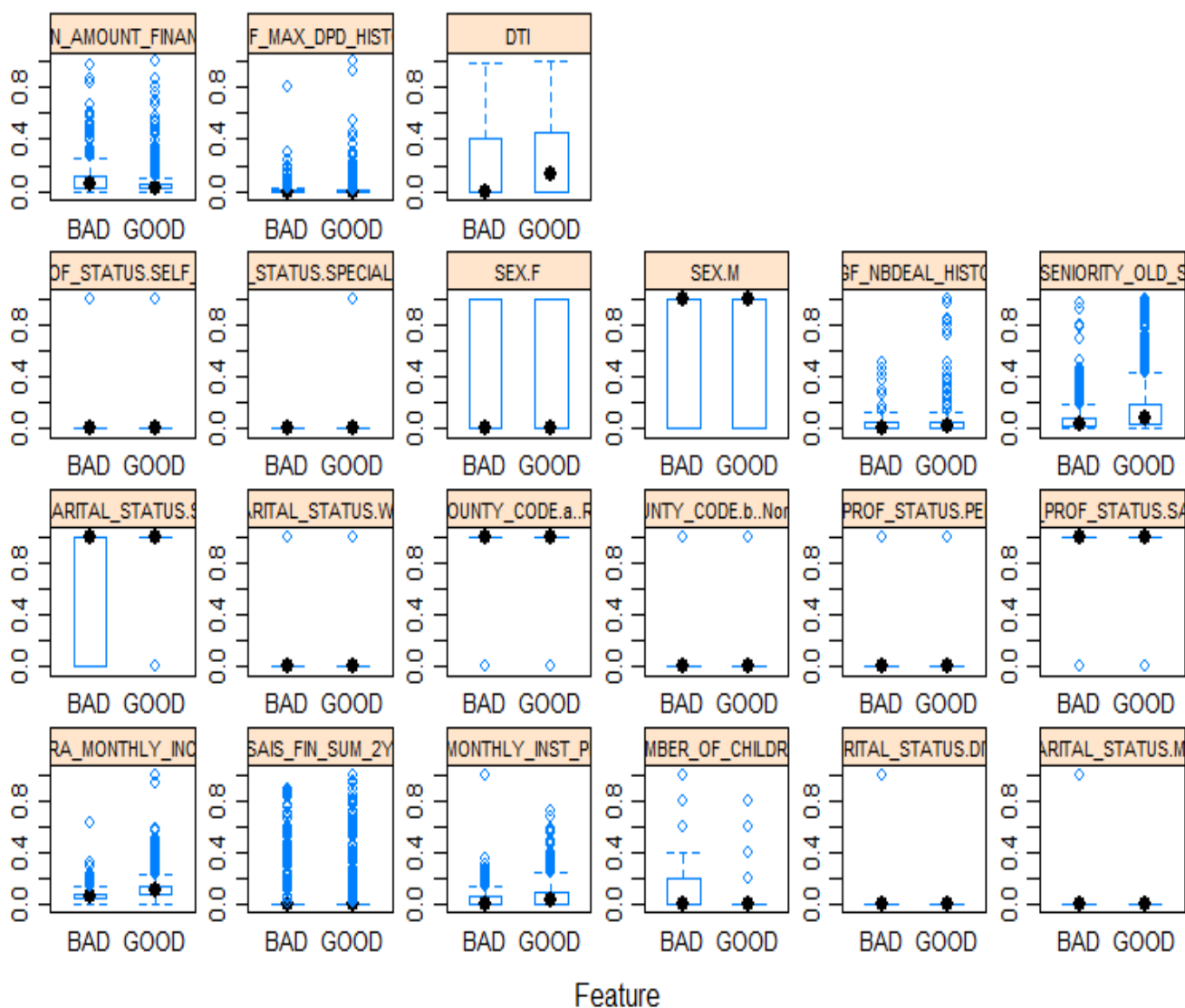
SODRA_MONTHLY_INCOME SAIS_FIN_SUM_2Y KB_MONTHLY_INST_PRDB NUMBER_OF_CHILDREN GRT_MARITAL_STATUS.DIVORCED
Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.0000 Min. :0.00000
1st Qu.:0.06702 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.00000
Median :0.09331 Median :0.00000 Median :0.02299 Median :0.0000 Median :0.00000
Mean :0.10582 Mean :0.02363 Mean :0.05632 Mean :0.0454 Mean :0.02928
3rd Qu.:0.12980 3rd Qu.:0.00000 3rd Qu.:0.08964 3rd Qu.:0.0000 3rd Qu.:0.00000
Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.0000 Max. :1.00000
GRT_MARITAL_STATUS.MARRIED GRT_MARITAL_STATUS.SINGLE GRT_MARITAL_STATUS.WIDOWED X_GRT_COUNTY_CODE.a..Registered
Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000
1st Qu.:0.00000 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:1.0000
Median :0.00000 Median :1.0000 Median :0.0000 Median :1.0000
Mean :0.02053 Mean :0.7905 Mean :0.1597 Mean :0.9924
3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000
Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000
X_GRT_COUNTY_CODE.b..Non.Registered SODRA_PROF_STATUS.PENSIONER SODRA_PROF_STATUS.SALARIED SODRA_PROF_STATUS.SELF_EMPLOYED
Min. :0.000000 Min. :0.00000 Min. :0.0000 Min. :0.000000
1st Qu.:0.000000 1st Qu.:0.00000 1st Qu.:1.0000 1st Qu.:0.000000
Median :0.000000 Median :0.00000 Median :1.0000 Median :0.000000
Mean :0.007605 Mean :0.08289 Mean :0.9049 Mean :0.001141
3rd Qu.:0.000000 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.000000
Max. :1.000000 Max. :1.00000 Max. :1.0000 Max. :1.000000
SODRA_PROF_STATUS.SPECIAL_OCCUPATION SEX.F SEX.M GF_NBDEAL_HISTO SODRA_SENIORITY_OLD_SAL_JOB
Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.00000 Min. :0.00000
1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.02990
Median :0.00000 Median :0.0000 Median :1.0000 Median :0.02439 Median :0.07641
Mean :0.01103 Mean :0.4494 Mean :0.5506 Mean :0.03511 Mean :0.14353
3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.04878 3rd Qu.:0.17940
Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.00000 Max. :1.00000
LOAN_AMOUNT_FINANCED GF_MAX_DPD_HISTO DTI HR_12
Min. :0.00000 Min. :0.000000 Min. :0.0000 BAD : 396
1st Qu.:0.01672 1st Qu.:0.000000 1st Qu.:0.0000 1st Qu.:0.0000 GOOD:2234
Median :0.03010 Median :0.000000 Median :0.1199
Mean :0.05890 Mean :0.011209 Mean :0.2422
3rd Qu.:0.06003 3rd Qu.:0.002103 3rd Qu.:0.4526
Max. :1.00000 Max. :1.000000 Max. :1.0000

```

12 pav. Mokymo duomenų bazė po duomenų transformacijos

Iš 12 pav. pateiktųjų duomenų galima nustatyti, kurie kintamieji ar kintamojo dalys bus mažiau naudingi modeliuose ir turės nedidelę prognozuojamąją galią. Žvelgiant į transformuotų duomenų vidurkį (angl. *Mean*) matoma, kad mažiausias ir didžiausias reikšmes turintys kintamieji (ar kintamojo dalys) bus mažiausiai reikšmingi, kadangi jie sudaro arba pernelyg mažą dalį duomenų arba sudaro pernelyg didelę dalį tam, kad galėtų būti naudojami išskiriant „gerus“ ir „blogus“ klientus. Mažiausias *Mean* reikšmes turi indikatorius, kad klientas vykdo individualią veiklą (0,001141), indikatorius, kad klientas neturi registruotos gyvenamosios vietos (0,007605) bei indikatorius, kad klientas yra statutinis valstybės tarnautojas (0,01103). Tuo tarpu didžiausias *Mean* reikšmes įgijo indikatorius, kad klientas turi registruotą gyvenamąją vietą (0,9924), klientas gauna pastovias pajamas (0,9049) bei indikatorius, kad kliento šeimyninis statusas yra vienišas.

Sutvarkius visus kintamuosius vizualiai galima patikrinti, kurie kintamieji atrodo reikšmingiausi. Vienas dažniausiai mokslinėje literatūroje naudojamų būdų tam patikrinti yra paprasčiausiai suskirstyti kintamąjį X pagal prognozuojamo kintamojo Y kategorijas ir jeigu tarp X kintamojo grupių yra reikšmingas poslinkis, vadinasi kintamasis turės svarbų vaidmenį modelio sudaryme. Tai galima pamatyti naudojant Rstudio funkcija *featurePlot* pasirinkus opciją *box*, kuri atvaizduoja dėžutės diagramas (angl. box plot). FeaturePlot funkcijos rezultatai atvaizduoti 13 pav.

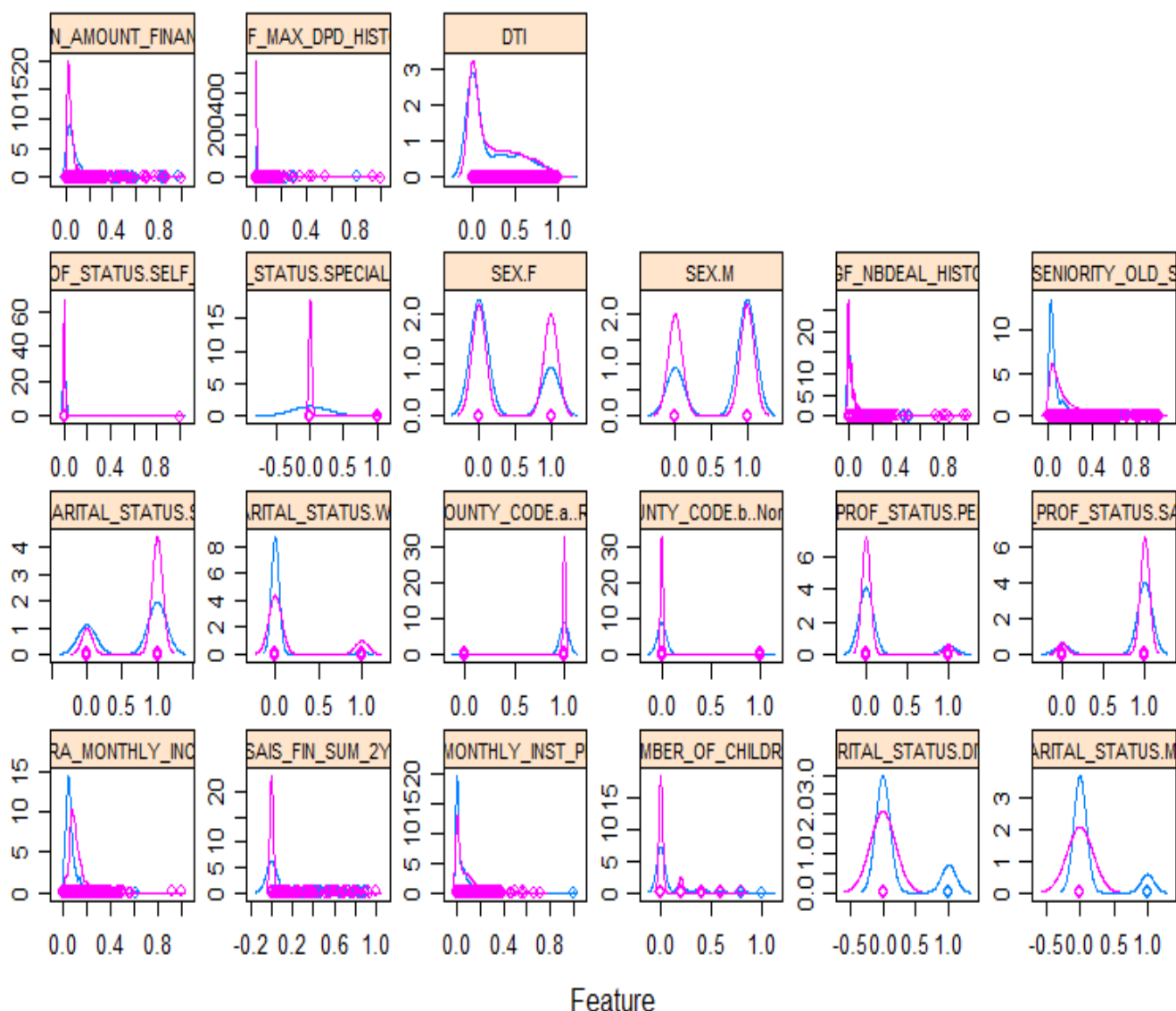


13 pav. Dėžutės diagramos, skirtos kintamųjų reikšmingumo vizualizacijai

Kiekviename iš 13 pav. atvaizduotų grafikų yra dvi dėžutės (mėlynos spalvos) kiekvienoje Y kategorijoje (GOOD ir BAD). Dėžutės viršus nurodo 25-ąjį percentilę, o apatinė dalis nurodo 75-ąją percentilę. Juodas taškas dėžutės viduje yra vidurkis. Taigi mėlynosios dėžutės nurodo intervalą, kuriame yra didžioji dalis duomenų taškų. Taip pat grafikuose matoma daug mėlynų

apskritimų, esančių virš arba po mėlynomis dėžutėmis. Šie apskritimai yra kintamojo ekstremumai (išskirtys). Iš 13 pav. grafikų matoma, kad dauguma klientų pajamas nurodančio kintamojo `SODRA_MONTHLY_INCOME` reikšmių yra akivaizdžiai pasislinkę į viršų, kai klientas yra „geras“, tokia pati situacija yra su kliento darbo stažą nurodančio kintamojo `SENIORITY_OLD_SAL_JOB` vidurkiu, taip pat šio kintamojo ekstremumai ties „gerais“ klientais yra gerokai aukščiau, nei su „blogais“ klientais. Skolinamą sumą nurodančio kintamojo `LOAN_AMOUNT_FINANCED` grafikas indikuoja, kad „blogi“ klientai dažniau skolinasi didesnes sumas. Ne toks intuityvus yra kliento įsiskolinimo ir pajamų santykį nusakančio kintamojo `DTI` grafikas. Iš šio grafiko galima daryti išvadą, kad „geresnių“ klientų įsiskolinimų dalis lyginant su jų pajamomis yra vidutiniškai didesnė, nei „blogų“ klientų.

Tam pat galima atlikti kitą panašų vizualia analize paremtą testą – panagrinti kintamųjų tankio funkcijas naudojant tą pačią Rstudio funkciją `featurePlot` ir pasirinkus opciją `density`.



14 pav. Kintamųjų tankio funkcijos.

Šiuo atveju, tikėtina, kad kintamasis bus reikšmingas modelyje, jeigu jo tankio funkcijos reikšmingai skirsis abiem kintamojo Y reikšmėms. Reikšmingesnius vizualius skirtumus galima pamatyti žvelgiant į kliento darbo stažą, šeimyninį statusą, mėnesines pajamas bei finansuojamą sumą. Tuo tarpu kliento turimų mėnesinių įsipareigojimų ir gaunamų mėnesinių pajamų santykis, didžiausias buvo kliento vėlavimo dienų skaičius bei registruotos gyvenamosios vietos indikatorius nerodo aiškių vizualių skirtumų tarp abiejų Y reikšmių

