VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
VERSLO VADYBOS FAKULTETAS
FINANSŲ INŽINERIJOS KATEDRA

Akvilė Ereminė

FINANSINIO SUKČIAVIMO ATPAŽINIMO TYRIMAS NAUDOJANT SKAITMENINĮ INTELEKTĄ

INVESTIGATION OF FINANCIAL FRAUD DETECTION BY USING COMPUTATIONAL INTELLIGENCE

Baigiamasis magistro darbas

Finansų inžinerijos studijų programa, valstybinis kodas 6211LX060
Finansinių technologijų (FinTech) specializacija
Finansų studijų kryptis

Vilnius, 2023

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

VERSLO VADYBOS FAKULTETAS

FINANSŲ INŽINERIJOS KATEDRA

TVIRTINU

Katedros vedėja

_____
(Parašas)

doc. dr. Algita Miečinskienė
(Vardas, pavardė)

_____
(Data)

Akvilė Ereminė

FINANSINIO SUKČIAVIMO ATPAŽINIMO TYRIMAS NAUDOJANT SKAITMENINĮ INTELEKTĄ

INVESTIGATION OF FINANCIAL FRAUD DETECTION BY USING COMPUTATIONAL INTELLIGENCE

Baigiamasis magistro darbas

Finansų inžinerijos studijų programa, valstybinis kodas 6211LX060

Finansinių technologijų (FinTech) specializacija

Finansų studijų kryptis

Vadovė doc. dr. Nijolė Maknickienė _____ _____
(Moksl. laipsnis/pedag. vardas, vardas, pavardė) (Parašas) (Data)

Konsultantas _____ _____ _____
(Moksl. laipsnis/pedag. vardas, vardas, pavardė) (Parašas) (Data)

Konsultantas _____ _____ _____
(Moksl. laipsnis/pedag. vardas, vardas, pavardė) (Parašas) (Data)

Lietuvių kalbos konsultantas_____ _____ _____
(Moksl. laipsnis/pedag. vardas, vardas, pavardė) (Parašas) (Data)

Vilnius, 2023

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
VERSLO VADYBOS FAKULTETAS
FINANSŲ INŽINERIJOS KATEDRA

Studijų kryptis: Finansai
Studijų programa: Finansų inžinerija , valstybinis kodas 6211LX060
Specializacija: Finansinės technologijos (FinTech)

TVIRTINU
Katedros vedėjas
Algita Miečinskienė
2023-12-15

MAGISTRO BAIGIAMOJO DARBO UŽDUOTIS

Nr. FTfm-22–5522

Vilnius

Studentas (-ė): Akvilė Ereminė

Baigiamojo darbo tema: Finansinio sukčiavimo atpažinimo tyrimas naudojant skaitmeninį intelektą

Baigiamojo darbo užbaigimo terminas pagal numatytą studijų kalendorinį grafiką.

BAIGIAMOJO DARBO UŽDUOTIS:

Baigiamojo magistro darbo uždaviniai:
1. Atlikti mokslinės literatūros analizę apie finansinio sukčiavimo reiškinį, jo būdus ir tikrąją kainą ekonomikai.
2. Apibrėžti skaitmeninio intelekto sąvoką ir palyginti skaitmeninio intelekto metodus bei pateikti jo galimas naudas, rizikas ir įtaką taikant jį finansinio sukčiavimo atpažinimo įrankiuose.
3. Išnagrinėti pasirinktų metodų taikymą finansinio sukčiavimo atpažinimo problemai spręsti.
4. Atlikti finansinio sukčiavimo atpažinimo tyrimą remiantis pasirinktais metodais ir palyginti gautų rezultatų finansinę naudą juos pritaikius finansinių paslaugų teikimo įmonės veikloje.

Vadovas docentas Nijolė Maknickienė

Vilniaus Gedimino technikos universitetas
Verslo vadybos fakultetas
Finansų inžinerijos katedra

ISBN | ISSN
Egz. sk. .....
Data .....-.....-.....

Antrosios pakopos studijų **Finansų inžinerijos** programos magistro baigiamasis darbas

Pavadinimas	**Finansinio sukčiavimo atpažinimo tyrimas naudojant skaitmeninį intelektą**
Autorius	**Akvilė Ereminė**
Vadovas	**Nijolė Maknickienė**

Kalba: lietuvių

Anotacija

Proaktyvus finansinio sukčiavimo atpažinimas vis dar išlieka viena aktualiausių temų globaliu mastu įmonėms teikiančioms finansines paslaugas, nes sėkmingų atvejų rodiklis, kai identifikuojamas finansinis sukčiavimas išlieka labai žemas. Siekdami kovoti su finansinio sukčiavimo problema bankai ir kitos finansų sektoriuje veikiančios įmonės yra įpareigotos stebėti, įvertinti ir praneši apie įtariamą nelegalią veiklą. Didelis duomenų kiekis, susijęs su finansinių paslaugų teikimo reikalavimų laikymusi, ir didėjantis nusikalstamų metodų sudėtingumas verčia finansų įstaigas nuolat ieškoti naujų priemonių, kurios padėtų įvykdyti reguliavimo įsipareigojimus. Taip pat svarbu paminėti, jog visuotinai sutariama, kad tradicinių statistikos metodų nebeužtenka ir dirbtinio intelekto pagrindu sukurti sprendimai yra būdas laiku aptikti ir sustabdyti finansinio sukčiavimo atvejį. Darbo tikslas – palyginti skaitmeninio intelekto metodų teikiamą finansinę naudą atpažįstant finansinį sukčiavimą. Šiam tikslui pasiekti buvo lyginami šie skaitmeninio intelekto metodai: Neuroniniai tinklai, Bajeso tinklai, Sprendimo medis, XGboost, Catboost, nes jie sugeba apdoroti didžiulius duomenų kiekius, padeda aptikti netipinę finansinę elgseną. Atlikus metodų vertinimo tyrimą, rezultatai parodė, kad metodai, kurie veikia sprendimo medžio veikimo logika davė geresnius rezultatus nei neuroninio tinklo veikimu paremti metodai. Taip pat didžiausia finansinė nauda įmonei būtų pritaikius savo rizikos valdymo sistemoje sprendimo medžio metodą. Šio tyrimo darbo apribojimai yra duomenys, adaptavimas kintančioje aplinkoje, techniniai resursai. Šis tyrimas galėtų būti naudingas įmonėms, kurios kuria sistemas skirtas pinigų plovimo ir teroristų finansavimo prevencijos sustiprinimui ar finansinių operacijų stebėsenos įrankius. Tęsiant šį tyrimą būtų naudinga pagalvoti apie galimybę įtraukti į duomenų rinkinį įvairesnius duomenų šaltinius tokius kaip pažink savo vartotoją (angl. Know your customer), tai leistų jau ankstesniame žingsnyje geriau identifikuoti kliento rizikos lygį.

Prasminiai žodžiai: Finansinis sukčiavimas, skaitmeninis intelektas, neuroniniai tinklai, Bajeso tinklai, sprendimo medis, XGboost, Catboost.