Vizualizavus X ir Y ryšius, galima tik preliminariai pasakyti, kurie kintamieji bus svarbus prognozuojant Y. Tiesiog daryti prielaidą, kad kintamieji nėra svarbūs, nes to nesimato grafiškai būtų neracionalu, nes kartais kintamieji su visiškai niekuo vizualiai neišsiskiriančia struktūra gali paaikškinti tokius Y aspektus, kurių negali vizualiai svarbūs kintamieji.

2.4. Modelių efektyvumo patikrinimo metodai.

Vienas iš modelių efektyvumo patikrinimo būdų – klasifikavimo matrica (6 lentelė). Klasifikavimo matrica pateikia daugiau informacijos negu paprasta modelio teisingai ir neteisingai prognozuotų reikšmių proporcija. Kreditingumo vertinime bankams yra labai svarbu turėti kuo mažesnę pirmos rūšies tikimybę, kai „blogas“ klientas yra laikomas „geru“. Bazelio bankų priežiūros komitetas (2005) pataria bankams detaliai analizuoti modelius juos testuojant ir tikrinant jų tikslumą, kadangi skolininkai daro tiesioginę įtaką banko kapitalo pakankamumo skaičiavimui.

6 lentelė. **Klasifikavimo matrica**

		Faktas	
		„Blogas“	„Geras“
Prognozė	„Blogas“	TP	FP
	„Geras“	FN	TN

Sutrumpinimai:

TP (angl. true positive) - teisingai klasifikuoti „blogi“ klientai

TN (angl. true negative) - teisingai klasifikuoti „geri“ klientai

FP (angl. false positive) - neteisingai klasifikuoti „blogi“ klientai

FN (angl. false negative) - neteisingai klasifikuoti „geri“ klientai

Taip pat modelių efektyvumui įvertinti naudojami jautrumo (angl. sensitivity) ir konkretumo (angl. specificity) testai, AUC, bei Cohen's kappa rodiklis. Jautrumo testas parodo, kokia dalis „blogų“ klientų buvo identifikuota teisingai, konkretumo testas parodo kokia dalis „gerų“ klientų buvo identifikuota teisingai. Cohen's kappa rodiklis yra skirtas modelio patikimumui įvertinti, rodiklis kinta ribose nuo 0 iki 1 ir yra interpretuojamas taip:

- 0 = modelis nepatikimas.
- 0– 0.20 = modelis labai mažai patikimas.
- 0.21 – 0.40 = modelis mažai patikimas.
- 0.41 – 0.60 = modelis vidutiniškai patikimas.
- 0.61 – 0.80 = modelis gana patikimas.
- 0.81 – 0.99 = modelis labai patikimas.
- 1 = modelis visiškai patikimas.

3. FIZINIŲ ASMENŲ KREDITINGUMO VERTINIMO ANALIZĖ IR TYRIMO REZULTATAI

Tyrimo rezultatų ir tiriamosios dalies analizės skyriuje pateikti visų anksčiau aprašytų modelių, naudojamų fizinių asmenų kreditingumo vertinime, rezultatai. Siekiant nustatyti ar klasikinis logistinės regresijos modelis vis dar yra optimaliausias ir tinkamiausias pasirinkimas finansų institucijoms, siekiančioms turėti pelningą portfelį, logistinės regresijos modelio rezultatai tyrime lyginami su vis labiau populiarėjančiais mašininio mokymosi modeliais. Taip pat tyrimo metu siekiama nustatyti, kuris iš analizuojamų modelių labiausiai tinkamas turimiems duomenims prognozuoti ir kurie kintamieji turi didžiausią įtaką prognozei. Visos analizės atliekamos naudojant klasifikavimo matricas, tikslumo (angl. accuracy), jautrumo, konkretumo ir patikimumo rodiklius bei vizualizacijas.

3.1. Kintamųjų svarba

Kadangi mašininio mokymosi modeliai geba patys atsirinkti, kurie kintamieji yra svarbiausi, nereikia daryti papildomos duomenų reikšmingumo analizės. Tačiau, tam, kad suprasti kaip veikia modelis ir ar jis logiškai paaiškinamas, labai svarbu žinoti kiek įtakos jam turi vienas ar kitas kintamasis. Priešingai nei tradiciniuose regresiniuose modeliuose, mašininio mokymosi modeliai negali pateikti koeficientų prie kintamųjų, todėl kintamųjų svarba pateikiama procentine išraiška. Visų modelių naudojamų kintamųjų svarba pateikiama 7 lentelėje.

7 lentelė. Modelyje naudojamų kintamųjų reikšmė modelio generacijoje

Kintamasis	LDA	LOG	NNET	GLM	KNN	RF
SODRA_MONTHLY_INCOME	100	100	100	100	100	X
SODRA_SENIORITY_OLD_SAL_JOB	64,9	64,9	32,3	31,9	64,9	X
LOAN_AMOUNT_FINANCED	44,8	44,8	65,6	71,3	44,8	X
GRT_MARITAL_STATUS.DIVORCED	35,1	35,1	42,5	0,2	35,1	X
GRT_MARITAL_STATUS.SINGLE	31,6	31,6	23,5	54,8	31,6	X
GRT_MARITAL_STATUS.MARRIED	24,5	24,5	20	0	24,5	X
GRT_MARITAL_STATUS.WINDOWED	27,9	27,9	28,9	0	27,9	X
SAIS_FIN_SUM_2Y	34,7	34,7	54,8	64,8	34,7	X
SEX.M	33,2	33,2	19,8	0	33,2	X
SEX.F	33,2	33,2	19,8	40,4	33,2	X

7 lentelės tęsinys. Modelyje naudojamų kintamųjų reikšmė modelio generacijoje

KB_MONTHLY_INST_PRDB	32,9	32,9	48,5	33,8	32,9	X
NUMBER_OF_CHILDREN	23,4	23,4	20,2	7,1	23,4	X
DTI	19,1	19,1	29,2	20,6	19,1	X
GF_MAX_DPD_HISTO	7,1	7,1	51,8	4,7	7,1	X
SODRA_PROF_STATUS.SALARIED	4,2	4,2	18,9	1	4,2	X
SODRA_PROF_STATUS.PENSIONIER	6,2	6,2	23,2	1	6,2	X
SODRA_PROF_STATUS.SELF_EMPLOYED	0	0	0	1	0	X
SODRA_PROF_STATUS.SPECIAL_OCCUPATION	2,1	2,1	1,7	1	2,1	X
GF_NBDEAL_HISTO	6,3	6,3	27,9	4,9	6,3	X
X_GRT_COUNTY_CODE.Registered	4	4	0,5	9,6	4	X
X_GRT_COUNTY_CODE.Not.Registered	4	4	0,5	0	4	X

Atlikta kintamųjų svarbos analizė parodė, kad svarbiausi kintamieji beveik visuose modeliuose yra kliento gaunamos pajamos, darbo stažas, skolinama suma, paskutinių 2-ų metų kredito istorija, šeimyninis statusas bei lytis. Neabejotinai pats svarbiausias kintamasis buvo kliento gaunamos pajamos. Šį kintamąjį naudojo visi modeliai, visose iteracijose. Nesunku ir logiškai suprasti, kad kliento mokumas ir gaunamos pajamos yra stipriai susiję veiksniai. Mažiausiai svarbūs kintamieji buvo profesinis statusas, įmonės vidinė kredito istorija, ir indikatorius, rodantis ar klientas turi registruotą gyvenamąją vietą.

Sprendimų miško modelis nepateikia galimybės įvertinti kintamųjų svarbą modelio sudaryme, todėl jis yra sunkiausiai paaiškinamas ir interpretuojamas. Tokiam modeliui reiktų pateikti tik kintamuosius, kurie analitikui atrodo logiškai paaiškinami ir suprantami.

3.2. Tyrimo rezultatai

Visų modelių klasifikavimo matricos pateikiamos 8 lentelėje.

8 lentelė. Visų modelių klasifikavimo matricos

LDA		Faktas	
		„Blogas“	„Geras“
Prognozė	„Blogas“	42	9
	„Geras“	56	549

8 lentelės tęsinys. Visų modelių klasifikavimo matricos

LOG		Faktas	
		„Blogas“	„Geras“
Prognozė	„Blogas“	59	18
	„Geras“	39	540
NNET		Faktas	
		„Blogas“	„Geras“
Prognozė	„Blogas“	75	14
	„Geras“	23	544
GLM		Faktas	
		„Blogas“	„Geras“
Prognozė	„Blogas“	63	18
	„Geras“	35	540
KNN		Faktas	
		„Blogas“	„Geras“
Prognozė	„Blogas“	67	16
	„Geras“	31	542
RF		Faktas	
		„Blogas“	„Geras“
Prognozė	„Blogas“	74	3
	„Geras“	24	555

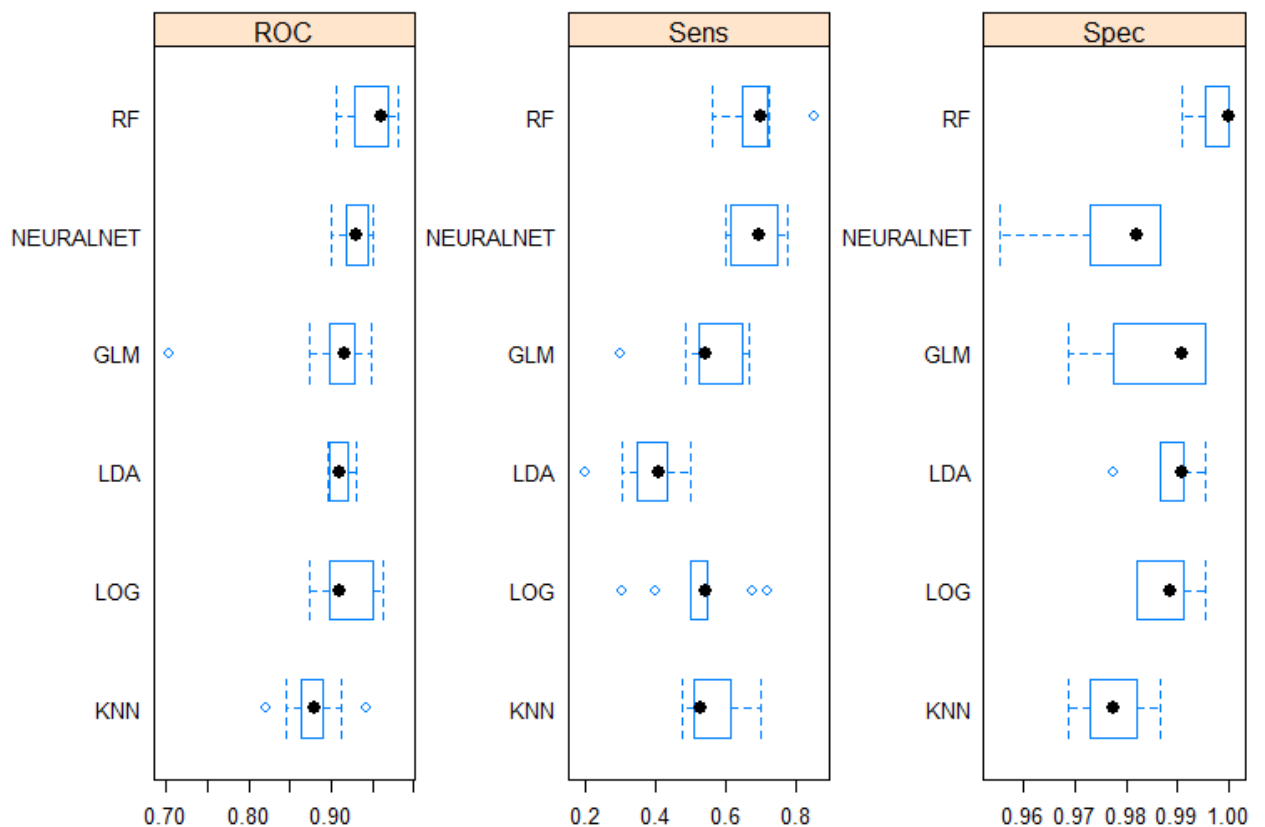
Iš klasifikavimo matricos matoma, kad didžiausią dalį antros rūšies klaidų („blogus“ klientus priskyrė prie „gerų“) padarė tiesinės diskriminacinės analizės modelis (LDA) – 56 ir logistinės regresijos (LOG) modelis - 39, tuo tarpu mažiausiai – neuroninių tinklų modelis (NNET) – 23 ir sprendimų miško modelis (RF) - 24. Daugiausiai pirmos rūšies klaidų („gerus“ klientus priskyrė prie „blogų“) padarė apibendrintas tiesinis modelis (GLM) – 18 ir logistinės regresijos (LOG) modelis – 18. Mažiausiai pirmos rūšies klaidų padarė sprendimų medžio (miško) – 3 ir tiesinės diskriminantinės analizės modelis (LDA) – 9.

Taigi klasifikavimo matrica rodo, kad saugiausias pasirinkimas kredito rizikos vertinimui yra naudoti neuroninių tinklų (89 prognozuoti „blogi“ klientai) arba k-artimiausio kaimyno modelius (83 prognozuoti „blogi“ klientai). Šie modeliai yra konservatyviausi ir dažniau laikys klientą „blogu“, jeigu nebus visiškai aišku ar jis „geras“ ar „blogas“. Taip pat iš klasifikavimo matricos galima matyti, kad nors k-artimiausio kaimyno modelis ir yra konservatyvus, tačiau jis nėra optimalus, nes net ~19% prognozuojamų „blogų“ klientų iš tiesų yra „geri“ klientai, tuo tarpu neuroninių tinklų modelio atveju – ~16%, o sprendimų miško modelio prognozė prie „gerų“ klientų priskiria vos ~4% prognozuojamų „blogų“ klientų.

Toliau, naudojant AUC, jautrumo, konkretumo, tikslumo, Kappa ir subalansuoto tikslumo rodiklius bus įvertinamas modelių efektyvumas. Modelių efektyvumo rodikliai pateikti 9 lentelėje ir grafiškai atvaizduoti 15 pav.

9 lentelė. Modelių efektyvumo rodikliai.

Modelis	AUC	Jautrumas	Konkretumas	Tikslumas	Kappa	Subalansuotas tikslumas
LDA	0,9107555	0,3889103	0,9892497	0,9009	0,5141	0,7062
LOG	0,9183379	0,5301923	0,9879104	0,9131	0,625	0,78489
NNET	0,9301289	0,6940385	0,9776345	0,9436	0,7693	0,8701
GLM	0,8952655	0,5508974	0,987474	0,9192	0,6576	0,80530
KNN	0,8800632	0,5607051	0,9771761	0,9284	0,6991	0,8275
RF	0,9503919	0,6866667	0,9973154	0,9588	0,8224	0,8749



15 pav. Grafinis AUC, jautrumo ir konkretumo rodiklių atvaizdavimas.

Iš rezultatų pateiktų 9 lentelėje matoma, kad sprendimų miško modelis (RF) parodė geriausius rezultatus AUC, konkretumo, tikslumo, kappa ir subalansuoto tikslumo rodikliuose. Tuo tarpu tiesinės diskriminantinės analizės modelio rezultatai buvo prasčiausi 4-iuose iš 6-ių kriterijų

(jautrumo, tikslumo, kappa ir subalansuoto tikslumo). Logistinės regresijos modelis (LOG) neparodė geriausio rezultato nei viename iš kriterijų, tačiau nei viename nebuvo ir prasčiausias. Iš 15 pav. pateikto grafinio modelio efektyvumo rodiklių atvaizdavimo taip pat aiškiai matyti, kad sprendimų miško modelis (RF) yra akivaizdus lyderis ROC, jautrumo ir konkretumo testuose, o kiti metodai varijuoja tarpusavyje ir geriau pasirodo vertinant vieną kriterijų, tačiau prasčiau atrodo žvelgiant į kitą.

IŠVADOS

1. Išanalizavus dažniausiai mokslinėje literatūroje analizuojamus fizinių asmenų kredito rizikos vertinimo modelius, iš nagrinėtos literatūros pastebėta, kad bankai dažniausiai renkami lengviau paaiškinamus, lengvai interpretuojamus ir paprastai pritaikomus modelius fizinių asmenų kreditingumo vertinimui. Geriausiai šiuos kriterijus atitinka logistinės regresijos modelis, kuris kiekvienam vertinamam klientui apskaičiuoja tikimybę, kad tas klientas bus „blogas“, t.y. nemokus. Logistinės regresijos modelyje aiškiai atvaizduojami jame naudojami kintamieji ir jų svarba tikimybių skaičiavime. Dėl šiuos priežasties logistinės regresijos modelį lengva panaudoti klientų reitingavimui, o jį kuriantiems ekspertams – paaiškinti modelio logiką vadovams. Tuo tarpu tokie vis labiau populiarėjantys mašininio mokymosi modeliai kaip sprendimų medis (miškas), neuroninių tinklų modelis ar k-artimiausio kaimyno modelis, vis dar vengiami praktikoje dėl savo aiškaus interpretavimo stygiaus. Taip pat šių modelių praktinį panaudojimą apsunkina kompiuterinės galios reiklumas.
2. Visi nagrinėti modeliai naudoja tam tikras praktikoje sunkiai įgyvendinamas prielaidas, tokias kaip tiesinės kintamųjų tarpusavio priklausomybės, duomenų pasiskirstymas pagal normalųjį skirstinį ar, atmetimo taisyklėmis paremto modelio atvejų, tiesiog vadovaujasi savo ar kitų patirtimi. Tačiau neturint pakankamai duomenų empirinei klientų portfelio analizei ekspertinio vertinimo modelis yra vienintelis pasirinkimas.
3. Rizikos apibrėžimo, sampratos analizė parodė, kad rizika yra tikimybė, kad banko pelningumas skirsis nuo to, kurį tikimasi gauti. Kuo ši tikimybė yra didesnė, tuo didesnis ir prisiimamos rizikos lygis. Pasirinktas netinkamas rizikos lygis, gali lemti banko bankrotą ir turėti stiprų poveikį visai rinkos ekonomikai. Literatūros analizė parodė, kad vienas pagrindinių bankinio sektoriaus pajamų šaltinių Lietuvoje yra fizinių asmenų finansavimas. Todėl nuo to, kaip bankai sugebės suvaldyti būtent šią riziką ir priklausys bankinio sektoriaus veiklos tęstinumas. Pagrindinės rizikos, su kuriomis susiduria bankinis sektorius atlikdamas fizinių asmenų kreditingumo vertinimą yra rinkos rizika, užsienio valiutos rizika, palūkanų normų rizika, operacinė rizika, likvidumo rizika, sisteminė rizika ir kredito rizika. Pastaroji yra bene svarbiausia, nes tiesiogiai prisideda prie įmonės pelningumo. Kredito rizika apibrėžiama kaip rizika, kad banko skolininkas neįgyvendins finansinių įsipareigojimų bankui pagal numatytas sutarties sąlygas.

4. Išanalizavus rizikos valdymo svarbą versle, bankiniame sektoriuje, galima daryti išvadą, kad tai yra viena svarbiausių veiksnių bankinio sektoriaus valdyme. Teisingas rizikos valdymas didina tikslų pasiekimo tikimybę, padeda identifikuoti grėsmes ir galimybes, gerina valdymo kokybę, optimizuoja išteklių paskirstymą, mažina nuostolius, greitina organizacijos reakciją į pokyčius.
5. Išanalizavus keletą mokslinių tyrimų apie fizinių asmenų kreditingumo vertinimą, buvo pastebėta, kad rezultatai yra gana priešaringi. Galima daryti prielaidą, kad fizinių asmenų kreditingumo vertinimo modelių rezultatai priklauso nuo įvairių aplinkybių, tokių kaip finansinės institucijos rinkos geografinė padėtis, klientų kultūriniai aspektai ir surenkamų duomenų kiekis bei kokybė. Informacijos apie tai, kokius modelius tinkamiausia naudoti finansines paslaugas teikiančioms įmonėms veiklą vykdančioms Lietuvoje rasta nebuvo. Tačiau bendro atsakymo, tinkančio visiems rinkoms dalyviams, gauti praktiškai neįmanoma, todėl buvo pasirinkta analizuoti konkrečios finansines paslaugas Lietuvoje teikiančios įmonės klientų mokumą.