Vilnius Gediminas Technical University
Faculty of Business Management
Department of Financial Engineering

ISBN | ISSN
Copies No. .....
Date .....-.....-.....

Master Degree Studies **Financial Engineering** study programme Master Graduation Thesis

Title	**Investigation of Financial Fraud Detection by Using Computational Intelligence**
Author	**Akvilė Ereminė**
Academic supervisor	**Nijolė Maknickienė**

Thesis language: Lithuanian

Annotation

Proactive identification of financial fraud is still one of the hot topics for financial services companies globally, as the result of low success rate of identifying financial fraud. In order to combat the problem of financial fraud, banks and other companies operating in the financial sector are obliged to monitor, evaluate and report suspected illegal activities. The large amount of data related to financial services compliance and the increasing sophistication of criminal methods are forcing financial institutions to constantly look for new tools to help them meet their regulatory obligations. It is also important to mention that there is a general consensus that traditional statistical methods are no longer sufficient and artificial intelligence-based solutions are a way detect and stop financial fraud on time. The aim of the work is to compare the financial benefits of computational intelligence methods in identifying financial fraud. To achieve this goal, the following computational intelligence methods were compared: neural networks, Bayesian networks, decision tree, XGboost, Catboost. After conducting a method evaluation study, the results showed that the methods based on decision tree operation logic gave better results than the methods based on neural network operation. Also, the biggest financial benefit for the company would be to apply the decision tree method in its risk management system. The limitations of this research work are data, adaptation in a changing environment, technical resources. This research could be useful for companies that develop systems to strengthen the prevention of money laundering and terrorist financing or tools for monitoring financial transactions. In the continuation of this research, it would be useful to think about the possibility of including more diverse data sources in the data set, such as Know your customer, which would allow to better identify whether the customer is risky or not in the previous step.

Keywords: Financial fraud, computational intelligence Neural networks, Bayesian networks, Decision tree, XGboost, Catboost.

SUTIKIMAS DĖL ASMENS DUOMENŲ NAUDOJIMO

2023-12-21
(Data)

Šiuo sutikimu aš, Akvilė Ereminė (toliau – Duomenų subjektas) sutinku, kad Vilniaus Gedimino technikos universitetas, juridinio asmens kodas 111950243, adresas Saulėtekio al. 11, LT-10223 Vilnius (toliau – Duomenų valdytojas), tvarkytų mano asmens duomenis kitų studentų mokymosi tikslu. T. y. tvarkytų (*pažymėkite tinkamą*):

☐ vardą, pavardę, bakalauro baigiamąjį darbą;
☐ bakalauro baigiamąjį darbą, nenurodant vardo, pavardės;
☐ vardą, pavardę, magistro baigiamąjį darbą; ×
☒ magistro baigiamąjį darbą nenurodant vardo, pavardės.

Šiuo tikslu tvarkomų asmens duomenų Duomenų valdytojas neperduos jokiems tretiesiems asmenims, studentams su baigiamaisiais darbais bus leidžiama susipažinti vidinėje informacinėje sistemoje. Duomenų subjekto asmens duomenys šiuo tikslu bus naudojami ne ilgiau nei 5 metai.

Šiuo sutikimu Duomenų subjektas patvirtina, kad yra supažindintas su šiomis teisėmis:

- Susipažinti su savo duomenimis ir kaip jie yra tvarkomi (teisė susipažinti);
- Reikalauti ištaisyti arba, atsižvelgiant į asmens duomenų tvarkymo tikslus papildyti asmens neišsamius asmens duomenis (teisė ištaisyti);
- Savo duomenis sunaikinti arba sustabdyti savo duomenų tvarkymo veiksmus (išskyrus saugojimą) (teisė sunaikinti ir teisė „būti pamirštam“);
- Reikalauti, kad asmens duomenų valdytojas apribotų asmens duomenų tvarkymą (teisė apriboti);
- Teise į duomenų perkėlimą (teisė perkelti);
- Nesutikti, kad būtų tvarkomi asmens duomenys, kai šie duomenys tvarkomi ar ketinami tvarkyti kitais tikslais;
- Pateikti skundą Valstybinei duomenų apsaugos inspekcijai;

Duomenų subjektas turi teisę bet kuriuo metu atšaukti savo sutikimą.

Akvilė Ereminė
[Duomenų subjekto vardas, pavardė, parašas]

Duomenų valdytojo rekvizitai:
Vilniaus Gedimino technikos universitetas
Juridinio asmens kodas: 111950243
Adresas: Saulėtekio al. 11, LT-10223 Vilnius
Tel. (8 5) 274 5030
Faks. (8 5) 270 0112
El. paštas: vilniustech@vilniustech.lt
PVM mokėtojo kodas: LT119502413
Duomenų apsaugos pareigūno tel. (8 5) 251 2191, el. paštas: dap@vilniustech.lt

TURINYS

LENTELIŲ SĄRAŠAS

PAVEIKSLŲ SĄRAŠAS

PRIEDŲ SĄRAŠAS

ĮVADAS

Tyrimo aktualumas. Proaktyvus finansinio sukčiavimo atpažinimas vis dar išlieka viena aktualiausių temų globaliu mastu įmonėms, vykdančioms finansines operacijas, įskaitant bankus, finansinių technologijų įmones, kriptovaliutų prekybos įmones ir kitoms šiame sektoriuje veikiančioms įmonėms. Ypač pastarieji ketveri metai buvo sudėtingi: „Brexit“, COVID-19 pandemija, ekonomikos nuosmukis, precedento neturinti infliacija ir karas Ukrainoje. Kiekvienas iš šių įvykių atvėrė kelią nusikalstamos veiklos pagausėjimui, tai pagrindžia faktas, jog finansinis sukčiavimas 2023 m. pirmame ketvirtyje pasiekė rekordinę sumą - 3,7 trilijoną dolerių (Finteth Global, 2023).

Siekdami kovoti su finansinio sukčiavimo problema bankai ir kitos finansų sektoriuje veikiančios įmonės yra įpareigotos stebėti, įvertinti ir praneši apie įtariamą nelegalią veiklą. Didelis duomenų kiekis, susijęs su finansinių paslaugų teikimo reikalavimų laikymusi, ir didėjantis nusikalstamų metodų sudėtingumas verčia finansų įstaigas nuolat ieškoti naujų priemonių, kurios padėtų įvykdyti reguliavimo įsipareigojimus. Siekiant šio tikslo technologijos, kurių veikimas paremtas dirbtiniu intelektu, tampa neatsiejamos nuo proaktyvaus finansinio sukčiavimo atpažinimo ir jos prevencijos. Prognozuojama, jog pasaulinė kovos su finansinių nusikaltimo atpažinimo įrankių rinka nuo 2020 iki 2025 m. augs 15,6 proc. (Anti-money Laundering Market, 2022).

Tyrimo problema – kaip įgyti skaitmeninį pranašumą prieš nusikalstamą finansinę veiklą vykdančius asmenis?

Tyrimo objektas – skaitmeninio intelekto metodai, kurie gali padėti atpažinti finansinį sukčiavimą.