6. Mokslinėje literatūroje dažniausiai analizuojami septyni statistiniai fizinių asmenų mokumo vertinimo modeliai: a) tiesinė diskriminantinė analizė; b) logistinė regresija; c) neuroninių tinklų modelis; d) apibendrintas tiesinis modelis; e) k-artimiausio kaimyno modelis; f) sprendimų medis (miškas); g) atmetimo taisyklėmis pagrįsta mokumo vertinimo sistema. Visi modeliai turi ir privalumų ir trūkumų ir bendro sutarimo, kuris yra tinkamas universaliai naudojimui nepriklausomai nuo geografinės, politinės, socialinės ir t.t. padėties valstybėje, kurioje veikia finansines paslaugas teikianti įmonė ar nuo pačios įmonės dydžio – nėra. Metodų empirinio taikymo rezultatų mokslinėje literatūroje buvo rasta labai mažai, tai rodo, kad fizinių asmenų kreditingumo vertinimo problema yra tikrai mažai analizuojama, nors tai yra svarbi ekonomikos dalis. Rasti empiriniai tyrimai mokslinėje literatūroje parodė labai prieštarigus rezultatus – visais trimis rasta atvejais tiksliausias modelis buvo vis kitas. Tiksliausi modeliai pagal mokslinę literatūrą: sprendimų medžio modelis, neuroninių tinklų modelis, logistinės regresijos modelis.
7. Atlikus empirinį tyrimą, buvo nustatyta, kad didžiausią prognozuojamąją galią tiesinės diskriminacinės analizės, k-artimiausio kaimyno ir logistinės regresijos modeliuose turi kliento pajamos, darbo stažas, finansuota suma ir šeimyninis statusas. Neuroninių tinklų modelis reikšmingiausiais kintamaisiais laikė kliento pajamas, finansuojamą sumą, kliento

kredito istoriją bei kliento didžiausią buvusį vėlavimo dienų skaičių. Apibendrintas tiesinis modelis svarbiausiais kintamaisiais laikė kliento pajamas, finansuojamą sumą, kliento kredito istoriją ir šeimyninį statusą. Tuo tarpu mažiausią prognozuojamąją galią tiesinės diskriminacinės analizės, k-artimiausio kaimyno ir logistinės regresijos modeliuose turėjo kliento registruotos gyvenamosios vietos indikatorius, didžiausias buvus kliento vėlavimo dienų skaičius ir profesinis statusas. Neuroninių tinklų modelis mažiausiai reikšmingais laikė kliento profesinį statusą, registruotos gyvenamosios vietos indikatorius ir lytį. Apibendrintas tiesiniame modelyje mažiausią prognozuojamąją galią turėjo kliento šeimyninis statusas, lytis ir registruotos gyvenamosios vietos indikatorius.

8. Vienas svarbiausių, tikslaus fizinių asmenų kreditingumo vertinimo kriterijų yra pakankama tyrimo imtis. Rekomenduojama, kad tyrimo imtyje būtų ir reikšmingas „blogų“ klientų skaičius, neturint pakankamai stebinių arba nepakankamai nemokių klientų, modelis bus netikslus. Taip pat rekomenduojama turėti kuo daugiau kintamųjų, apibūdinančių kliento profilį, kad modeliai turėtų galimybę atrinkti didžiausią prognozuojamąją galią turinčius kintamuosius.
9. Empirinis tyrimas naudojant vienos iš Lietuvoje veikiančių finansines paslaugas teikiančių įmonių duomenis, parodė, kad geriausius rezultatus galima pasiekti naudojant sunkiau interpretuojamus modelius. Sprendimų medžio (miško) modelis parodė geriausius rezultatus penkiuose iš šešių (AUC – 0,95, konkretumo – 0,99, tikslumo – 0,95, kappa – 0,82 ir subalansuoto tikslumo – 0,87), modelio efektyvumo vertinimui naudotų kriterijų. Tuo tarpu tiesinės diskriminantinės analizės modelis buvo prasčiausias keturiuose iš šešių kriterijų (jautrumas – 0,39, tikslumas – 0,90, kappa – 0,51, subalansuotas tikslumas – 0,71). Dažniausiai praktikoje naudojamas ir plačiausiai mokslinėje literatūroje nagrinėjamas – logistinės regresijos modelis neparodė geriausių rezultatų nei viename iš kriterijų, tačiau nei viename nebuvo ir prasčiausias (AUC – 0,92, jautrumas – 0,53, konkretumas – 0,988, tikslumas – 0,91, kappa – 0,63, subalansuotas tikslumas – 0,78). Logistinės regresijos modelio lengvas pritaikymas ir interpretavimas, autoriaus nuomone, yra reikšmingesni, nei skirtumas tarp modelių efektyvumo, todėl šis modelis pelnytai yra populiariausias fizinių asmenų kreditingumo vertinimo metodų sąrašė. Tačiau siekiant maksimalaus rezultato ir turint pakankamą kompiuterinės galios išteklių rezervą, analizuojamu atveju, geriausias pasirinkimas būtų naudoti sprendimų miško modelį, tačiau rekomenduojama kruopščiai

atsirinkti šiame modelyje naudojamus kintamuosius, kadangi modelis nesuteikia galimybės logiškai įvertinti kintamųjų prognozuojamosios galios.

**ASSESSMENT OF METHODS FOR CREDITWORTHINESS EVALUATION OF
CONSUMER CREDIT IN THE CASE OF ONE OF THE COMPANIES
PROVIDING FINANCIAL SERVICES IN LITHUANIA**

Eimantas PALIONIS Paper for the Master's degree

Business economics Master's Program

Vilnius University, Faculty of Economics and Business

administration, Management Department Supervisor – prof. dr. A.

Paškevičius Vilnius, 2019

SUMMARY

56 pages, 9 charts, 15 pictures, 62 references.

Credit rating methods were first introduced in 1940 and over the years they have become an increasingly important tool for successful financial institutions. In the 1960's when credit cards were created, credit rating methods became particularly important to banks. Nowadays, creditworthiness assessment models are widely used as tools for making consumer credit financing or non-financing decisions, or even predicting bankruptcy of companies.

The main purpose of creditworthiness assessment is to build a model that establishes a customer's creditworthiness rating using historical data. Credit rating models are significantly improving with the growing amount of information collected. An effective credit rating system can help protect financial sector companies from bankruptcy, maximize profits and minimize the risk of customer insolvency. A wide number of different credit assessment models are currently being developed and they produce different results across different customer segments.

The subject of this master's thesis is the market of consumer credit for natural persons.