Tyrimo tikslas – palyginti skaitmeninio intelekto metodų teikiamą finansinę naudą atpažįstant finansinį sukčiavimą.

Uždaviniai tikslui pasiekti:

1. Atlikti mokslinės literatūros analizę apie finansinio sukčiavimo reiškinį, jo būdus ir tikrąją kainą ekonomikai.
2. Apibrėžti skaitmeninio intelekto sąvoką ir palyginti skaitmeninio intelekto metodus bei pateikti jo galimas naudas, rizikas ir įtaką taikant jį finansinio sukčiavimo atpažinimo įrankiuose.
3. Išnagrinėti pasirinktų metodų taikymą finansinio sukčiavimo atpažinimo problemai spręsti.

4. Atlikti finansinio sukčiavimo atpažinimo tyrimą remiantis pasirinktais metodais ir palyginti gautų rezultatų finansinę naudą juos pritaikius finansinių paslaugų teikimo įmonės veikloje.

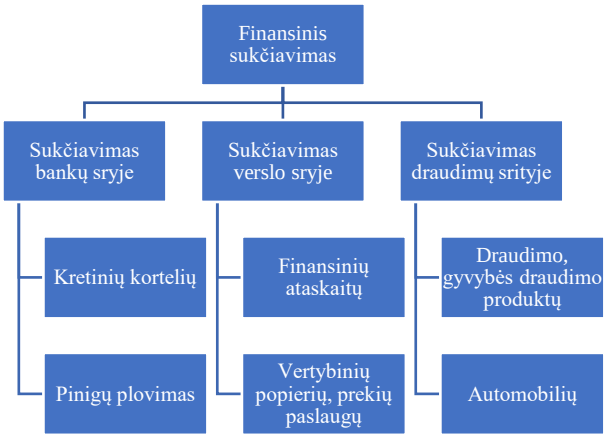
Tyrimo metodai: lietuvių ir užsienio mokslinės literatūros šaltinių analizė. Naudojant „Python“ programą analizuojami duomenys, ieškoma anomalijų, lyginami finansinio sukčiavimo atpažinimo atvejams taikomi skaitmeninio intelekto metodai: neuroninių tinklų, bajeso tinklo, sprendimo medžio, „XGBoost“ ir „CatBoost“ pasirinkti algoritmai.

1. FINANSINIO SUKČIAVIMO ATPAŽINIMO TEORINIAI ASPEKTAI

1.1. Finansinio sukčiavimo samprata

Finansinis sukčiavimas – tai veikla, kai asmuo ar organizacija sąmoningai apgaudinėja kitus, siekdami neteisėtos finansinės naudos (Baesens et al., 2015). Šis nusikaltimas padaro žalą asmenims, bendrovėms ir net visai finansų sistemai.

Mokslinėje literatūroje finansinis sukčiavimas suskirtomas į šiuos tipus (žr. 1 pav.) (Teichmann, 2019, Jain, Shinde, S., 2019; Gupka, 2021):



1 pav. Sukčiavimo tipai

Šaltinis: Teichmann (2019), Jain, Shinde (2019), Gupka (2021)

Finansinis sukčiavimas gali būti vykdomas įvairiomis formomis, ir visada jis yra skirtas asmeninei finansinei naudai gauti, naudojant sukčiavimą, apgaulę ar kitus neteisėtus metodus.

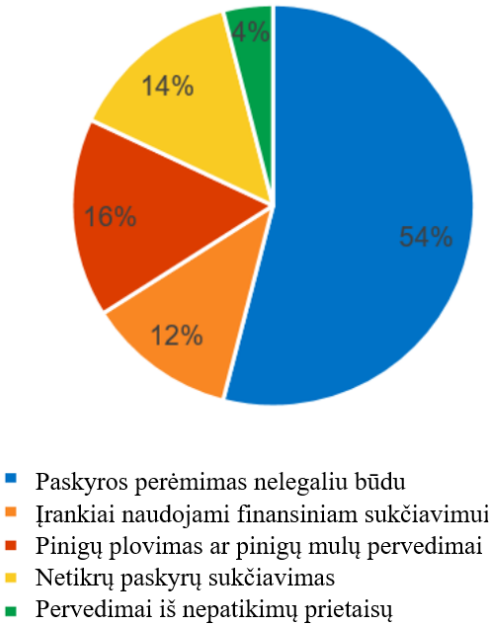
Mokslinėje literatūroje dažniausiai pateikiamos šios nusikalstamos finansinės veikos formos (Baesens, 2017; Teichmann, 2019; Gupka, 2021; Ch, Gadekallu, Abidi 2021):

- **Sukčiavimas**, tai veiksmas, kai asmuo apgaudinėja kitus, siekdamas gauti neteisėtą finansinę naudą. Tai gali apimti melagingus investicinius planus, netikrų produktų ar paslaugų pardavimą, draudimo sukčiavimą.
- **Pinigų plovimas**, tai veiksmas, kai asmuo ar organizacija bando legalizuoti neteisėtai įgytus pinigus ar turtą. Tai dažnai susiję su finansinėmis schemomis, kuriose pinigai perkeliami per įvairias sąskaitas ar transakcijas, siekiant išvengti jų kilmės atsekimo.
- **Prekyba vertybiniais popieriais, pasinaudojant viešai neatskleista informacija**, tai veiksmas, kai asmuo naudoja konfidencialią informaciją apie įmonės ar rinkos veiklą, kad gautų neteisėtą pelną iš akcijų ar

kitų finansinių priemonių prekybos. Tokios veiklos metu asmuo naudoja viešai neprieinamą informaciją, kurią įgyja dėl savo pareigų ar ryšių su įmone, ir išnaudoja šią informaciją finansiniam pranašumui.

- **Klastojimas**, tai veiksmas, kai asmuo klastoja oficialius dokumentus – pasą, banko sąskaitos išrašus arba finansinius dokumentus, siekdamas apgauti kitus.
- **Internetiniai sukčiavimai**, tai veiksmas, kai asmuo naudoja internetą ir elektronines komunikacijas siekdamas apgauti ir neteisėtais būdais gauti finansinės naudos.

„Kaspersky“[1] (2020) pateikia (žr. 2 pav.) kiek kitokius nelegalios veiklos tipus. Toliau pateiktame paveiksle matome, jog daugiau nei pusę finansinio sukčiavimo yra atliekama įsilaužiant į paskyras.



2 pav. Nelegalios veiklos tipų pasiskirstymas

Šaltinis: Kaspersky (2020)

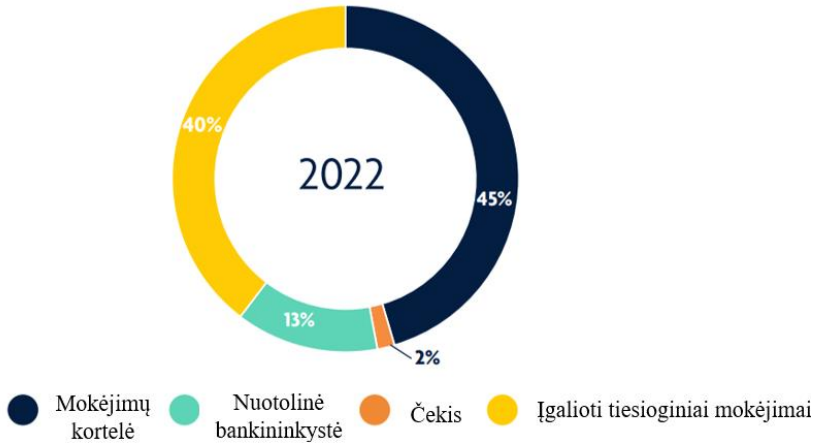
Analizuojant 2020 m. sausio – gruodžio mėn. duomenis, tokių incidentų dalis padidėjo nuo 34 proc. 2019 m. iki 54 proc. 2020 m. (Kasperovski, 2020). Atlikus analizę pasitvirtino, kad vis dar nusikaltėlių dažniausiai naudojami du bendri būdai gauti prieigą prie paskyrų. Pirmoji taktika rodo, kad sukčiai apsimeta „gelbėtojais“ ir saugumo ekspertais ir įgyvendina scenarijus, kad „išgelbėtų“ vartotojus. Jie skambina apsimetę apsaugos pareigūnais banko klientams ir praneša apie įtartinus mokesčius ar mokėjimus bei siūlo savo pagalbą. Antrasis

[1] „Kaspersky Lab“ yra Jungtinėje Karalystėje registruota bendrovė, kurios padaliniai veikia 30 šalių visame pasaulyje. „Kaspersky Lab“ produktus ir technologijas naudoja daugiau nei 300 mln. galutinių vartotojų visame pasaulyje ir daugiau nei 250 tūkst. įmonių.

pavyzdys – kibernetiniai nusikaltėliai veikia kaip „investuotojai“. Šis scenarijus apima sukčius, apsimetančius investicinės bendrovės darbuotojais arba banko investicijų konsultantais. Jie skambina klientams, siūlo greitą būdą užsidirbti pinigų investuojant į kriptovaliutą ar akcijas tiesiai iš kliento sąskaitos, nesikreipiant į banko skyrių.

Kitą dalį sudaro pinigų plovimas, fiktyvios paskyros, neteisėti sukčiavimo būdai.

Didžiosios Britanijos metinėje finansinio nusikaltimo apžvalgoje 2023 m. įvardyta, kad su mokėjimo kortelėmis susiję finansų sukčiavimai sudaro 45 proc. Kitą didelę finansinio sukčiavimo dalį sudaro įgalioti tiesioginiai mokėjimai 40 proc (žr. 3 pav.).



3 pav. Finansinio nuostolio pasiskirstymas pagal finansinio sukčiavimo tipus

Šaltinis: UK finance. ANNUAL FRAUD REPORT The definitive overview of payment industry fraud in 2022 (2020)

Apibendrinant pateiktą informaciją mokslinėje literatūroje ir oficialiuose finansinių sukčiavimo metinėse ataskaitose galima teigti, kad finansinis sukčiavimas pasireiškia įvairiomis formomis. Taip pat pastebima, kad finansiniai sukčiai puikiai prisitaikė tiek prie finansinių įmonių siūlomų naujų finansinių produktų vis didėjančios įvairovės, tiek prie pasikeitusio vartotojų pirkimo įpročių, kai vis daugiau atskaitoma internetu, išmaniuoju telefonu.

1.2. Finansinio sukčiavimo mastas ir jo kaina

Šiame skyriuje analizuojant pasirinktą mokslinę literatūrą, mokslinius žurnalus, statistinę analizę bei kitus internetinius šaltinius – aptariama tikroji finansinio sukčiavimo kaina.

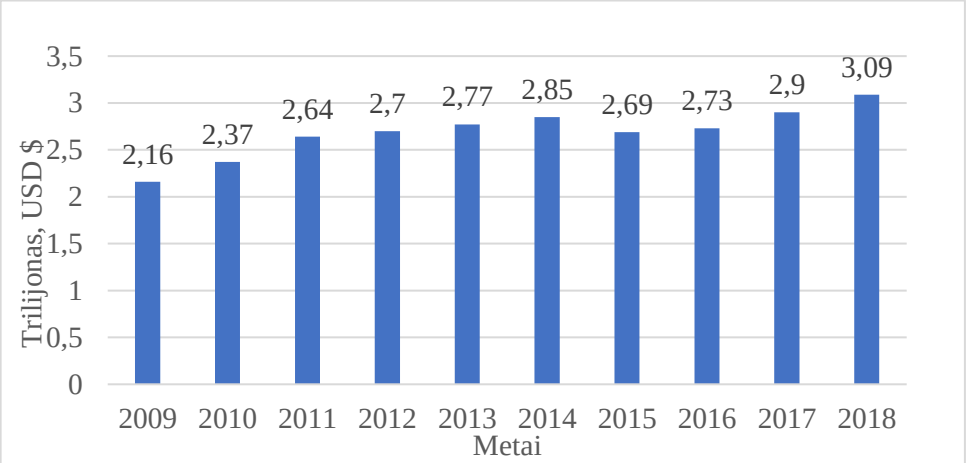
Analizuojant šią temą giliau norisi suprast šios nelegalios veiklos tikrąjį mastą. Tačiau tikrojo finansinio sukčiavimo masto ir tikrosios žalos (išlaidų ekonomikai) nustatyti neįmanoma, dėl kelių priežasčių (Pol, 2020):

Nuolat kintančių nusikaltėlių vykdomų schemų, kurias nusikaltėliai naudoja

siekdami apeiti esamus reguliavimo mechanizmus ir atitikties procesus. Išaugusio duomenų srauto finansinių institucijų paslaugoms persikėlus į skaitmeninę erdvę, dėl šios situacijos finansinės įstaigos priverstos ieškoti naujų technologijų būdų apdorojant duomenis.Reguliavimo ir nuolatinio lenktyniavimo finansų institucijų tarpusavyje siekiant sukurti reguliavimo ir atitikties metodus, kurių taikymo sritis būtų plati, pakankamai naujoviška ir veiksmingi, kad būtų galima stebėti ir prireikus laiku sustabdyti neteisėtą veiklą.

Šiais metais „Finteth Global“ duomenimis, 2023 m. pirmame ketvirtyje finansinis sukčiavimas sudarė 3,7 trilijonų USD dolerių. Kiekvienas sukčiavimo doleris JAV finansinėms tarnyboms kainuoja 4,23 USD dolerių. Į šiuos mokesčius įeina teisinės, tyrimo ir išieškojimo išlaidos (Kadar, 2023). Taip pat įdomi tendencija pastebima, jog ypač išaugo mobiliosios bankininkystės sukčiavimo dalis tarp visų sukčiavimo išlaidų (Kroll, 2023).

Iki šios dienos finansinis sukčiavimas vis dar auga. Žemiau pateiktame 4 paveiksle (žr. 4 pav.) šis faktas patvirtinamas. Tiriamu periodu nuo 2009 iki 2018 m. pajamos iš nusikalstamos veiklos nuosekliai auga, o tai reiškia, kad problema yra įsisenėjusi ir blogėjanti.