The Scientific Problem: Major problem faced by banks and other financial institutions in Lithuania - how to determine which customer will be “good” and which will be “bad” before providing financial services, how long does it take to determine that the customer is “bad”, which variables best describe a “bad” and “good” customer, and which method to use to get the best possible result. The creditworthiness evaluation of companies is a widely discussed topic in Lithuania, while the analysis of the creditworthiness evaluation of natural persons, although not less important, is not so widely covered. Companies providing financial services in Lithuania often have to buy models for assessing the creditworthiness of natural persons from third parties, because they themselves have very little understanding about creating such models. Therefore, the main problem is to determine which statistical model is best suited for companies providing financial services in the Lithuanian market.

Aim of the work: to evaluate the creditworthiness assessment methods used in practice, to understand the importance of risk assessment in business and in the financial sector in particular, to determine which method is best for predicting customer insolvency for a particular financial services company operating in Lithuania.

The work consists of literature analysis, methodology and empirical research results. The literature analysis defines the most often used statistical models for natural persons' creditworthiness evaluation, concept of risk itself and types of risks faced in financial sector. In the empirical analysis section analysis is performed on clients from one of the companies providing financial services in Lithuania. Variable significance tests are made, and models compared to each other. Finally, the paper presents the results of empirical research and their evaluation.

The results of the empirical analysis reveal that different models are using different variables to forecast whether or not client will be insolvent and in particular case Random forest model showed the best results while Linear discriminant analysis provided the worst test results.

It is recommended that financial institutions would take into account that significant number of variables is important when trying to make a creditworthiness evaluation model. The results showed that models, which are hard to interpret provided the best results. It is recommended to use such models with caution, since it is very hard to understand if it is in line with business plan. There is no universal creditworthiness evaluation model so financial institutions are recommended to develop models themselves so they would represent their client profile.

LITERATŪROS SĄRAŠAS

1. Agresti, A. (2007). An Introduction to categorical data analysis. 2nd Edition, John Wiley and Sons, Inc., New York.
2. Agresti, A. (2013) Categorical Data Analysis. 3rd Edition, John Wiley & Sons Inc., Hoboken.
3. Aleknevičienė V. (1997) Investicijų rizikos valdymas (žemės ūkio produktus gaminančių ir perdirbančių įmonių pavyzdžiu): daktaro disertacija, socialiniai mokslai, ekonomika, LŽŪU.
4. Altman, E. I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, Vol 23(4) p. 589–609.
5. BACK, B., LAITINEN, T., SERE, K. & WEZEL, M. V. (1996) Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms. Turku, Finland: Turku Centre for Computer Science.
6. Bank for International Settlements (2005). Update on work of the Accord Implementation Group related to validation under the Basel II Framework. Basel Committee on Banking Supervision. Prieiga per internetą: https://www.bis.org/publ/bcbs_n14.htm
7. Bareikaitė, E., Martinkutė-Kaulienė, R. (2014) Liquidity risk and its management in lithuanian banking system. *Science: Future of Lithuania*. Feb2014, Vol. 6 Issue 1, p64-71. 8p. f
8. Bazelio bankų rizikos komitetas. (2001) Risk management practices and regulatory capital: cross-sectoral comparison. *The Joint Forum*. Prieiga per internetą: <https://www.bis.org/publ/joint04.pdf>
9. Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: a critical evaluation. *BMC medical informatics and decision making*, 16 Suppl 3(Suppl 3), 74. doi:10.1186/s12911-016-0318-z
10. Bernstein, P. L. (1998) Against the Gods: The Remarkable Story of Risk. *Wiley* p. 383
11. CHUANG, C.L. and R.H. LIN (2009) Constructing a reassigning credit scoring model, *Expert Systems with Applications*, 36, 1685–1694.
12. COATS, P.K. and L.F. FANT (1993) Recognizing financial distress patterns using a neural network tool, *Financial Management*, 2, 142–155.
13. DESAI, V.S., J.N. CROOK and G.A. OVERSTREET (1996) A comparison of neural networks and linear scoring models in the credit union environment, *European Journal of Operational Research*, 95, 24–37.
14. Dionne, G. (2013) Risk Management: History, Definition, and Critique. *Wiley Online Library*. Prieiga per internetą: <https://onlinelibrary.wiley.com/doi/full/10.1111/rmir.12016>

15. Doumpos, M., & Zopounidis, C. (2014). Credit Scoring Multicriteria Analysis in Finance (pp. 43-59): Springer.
16. EBA (2016). Guidelines on the application of the definition of default under Article 178 of Regulation (EU) No 575/201. Prieiga per internetą: <https://eba.europa.eu/documents/10180/1597103/Final+Report+on+Guidelines+on+default+definition+%28EBA-GL-2016-07%29.pdf/004d3356-a9dc-49d1-aab1-3591f4d42cbb>
17. Ertekin, O. (2010). Ipma2010 Presentation. *Slideshare*. Prieiga per internetą: <https://www.slideshare.net/omerertekin/ipma2010-presentation-omer-ertekin>.
18. European banking authority. Market risk. Prieiga per internetą: <https://eba.europa.eu/regulation-and-policy/market-risk>
19. Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, Vol. 7, 179-188.
20. GARP (2012), Foundations of risk management, *GARP*, p. 2017
21. Genriha, Irina & Voronova, Irina. (2012). Methods for Evaluating the Creditworthiness of Borrowers. *Ekonomika un uzņēmējdarbība*. Nr.22, 2012, 42.-49.lpp. ISSN 1407-7337. e-ISSN 2255-8756.. 42-49.
22. Gouvêa, Maria & Bacconi, Eric. (2007). Credit Risk Analysis Applying Logistic Regression, Neural Networks and Genetic Algorithms Models. Prieiga per internetą: https://pomsmeetings.org/ConfProceedings/007/CDProgram/Topics/full_length_papers_files/007-0210.pdf
23. HAND, D. J. & HENLEY, W. E. (1997) Statistical classification methods in consumer credit scoring: a review. *J. R. Stat. Assoc., Ser. A*, 160, 523–541.
24. Hillson, D. (2016) Enterprise Risk Management: Managing Uncertainty and Minimising Surprises. Prieiga per internetą: <http://www.risk-doctor.com/uploaded/AdvisingUpwardsHillsonChapter2EnterpriseRiskMgt.pdf>