4 pav. Jungtinių Tautų įvertintos pajamos globaliu mastu iš nusikalstamos veiklos 3,6 % BVP

Šaltinis: Pol (2020)

Taip pat svarbu suprasti, kiek dar papildomai išlaidų sugeneruoja finansinio sukčiavimo aptikimas. Iš žemiau pateiktos lentelės (žr. 1 lentelė) matome, jog lėšos, kurios buvo fiksuotos kaip iš nelegalios veiklos ir konfiskuotos lėšos iš nelegalios veiklos skiriasi šimtus ir net tūkstančių kartų ir faktas yra tai, kad finansinio sukčiavimo nesuvaldome. Taip pat finansinio sukčiavimo atpažinimui išleidžiama milžiniškas sumas. Svarbu atkreipti dėmesį, jog kovos su pinigų plovimu sėkmės rodiklis yra labai žemas t. y. Europoje siekia 0,55 proc., o bendrai pasaulyje tik 0,05 proc., kai tuo tarpu sustabdyti šiai nelegalei veiklai yra išleidžiami milžiniški pinigai.

1 lentelė. Finansinio sukčiavimo prevencijos politikos patiriamos išlaidos

	Europa (Eur)	**Pasaulinis (USD)**
Kasmet sugeneruojamos lėšos iš nusikalstamos veiklos	110B	3T
Kasmet konfiskuojamos neteisėtos lėšos / turtas	1,2B	3B
Konfiskuotų neteisėjų lėšų arba turto dalis iš pinigų plovimo	50 proc.	50 proc.
Kovos su pinigų plovimu sėkmės rodiklis	0,55 proc.	0,05 proc.
Atitikties išlaidos	144B	304B
Atitikties išlaidų tenkančių pinigų plovimui	50 proc.	50 proc.

Šaltinis: Pol (2020)

Prognozuojama, jog pasaulinė kovos su finansinių nusikaltimo atpažinimo įrankių rinka nuo 2020 m. iki 2025 m. augs 15,6 proc., tai reiškia, kad nuo 2,2 mlrd. USD 2020 m. iki 2025 m. rinkos vertė 4,5 mlrd. USD (Anti-money Laundering Market, 2022). Šis faktas rodo, kad globaliu mastu suprantama, jog be naujausių technologijų finansinį sukčiavimą pažaboti neįmanoma.

Taip pat reikia pridurti, kad institucijos, įmonės deda dideles pastangas ir Didžiosios Britanijos finansinio sukčiavimo metinės ataskaitos duomenimis (žr. 2 lentelė) 2022 m. finansinio sukčiavimo fiksuotų atvejų skaičius drastiškai krito, palyginus su 2021 m. Finansinio sukčiavimo internetinio banko paslaugų skaičius nuosekliai augo nuo 2013 m. ir didžiausias fiksuotas 2021 m. Kai tuo tarpu 2022 m. matomas šioje srityje drastiškas atvejų sumažėjimas.

2 lentelė. Fiksuoti sukčiavimo atvejai

Bankininkystės kanalai	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Internetinė bankininkystė	13799	16041	19691	20088	21745	20904	25849	55995	72557	32036
Telefonijos bankininkystė	5596	5778	11380	10495	9577	7937	11199	7490	4623	3076
Mobilioji bankininkystė	N/A	N/A	2235	2809	3424	2956	6872	10155	11270	12361
Iš viso	19395	21819	33306	33392	34746	31797	43920	73640	88450	47473

Šaltinis: UK finance. ANNUAL FRAUD REPORT The definitive overview of payment industry fraud in 2022 (2022)

Kalbant apie banko paslaugų teikimą telefonu ir šioje srityje fiksuotus finansinio sukčiavimo atvejus matome, jog atvejų skaičius yra svyruojanti ir nuo 2019 m. sukčiavimo atvejų mažėja. Tačiau, reikia suprasti, kad būtent ši sritis bendrai bankinių paslaugų apimtyje

yra mažėjanti. Atvirkščiai nei mobilioji bankininkystė, kuri ypač išaugo COVID- 19 laikotarpiu, kai didžioji dalis pirkinių ir paslaugų atsiskaitymo perėjo į elektroninę erdvę.

Svarbus faktas tai, kad šis bendras finansinių sukčiavimo atvejų kiekio mažėjimas nėra atsitiktinis. Didžiosios Britanijos finansinio sukčiavimo metinėje ataskaitoje, kurioje buvo klausiama didžiausių šios valstybės bankus, finansinių technologijų įmonių, kriptovaliutų prekybos įmonių ir kitų šiame sektoriuje veikiančių vadovų apie, tai ar investuojama į naujausias technologijas ir net 88 proc. atsakė, jog įmonės investuoja į naujausias technologijas (UK finance. ANNUAL FRAUD REPORT The definitive overview of payment industry fraud in 2022). Taip pat labai panašūs atsakymai buvo gauti ir globaliu mastu atliktoje Kroll Fraud and Financial Crime Report sukčiavimo ir finansinių nusikaltimų ataskaitoje, kurioje dalyvavo 400 darbuotojai užimantys vadovaujančias pareigas sukčiavimo ir rizikos operacijų, pinigų plovimo, sukčiavimo strategijos, sukčiavimo analizės, technologijų ir duomenų srityse (Kroll, 2023). Apklausos duomenimis beveik vienbalsiai paklausus, kas apklaustųjų manymu gali sustabdyti finansinį sukčiavimą buvo įvardinta naujausios technologijos, kurios gali apdoroti didelius duomenis bei galinčios mokintis.

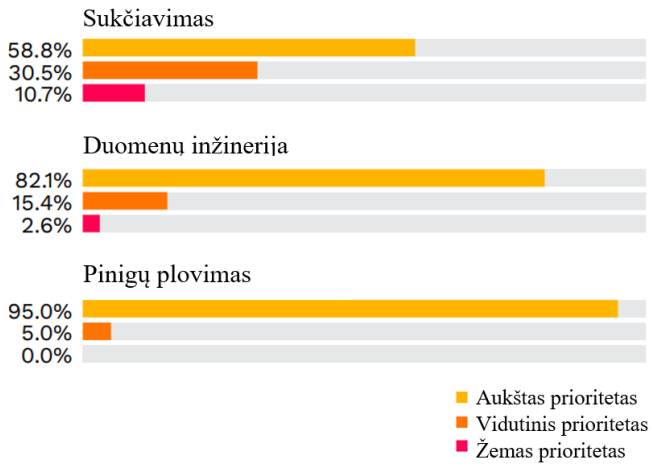
Apibendrinant, finansinio sukčiavimo mastas yra didžiulis, išleidžiamos milžiniškos sumos kovoti su šia nelegalia veiklos ir visgi sėkmingų atvejų, kai identifikuojamas finansinis sukčiavimas yra labai žemas. Tačiau, situaciją bandoma taisyti su naujausių technologijų pagalba. Remiantis tarptautinėmis ataskaitomis, kurių metu apklausiami šios srities ekspertai, vadovaujančias pareigas einantys asmenys, visi bendrai sutaria, kad šią nelegalią veiklą gali padėti sustabdyti naujausios technologijos, paremtos dirbtiniu intelektu.

1.3. Naujausių technologijų pritaikymo svarba siekant atpažinti finansinį sukčiavimą

Šiame skyriuje analizuojant pasirinktą mokslinę literatūrą, mokslinius žurnalus, statistinę analizę bei kitus internetinius šaltinius – aptariama dirbtinio intelekto įrankių pritaikymo svarba ieškant galimybių atpažinti finansinį sukčiavimą.

Jungtinėse Amerikos Valstijose atliktame tyrime (The State Of Fraud And Financial Crime In The U.S., 2022) dalyvavo 200 finansinių institucijų darbuotojai, kurių turtas sudarė ne mažiau kaip 5 mlrd. dolerių. Apklausti darbuotojai užėmė vadovaujančias pareigas sukčiavimo ir rizikos operacijų, pinigų plovimo, sukčiavimo strategijos, sukčiavimo analizės, technologijų ir duomenų srityse. 177 iš apklausos respondentų buvo atsakingi už sukčiavimo atpažinimo ir prevencijos valdymą iš kurių 20 buvo atsakingi už pinigų plovimo prevencijos valdymą. Kiti respondentai buvo atsakingi už duomenų mokslą ir technologijas.

Tyrimo dalyvių buvo klausiama įvardinti prioritetus pagal skirtingas sritis, kuriose vadovaujančias pareigas užimantys darbuotojai teikia naujų sprendimų kūrimui arba aptikimo tobulinimui ir prevencijos sistemos, skirtoms kovai su sukčiavimu ir finansiniais nusikaltimais pagal specializacijos sritis (žr. 7 pav.).



5 pav. Naujovių ir tobulėjimo svarba pagal respondentų veiklos sritis

Šaltinis: The State Of Fraud And Financial Crime In The U.S. (2022)

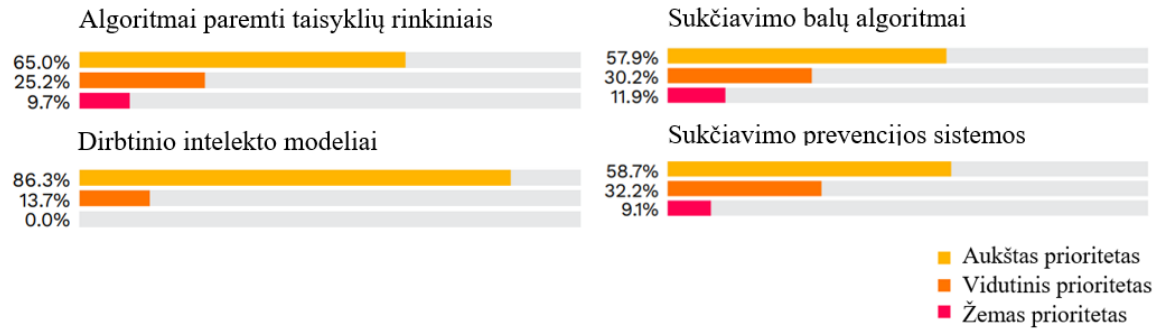
Tyrimo duomenys atskleidžia, kad vadovaujančias pareigas einantys darbuotojai pinigų plovimo prevencijos srityje beveik vienbalsiai t. y. 95 proc. mano, jog naujausių technologijų diegimas finansinio sukčiavimo atpažinimo tyrime yra aukščiausias prioritetas. Taip pat šios nuomonės laikosi vadovai dirbantys finansinio sukčiavimo srityje ir net 82 proc. mano, jog naujausios technologijos ir inovacijos yra taip pat viena iš galimybių, galinčių padėti kovoti su finansiniais nusikaltimais. Tuo tarpu duomenų mokslo vadovų tik daugiau nei 59 proc. mano, jog tai prioritetinė sritis.

Įmonės jau taikančios modernias technologijas buvo linkusios teikti ir toliau pirmenybę integracijai šiuolaikinėms pinigų plovimo (toliau – AML) bei kovos su sukčiavimo technologijomis. Įmonės, kurios nebuvo integravusios naujų kovos su sukčiavimu ir AML sprendimų, manė, jog naujausių technologijų įdiegimas yra sudėtingas procesas, kuris neužtikrins finansinio sukčiavimo atpažinimo mažėjimą. Apklausti vadovai, kurie nepriėmė šiuolaikinių AML ir kovos su sukčiavimu strategijų, tai padarė ne todėl, kad nebuvo technologinių sprendimus, kuriuos būtų galima integruoti su esamomis technologijomis, bet todėl, kad buvo įsitikinę, kad tai padaryti bus per sunku.

Taip pat atlikta apklausa nustatė, kad 66 proc. vadovų išreiškė susirūpinimą, kad reguliavimo standartai yra per sudėtingi ir jų sprendimas gali neatitikti visų jų atitikties poreikių.

Bendrai 58 proc. apklaustųjų manė, kad šiuolaikinės sukčiavimo schemos buvo pernelyg sudėtingos ir joks sprendimas negalėjo būti veiksmingas.

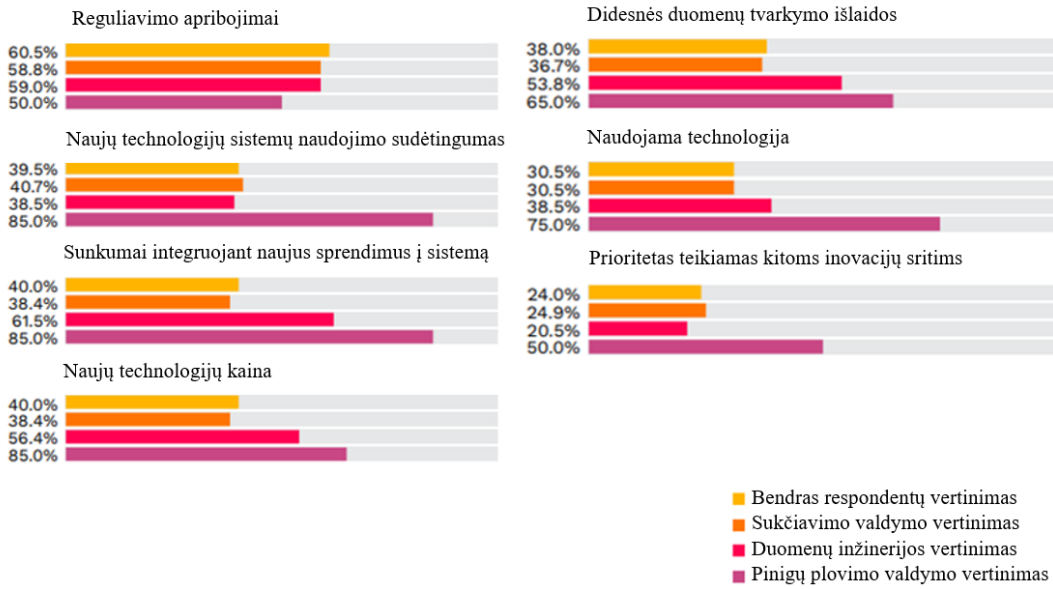
Kalbant apie technologijas, kurioms aukščiausio lygio vadovai teikia prioritetą ieškodami alternatyvų ar jau taiko veikloje (žr. 6 pav.) 86 proc. įmonių, naudojančių mašininį mokymąsi arba dirbtinio intelekto įrankius, mano jog tai prioritetinė sritis bei teigia, jog jų taikymas duoda teigiamų rezultatų kovojant su sukčiavimu. Taip pat 78 proc. apklaustųjų mano, jog debesijos pagrindu veikiančios sukčiavimo platformos lengvina jų kiekvienos dienos veiklą dėl integruotų išmanių vizualizacijos priemonių, kurios apdoroja didelius duomenų srautus.