25. Horcher, K. A. (2005). Essentials of Financial Risk Management. *John Wiley & Sons, Inc.* Prieiga per internetą: <https://ebookcentral.proquest.com/lib/viluniv-ebooks/reader.action?docID=227407>
26. HUANG, Z., H. CHEN, C.J. HSU, W.H. CHEN and S. WU (2004) Credit rating analysis with support vector machines and neural networks: a market comparative study, *Decision Support Systems*, 37, 543–558
27. Huseyin Ince & Bora Aktan (2009) A comparison of data mining techniques for credit scoring in banking: A managerial perspective, *Journal of Business Economics and Management*, 10:3, 233-240

28. Ivaškevičius D., Sakalas A. (1997) Bankų vadyba. *Kaunas, Technologija*, p. 239
29. JASIENĖ, M. (2002) Pinigų normos ir palūkanų normų rizika. *Vilnius: Vilniaus universitetas*. 61p. ISSN 1392-1258
30. JENSEN, H.L. (1992) Using neural networks for credit scoring, *Managerial Finance*, 18, 15–26.
31. Kazi Rashedul Hasan (2016). Development of a Credit Scoring Model for Retail Loan Granting Financial Institutions from Frontier Markets. *International Journal of Business and Economics Research*. Vol. 5, No. 5, 2016, pp. 135-142. doi: 10.11648/j.ijber.20160505.11
32. Knight, F. H. (1921) Risk, Uncertainty and Profit. *Harper and Row, New York*.
33. Lessmann S, Baesens B., Seow H. and Lyn C. Thomas *European Journal of Operational Research*, 2015, vol. 247, issue 1, 124-136
34. Lietuvos bankas. Nutarimas dėl kapitalo pakankamumo skaičiavimo bendrųjų nuostatų. Prieiga per internetą: <https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.289756>
35. Lietuvos Bankas. Pagrindiniai bankų veiklos rodikliai. Prieiga per internetą: <https://www.lb.lt/lt/pagrindiniai-banku-veiklos-rodikliai> .
36. Lietuvos banko terminų bazė. 2006 m. lapkričio 9 d. Lietuvos banko valdybos nutarimas Nr. 138 „Dėl kapitalo pakankamumo skaičiavimo bendrųjų nuostatų“ (Žin., 2006, Nr. 142-5442) Prieiga per internetą: <http://www.rastija.lt/LBTB/Lietuvos-banko-termin%C5%B3-baz%C4%97/bir%C5%BEos-preki%C5%B3-kainos-rizika>
37. Mackevičius, J., Silvanavičiūtė, S. (2006) Įmonių Bankroto Prognozavimo Modelių Tinkamumo Nustatymas. *Business: Theory & Practice*, [s. l.], v. 7, n. 4, p. 193–202
38. Martinkutė-Kaulienė, R., Stasytė, V. (2018), Rizikos valdymas, *Technika*, p. 215
39. McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. 2nd Edition, Chapman and Hall, London.
40. MESTER, L. J. (1997) What Is the Point of Credit Scoring? Philadelphia, USA: Federal Reserve Bank of Philadelphia.
41. Mirković, V., Boban, D., Boris, S. (2013) Market risk management in banks. *ResearchGate* Prieiga per internetą: https://www.researchgate.net/publication/321749017_MARKET_RISK_MANAGEMENT_IN_BANKS
42. Nance, D. R., Smith, C.W., Smithson, C. W. (1993) On the determinants of corporate hedging. *The Journal of Finance* Vol. 48, No. 1 (Mar., 1993), p. 267-284 Prieiga per internetą: https://www.jstor.org/stable/2328889?seq=1#metadata_info_tab_contents

43. Olteanu, A. (2012) Operational risk management. *Constanta Maritime University, Romania* p.335-338.
44. Pan, Ruilin & Yang, Tingsheng & Cao, Jianhua & Lu, Ke & Zhang, Zhanchao. (2015). Missing data imputation by K nearest neighbours based on grey relational structure and mutual information. *Applied Intelligence*. 43. 10.1007/s10489-015-0666-x.
45. Petria, N., Petria, L. (2009) Operational risk management and Basel II. „*Constantin Brancoveanu*“ *University, Rm. Vâlcea*. p. 96-100.
46. PIRAMUTHU, S. (1999) Financial credit-risk evaluation with neural and neurofuzzy systems, *European Journal of Operational Research*, 112, 310–321.
47. Ranganatha P.n, C. S. Pramesh and Aggarwal R. Common pitfalls in statistical analysis: Logistic regression *Perspect Clin Res*. 2017 Jul-Sep; 8(3): 148–151.
48. Reichert, A.K., Cho, C.C. and Wagner, G.M. (1983). An examination of Conceptual Issues Involved in Developing Credit-scoring Models. *Journal of Business and Economic Statistics*, Vol. 1, 101-114.
49. Rescher, N. (1983), Risk: A Philosophical Introduction to the Theory of Risk Evaluation and Management, *University of Pittsburgh* p. 35-7.
50. SANDS, E., SPRINGATE, G., VAR, T. (1983) Predicting Business Failure. *CGA Magazine*, May, p. 24–27.
51. Stasytytė, V., Aleksienė, L. (2015) Įmonės veiklos rizikos vertinimas ir valdymas mažose ir vidutinėse įmonėse. *Business: Theory & Practice*. jun2015, Vol. 16 Issue 2, p140-148. 9p.
52. Statman, M. (1987) How many stocks make a diversified portfolio? *Journal of financial and quatitative analysis* Vol 22, NO. 3 September 1987. Prieiga per internetą: https://www.jstor.org/stable/2330969?seq=1#metadata_info_tab_contents
53. SUMMERS, B. & RIDING, A. (1996) Networks That Learn and Credit Evaluation. Arkansas, USA: University of Central Arkansas.
54. Sustersic, Maja and Mramor, Dusan and Zupan, Jure, Consumer Credit Scoring Models with Limited Data (March 1, 2007). EFA 2007 Ljubljana Meetings Paper. Prieiga per internetą: <https://ssrn.com/abstract=967384> arba <http://dx.doi.org/10.2139/ssrn.967384>
55. TAFFLER, R. J. (1984) Empirical models for the monitoring the UK corporations. *Journal of Banking and Finance*, Vol 8, p. 199–227.
56. Trinkle, B. S. and Baldwin, A. A. (2016) ‘Research Opportunities for Neural Networks: The Case for Credit’, *Intelligent Systems in Accounting, Finance & Management*, 23(3), pp. 240–254. doi: 10.1002/isaf.1394.
57. Urniežius, R. (2001) Rizika. *Vilnius : Mintis*, 2001. 183 p.

58. Vaitkevičiūtė, V. (2001) Tarptautinių terminų žodynas. *Vilnius: Žodynas, 2001.*
59. VELLIDO, A., P.J.G. LISBOA and J. VAUGHAN (1999) Neural networks in business: a survey of applications 1992–1998, *Expert Systems with Applications*, 17, 51–70.
60. Wiginton, J. C. (1980). A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour. *Journal of Financial and Quantitative Analysis*, Vol. 15, 757-770.
61. ZAVGREN, Ch. (1985) Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis. *Journal of Business Finance & Accounting*, Vol 12 (1), p. 19–45
62. Zenzerović R. (2011) CREDIT SCORING MODELS IN ESTIMATING THE CREDITWORTHINESS OF SMALL AND MEDIUM AND BIG ENTERPRISES *Croatian Operational Research Review (CRORR)*, Vol. 2, 143

Priedai

1. Tiesinės diskriminantinės analizės modelis