6 pav. Naujų technologijų pritaikymo svarba

Šaltinis: The State Of Fraud And Financial Crime In The U.S. (2022)

Svarbu suprasti ir skaitmeninio intelekto pritaikymo finansinio sukčiavimo atpažinimo tyrime stabdančius veiksnius. Tyrime dalyvavę respondentai sutinka, kad šie veiksniai (žr. 7 pav.), slopina naujoves arba naujų funkcijų įtraukimą į jau esamas sistemas. Pinigų plovimo prevencijoje vadovaujančias pareigas užimančių asmenų net 85 proc. mano, jog naujų technologijų sistemų naudojimas yra sudėtingas bei pripažįsta, jog susiduria su sunkumais integruojant naujus sprendimus į esamas sistemas. Taip pat net 75 proc. respondentų įvardija, jog naudojama sistema, technologija yra veiksnys stabdantis naujų technologijų pritaikymą veikloje.



7 pav. Naujų technologijų diegimą slopinantys veiksniai

Šaltinis: The State Of Fraud And Financial Crime In The U.S. (2022)

Apibendrinant tyrimą galima teigti, jog aukščiausio lygmens vadovai aiškiai supranta, kad dirbtinio intelekto pagrindu sukurti sprendimai yra būdas suvaldyti finansinę nusikalstamą veiklą. Taip pat įmonės teikiančios finansines paslaugas ieško būdų kaip kovoti su vis sudėtingesnėmis finansinių nusikaltimų formomis, taikydamos naujų technologijų sprendimus tokius kaip dirbtinis intelektas. Tyrimo rezultatai paviešino veiksnius, kurie visgi stabdo naujų technologijų diegimą pinigų plovimo prevencijoje t. y. nusistatymas, jog dirbtinio intelekto metodai atrodo sudėtingi ir sunkiai pritaikomi.

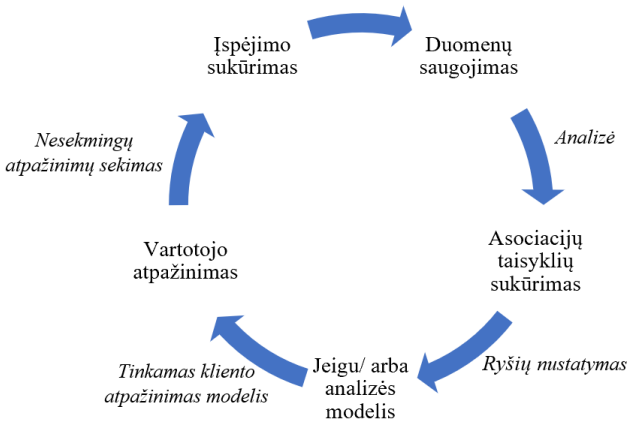
1.4. Finansinio sukčiavimo atpažinimo technologijos

Šiame skyriuje analizuojant pasirinktą mokslinę literatūrą, mokslinius žurnalus, straipsnius bei kitus internetinius šaltinius – aptariamas sukčiavimo ciklas bei naudojamos sukčiavimo atpažinimo technologijos, bei analizuojami, lyginami skaitmeninio intelekto metodai.

Kadangi finansinis sukčiavimas išlieka aktuali tema globaliu mastu ir šis reiškinys turi neigiamų pasekmių visai ekonomikai, ypač svarbu sekti, kontroliuoti ir imtis prevencinių veiksmų siekant maksimaliai sumažinti šią nusikalstamą veiklą (Mantegna & Giudici, 2018). Vidaus auditas ir nuolatinė stebėsena yra pagrindinės priemonės, leidžiančios aptikti didžiausias grėsmes, kurios šiais laikai yra užkoduotos didžiuojuosi duomenų kiekiuose (Deloitte, 2022). Šiuolaikinėse duomenų bazėse saugomos informacijos kiekis yra milžiniškas

ir norint laiku nustatyti finansinį sukčiavimą ir veiksmingai priimti sprendimus, turima informacija turi būti išvalyta, struktūruota, apibendrinta (Kobadilov et al., 2020).

Tam, kad būtų galima efektyviau aptikti įtartiną veiklą buvo sukurtas automatizuotų veiklų rinkinys, kurio tikslas aptikti ir blokuoti nestandartinę veiklą (žr. 8 pav.). Pateiktame sukčiavimo atpažinimo cikle pradžia laikoma tinkamas kliento prieigos kontrolės metodas. Šiais laikais turėti tinkamą autentifikacijos metodą yra būtina, nes 2021 metų US atlikto tyrimo duomenis nuo 2021 m. iki 2022 m. prieigos nusavinimas be asmens sutikimo išaugo 65 proc. (Finteth Global, 2023).



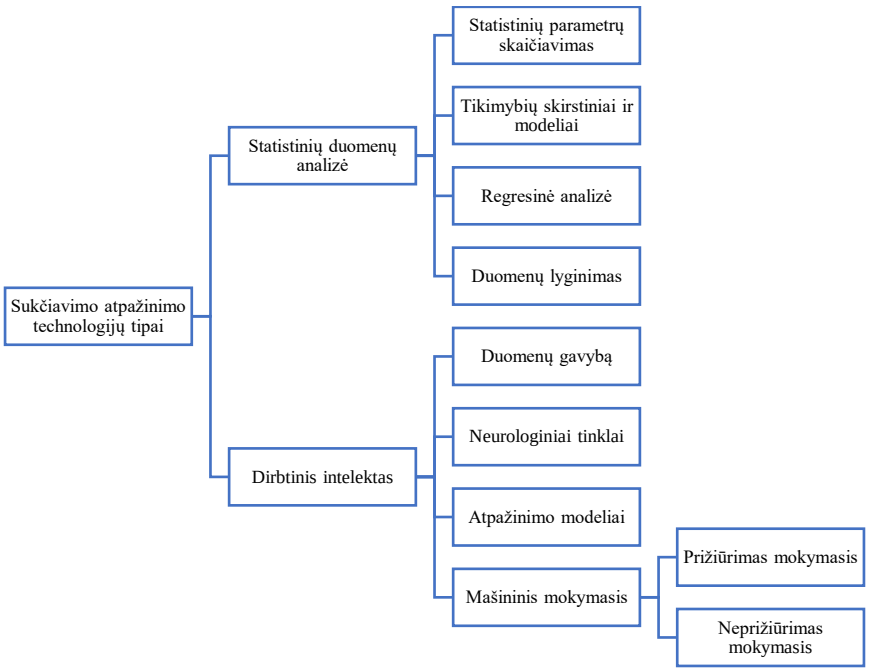
8 pav. Sukčiavimo atpažinimo ciklas

Šaltinis: Kanade (2021)

Kitas žingsnis yra nesėkmingų atpažinimų sekimas ir įspėjimo kontrolės sukūrimas, kuris inicijuotų konkrečius veiksmus.