```
> ctrl <- trainControl(method="cv",
+                      summaryFunction=twoClassSummary,
+                      classProbs=T,
+                      savePredictions = T)
> model_lda = train(HR_12 ~ ., data=trainData, method='lda', tuneLength=5, trControl=ctrl)
There were 12 warnings (use warnings() to see them)
> model_lda
Linear Discriminant Analysis

2630 samples
 21 predictor
  2 classes: 'BAD', 'GOOD'

No pre-processing
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 2367, 2366, 2367, 2368, 2368, 2367, ...
Resampling results:

   ROC      Sens      Spec
0.9107555 0.3889103 0.9892497

>
> predicted <- predict(model_lda, testData4)
> head(predicted)
[1] BAD BAD BAD BAD BAD BAD
Levels: BAD GOOD
>
> # Compute the confusion matrix
> confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')
Confusion Matrix and Statistics

          Reference
Prediction BAD GOOD
   BAD     42    9
   GOOD    56   549
```

2. Logistinės regresijos modelis

```
> library(Liblinear)
> model_log = train(HR_12 ~ ., data=trainData, method='regLogistic', tuneLength=5, verbose=F, trControl=ctrl)
Warning message:
In train.default(x, y, weights = w, ...) :
  The metric "Accuracy" was not in the result set. ROC will be used instead.
> model_log
Regularized Logistic Regression

2630 samples
 21 predictor
  2 classes: 'BAD', 'GOOD'

> predicted <- predict(model_log, testData4)
> head(predicted)
[1] BAD BAD BAD BAD BAD BAD
Levels: BAD GOOD
>
> # Compute the confusion matrix
> confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')
Confusion Matrix and Statistics

          Reference
Prediction BAD GOOD
   BAD     59   18
   GOOD    39   540
```

3. Neuroninių tinklų modelis

```

> model_nnet <- train(HR_12 ~ ., data = trainData, method = 'nnet', preProcess = c('center', 'scale'), tuneLength=5, trcontrol=ctrl,
+ tuneGrid=expand.grid(size=c(10), decay=c(0.1)))
# weights: 231
initial value 1223.133808
iter 10 value 512.179645
iter 20 value 431.343344
iter 30 value 399.251221
iter 40 value 375.692563
iter 50 value 356.520910
iter 60 value 348.182289
iter 70 value 342.504913
iter 80 value 338.966602
iter 90 value 335.847438
iter 100 value 333.298912
final value 333.298912
stopped after 100 iterations

> model_nnet
Neural Network

2630 samples
 21 predictor
  2 classes: 'BAD', 'GOOD'

Pre-processing: centered (21), scaled (21)
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 2368, 2368, 2366, 2368, 2366, 2366, ...
Resampling results:

   ROC      Sens      Spec
0.9230099 0.6742308 0.9789558

Tuning parameter 'size' was held constant at a value of 10
Tuning parameter 'decay' was held constant at a value of 0.1
>
> predicted <- predict(model_nnet, testData4)
> head(predicted)
[1] BAD BAD BAD BAD BAD BAD
Levels: BAD GOOD
>
> # Compute the confusion matrix
> confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')
Confusion Matrix and Statistics

          Reference
Prediction BAD GOOD
BAD      72   13
GOOD     26  545

```

4. Apibendrintas tiesinis modelis

```

> model_glm = train(HR_12~., data=trainData, method='glm', tuneLength=5, trControl=ctrl)
There were 41 warnings (use warnings() to see them)
> model_glm
Generalized Linear Model

2630 samples
 21 predictor
  2 classes: 'BAD', 'GOOD'

No pre-processing
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 2367, 2366, 2366, 2367, 2367, 2367, ...
Resampling results:

   ROC      Sens      Spec
0.9190709 0.5653205 0.9870115

>
> predicted <- predict(model_glm, testData4)
Warning message:
In predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
  prediction from a rank-deficient fit may be misleading
> head(predicted)
[1] BAD BAD BAD BAD BAD BAD
Levels: BAD GOOD
>
> # Compute the confusion matrix
> confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')
Confusion Matrix and Statistics

      Reference
Prediction BAD GOOD
BAD      63   18
GOOD     35  540

```

5. K-artimiausio kaimyno modelis

```

> model_kknn = train(HR_12 ~ ., data=trainData, method='kknn', tuneLength=15, trControl=ctrl)
Warning message:
In train.default(x, y, weights = w, ...) :
  The metric "Accuracy" was not in the result set. ROC will be used instead.
> model_kknn
k-Nearest Neighbors

2630 samples
 21 predictor
  2 classes: 'BAD', 'GOOD'

No pre-processing
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 2367, 2366, 2368, 2366, 2368, 2367, ...
Resampling results across tuning parameters:

 kmax  ROC      Sens      Spec
 5     0.8598707 0.5855769 0.9767197
 7     0.8635538 0.5830128 0.9762732
 9     0.8794553 0.5704487 0.9803011
11     0.8803292 0.5627564 0.9807495
13     0.8805989 0.5601923 0.9807495
15     0.8822374 0.5576282 0.9816464
17     0.8822374 0.5576282 0.9816464
19     0.8822374 0.5576282 0.9816464
21     0.8822374 0.5576282 0.9816464
23     0.8822374 0.5576282 0.9816464
25     0.8822374 0.5576282 0.9816464
27     0.8822374 0.5576282 0.9816464
29     0.8822374 0.5576282 0.9816464
31     0.8822374 0.5576282 0.9816464
33     0.8822374 0.5576282 0.9816464

```

```

> predicted <- predict(model_kknn, testData4)
> head(predicted)
[1] BAD BAD BAD BAD BAD BAD
Levels: BAD GOOD
>
> # Compute the confusion matrix
> confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')
Confusion Matrix and Statistics

```

```

      Reference
Prediction BAD GOOD
BAD      67    16
GOOD     31   542

```

6. Sprendimų medžio (miško) modelis

```

> library(ranger)
warning message:
package 'ranger' was built under R version 3.6.1
> model_rf = train(HR_12 ~ ., data=trainData, method='ranger', tuneLength=15, trControl=ctrl)
warning message:
In train.default(x, y, weights = w, ...) :
  The metric "Accuracy" was not in the result set. ROC will be used instead.
> model_rf
Random Forest

2630 samples
 21 predictor
  2 classes: 'BAD', 'GOOD'

No pre-processing
Resampling: Cross-validated (10 fold)
Summary of sample sizes: 2367, 2367, 2367, 2367, 2367, 2368, ...
Resampling results across tuning parameters:

```

mtry	splitrule	ROC	Sens	Spec
2	gini	0.9462939	0.6364744	1.0000000
2	extratrees	0.9145100	0.3359615	0.9995516
3	gini	0.9488413	0.6819231	0.9986587
3	extratrees	0.9335753	0.3787821	0.9995516
4	gini	0.9505167	0.6895513	0.9973114
4	extratrees	0.9407174	0.4419231	0.9986567
6	gini	0.9496712	0.7021154	0.9932816
6	extratrees	0.9460386	0.6214744	0.9941784
7	gini	0.9474855	0.7071154	0.9923847
7	extratrees	0.9470532	0.6442308	0.9937300
8	gini	0.9462607	0.7146795	0.9910414
8	extratrees	0.9449916	0.6644231	0.9928331
10	gini	0.9445258	0.7096154	0.9901405
10	extratrees	0.9459673	0.6821154	0.9914898
11	gini	0.9441190	0.7096795	0.9883508
11	extratrees	0.9456236	0.6871795	0.9910414
12	gini	0.9431358	0.7147436	0.9879024
12	extratrees	0.9453764	0.6921154	0.9919363
14	gini	0.9432563	0.7071154	0.9865571
14	extratrees	0.9449344	0.6972436	0.9910414
15	gini	0.9411254	0.7222436	0.9874540
15	extratrees	0.9475248	0.7047436	0.9896961
16	gini	0.9418670	0.7171795	0.9865571
16	extratrees	0.9474382	0.6971795	0.9905930
18	gini	0.9408877	0.7272436	0.9865571
18	extratrees	0.9461315	0.7073718	0.9910394
19	gini	0.9418560	0.7272436	0.9861107
19	extratrees	0.9442956	0.7098718	0.9883508
21	gini	0.9394328	0.7247436	0.9865571
21	extratrees	0.9456298	0.7022436	0.9887992

```

> predicted <- predict(model_rf, testData4)
> head(predicted)
[1] BAD BAD BAD BAD BAD BAD
Levels: BAD GOOD
>
> # Compute the confusion matrix
> confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')
Confusion Matrix and Statistics

          Reference
Prediction BAD GOOD
BAD      74     3
GOOD    24   555

```

7. Visas Rstudio programos kodas

Load the caret package

```
library(caret)
```

```
library(readxl)
```

```
Data_vol2 <- read_excel("D:/Chrome downloads/Data_vol2.xlsx",
                        col_types = c("text", "numeric", "numeric",
                                      "numeric", "numeric", "text", "text",
                                      "text", "text", "numeric", "numeric",
                                      "numeric", "numeric", "numeric"))
```

```
View(Data_vol2)
```

```
credit <- Data_vol2
```

```
credit$HR_12 <- as.factor(credit$HR_12)
```

```
credit$GRT_MARITAL_STATUS <- as.factor(credit$GRT_MARITAL_STATUS)
```

```
credit$X_GRT_COUNTY_CODE <- as.factor(credit$X_GRT_COUNTY_CODE)
```

```
credit$SODRA_PROF_STATUS <- as.factor(credit$SODRA_PROF_STATUS)
```

```
credit$SEX <- as.factor(credit$SEX)
```

Structure of the dataframe

```
str(credit)
```

```
summary(credit)
```

Create the training and test datasets

```
set.seed(100)
```

Step 1: Get row numbers for the training data

```
trainRowNumbers <- createDataPartition(credit$HR_12, p=0.8, list=FALSE)
```

```

# Step 2: Create the training dataset
trainData <- credit[trainRowNumbers,]
summary(trainData)

# Step 3: Create the test dataset
testData <- credit[-trainRowNumbers,]
summary(testData)

# Store X and Y for later use.
x = trainData[, 2:14]
y = trainData$HR_12

library(skimr)
skimmed <- skim_to_wide(trainData)
skimmed[, c(1:14)]

# Create the knn imputation model on the training data
preProcess_missingdata_model <- preProcess(trainData, method='knnImpute')
preProcess_missingdata_model

# Use the imputation model to predict the values of missing data points
library(RANN) # required for knnImpute
trainData <- predict(preProcess_missingdata_model, newdata = trainData)
anyNA(trainData)
summary(trainData)

# One-Hot Encoding
# Creating dummy variables is converting a categorical variable to as many binary variables as here
are categories.
dummies_model <- dummyVars(HR_12 ~ ., data=trainData)

# Create the dummy variables using predict. The Y variable will not be present in trainData_mat.
trainData_mat <- predict(dummies_model, newdata = trainData)

```

```

## Convert to dataframe
trainData <- data.frame(trainData_mat)

## See the structure of the new dataset
str(trainData)
summary(trainData)

preProcess_range_model <- preProcess(trainData, method='range')
trainData <- predict(preProcess_range_model, newdata = trainData)

# Append the Y variable
trainData$HR_12 <- y

apply(trainData[, 1:21], 2, FUN=function(x){c('min'=min(x), 'max'=max(x))})
summary(trainData)

featurePlot(x = trainData[, 1:21],
            y = trainData$HR_12,
            plot = "box",
            strip=strip.custom(par.strip.text=list(cex=.7)),
            scales = list(x = list(relation="free"),
                          y = list(relation="free")))

featurePlot(x = trainData[, 1:21],
            y = trainData$HR_12,
            plot = "density",
            strip=strip.custom(par.strip.text=list(cex=.7)),
            scales = list(x = list(relation="free"),
                          y = list(relation="free")))

# Set the seed for reproducibility
set.seed(100)

# Step 1: Impute missing values

```

```

testData2 <- predict(preProcess_missingdata_model, testData)
str(testData2)
# Step 2: Create one-hot encodings (dummy variables)
testData3 <- predict(dummies_model, testData2)
str(testData3)
# Step 3: Transform the features to range between 0 and 1
testData3 <- data.frame(testData3)
testData4 <- predict(preProcess_range_model, testData3)

#testData4 <- apply(testData3[, 1:21], 2, FUN=function(x){c('min'=min(x), 'max'=max(x))})

# View
head(testData4[, 1:21])

#model_adaboost = train(HR_12 ~ ., data=trainData, method='adaboost', tuneLength=2)
#model_adaboost
#####
ctrl <- trainControl(method="cv",
                      summaryFunction=twoClassSummary,
                      classProbs=T,
                      savePredictions = T)
model_lda = train(HR_12 ~ ., data=trainData, method='lda', tuneLength=5, trControl=ctrl)
model_lda

predicted <- predict(model_lda, testData4)
head(predicted)

# Compute the confusion matrix
confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')

#variable importance
varimp_LDA <- varImp(model_lda)
plot(varimp_LDA, main="Variable Importance with LDA")

```



```

#####
library(LiblineR)
model_log = train(HR_12 ~ ., data=trainData, method='regLogistic', tuneLength=5, verbose=F,
trControl=ctrl)
model_log

predicted <- predict(model_log, testData4)
head(predicted)

# Compute the confusion matrix
confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')

varimp_log <- varImp(model_log)
plot(varimp_log, main="Variable Importance with LOGREG")
#####
numFolds <- trainControl(method = 'cv', number = 25, classProbs = TRUE, verboseIter = TRUE,
summaryFunction = twoClassSummary,
preProcOptions = list(thresh = 0.75, ICAcomp = 3, k = 5))
model_nnet <- train(HR_12 ~ ., data = trainData, method = 'nnet', preProcess = c('center', 'scale'),
tuneLength=5, trControl=ctrl,
tuneGrid=expand.grid(size=c(10), decay=c(0.1)))
model_nnet

predicted <- predict(model_nnet, testData4)
head(predicted)

# Compute the confusion matrix
confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')

varimp_nnet <- varImp(model_nnet)
plot(varimp_nnet, main="Variable Importance with NNET")
#####
#control <- trainControl(method="repeatedcv", number=10, repeats=5)
model_glm = train(HR_12~., data=trainData, method='glm', tuneLength=5, trControl=ctrl)

```

```

model_glm

predicted <- predict(model_glm, testData4)
head(predicted)

# Compute the confusion matrix
confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')

varimp_glm <- varImp(model_glm)
plot(varimp_glm, main="Variable Importance with GLM")
#####
library(igraph)
library(kknn)
model_kknn = train(HR_12 ~ ., data=trainData, method='kknn', tuneLength=15, trControl=ctrl)
model_kknn

predicted <- predict(model_kknn, testData4)
head(predicted)

# Compute the confusion matrix
confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')

plot(model_mars, main="Model Accuracies with kknn")

varimp_kknn <- varImp(model_kknn)
plot(varimp_kknn, main="Variable Importance with KKNN")
#####
library(ranger)
model_rf = train(HR_12 ~ ., data=trainData, method='ranger', tuneLength=15, trControl=ctrl)
model_rf

predicted <- predict(model_rf, testData4)
head(predicted)

```

```

# Compute the confusion matrix
confusionMatrix(reference = testData$HR_12, data = predicted, mode='everything', positive='BAD')

plot(model_rf, main="Model Accuracies with RF")

#Importance not available
varimp_rf <- varImp(model_rf)
plot(varimp_rf, main="Variable Importance with RF")
#####
# Compare model performances using resample()
models_compare <- resamples(list(LDA=model_lda, LOG=model_log,
                                NEURALNET=model_nnet, GLM=model_glm,
                                KNN = model_kknn, RF=model_rf
                                ))

# Summary of the models performances
summary(models_compare)

# Draw box plots to compare models
scales <- list(x=list(relation="free"), y=list(relation="free"))
bwplot(models_compare, scales=scales)

```