Svarbu suprasti, kad sukčiavimo aptikimas – tai procesas, ar visuma veiksmų, kurių imamasi siekiant aptikti ir užblokuoti sukčių bandymą apgaule įgyti pinigų ar turto (Kanade, 2021). Dažnu atveju, tai daugiasluoksnis veiksmas, todėl finansinėms institucijos jį atpažinti tampa vis sunkiau ir klasikinių technologijų nebepakanka.

Organizacijos diegia šiuolaikines sukčiavimo aptikimo ir prevencijos technologijas bei rizikos valdymo strategijas, siekdamos kovoti su augančiais nesąžiningais sandoriais įvairiose platformose (Kanade, 2021; Mantegna, Giudici, 2018). Mokslinėje literatūroje išskiriami šie technologijų metodai, naudojami sukčiavimui nustatyti (žr. 9. pav.) (Kanade, 2021):

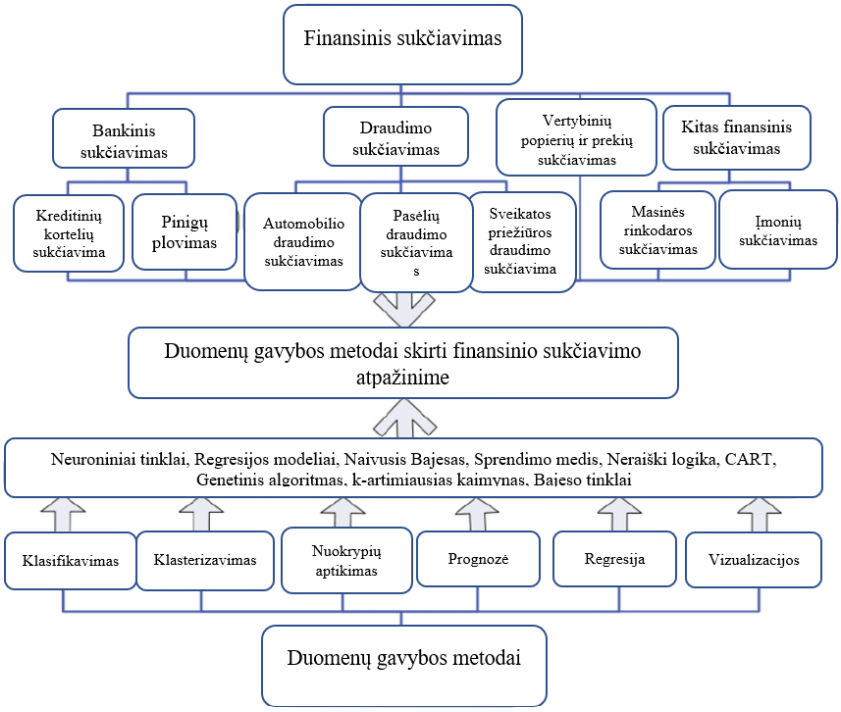


9 pav. Sukčiavimo atpažinimo technologijų tipai

Šaltinis: Kanade (2021)

Sukčiavimo aptikimo statistinė duomenų analizė atlieka įvairias statistines operacijas, tokias kaip sukčiavimo duomenų rinkimas, sukčiavimo nustatymas ir sukčiavimo patvirtinimas, atliekant išsamius tyrimus. Augant duomenų kiekiui tik statistinės analizės metodai nebepajėgia greitai ir laiku atpažinti sukčiavimo atvejus. Kalbant apie technologijas, kurių veikimas pagrįstas dirbtiniu intelektu, jos gali būti pritaikomos ne tik atpažinimui iš istorinių duomenų bet ir sukčiavimo prevencijai, kadangi mokydamasis iš istorinių duomenų, gali greičiau pastebėti atsikartojimą, kas praeityje buvo pripažinta kaip finansinis sukčiavimas (West, Bhattacharya, 2015). Šios sukčiavimo atpažinimo technologijos padeda įmonėms sustiprinti vidinį saugumą ir supaprastinti verslo procesus. Didėjant metodų pritaikomumui, tikslumui bei efektyvumui, dirbtinis intelektas tapo pagrindine technologija, padedančia užkirsti kelią sukčiavimui (Duhart, 2016).

Žemiau pateiktame paveiksle (žr. 10 pav.) matome finansinio sukčiavimo tipus ir taikomas finansinio sukčiavimo atpažinimo technologijas. Siekiant atpažinti nelegalią veiklą yra atliekami šie veiksmai: klasifikuojama, klasterizuojama, daromos prognozės, vizualizacijos.



10 pav. Finansinio sukčiavimo sritys ir sukčiavimo atpažinimo technologijų taikymas

Šaltinis: Sabau (2012)

Kadangi, finansinis sukčiavimas yra sudėtingas, daugiasluoksnis procesas, todėl jo aptikimui naudojami įvairūs metodai nuo neuroninių tinklų, sprendimo medžių, kurie pasižymi stipriu klasterizavimu iki fuzzy logic, kuri gali būti naudojama tikrinant ar nėra suklastos nuotraukos, dokumentai (Teichmann, 2019).

Sukčiavimo veikla skiriasi priklausomai nuo finansinės veiklos formos, kurios buvo aptartos 1.1 skyriuje. Toliau darbe bus koncentruojamasi į finansinį sukčiavimą, kuris pasitaiko bankų, finansinių technologijų srityje veikiačių įmonių veikloje.

Apibendrinant, svarbu paminėti ir audito kompanijos „Deloitte“ (2022) išvadas, kuriose teigiama, kad sudėtingėjant finansiniams produktams, sparčiai diegiant naujas technologijas, klasikinės sistemos nebeatliepia finansų institucijoms keliamų reikalavimų, todėl būtent dirbtinis intelektas geba pagerinti įvairias galimybes siekiant: atpažinti modelius, numatyti ateities įvykius; sukurti tinkamas taisykles, priimti tikslingus sprendimus. Taip pat svarbu akcentuoti, kad dirbtinio intelekto modelių įvairovė tuo pačiu kelia klausimą, kuriuos metodus reikėtų taikyti siekiant efektyviai atpažinti finansinį sukčiavimą.

1.5. Skaitmeninio intelekto sanprata

Šiame darbe nagrinėjama, kaip skaitmeninio intelekto (angl. computational intelligence) (toliau – CI) technologijos gali prisidėti prie finansinio sukčiavimo atpažinimo.

Kadangi, dirbtinio intelekto (angl. artificial intelligence) (toliau – AI) technologijos susiduria su problemomis, tokiomis kaip didelis taisyklių rinkinys, kuris tampa vis sudėtingesnis bei taisyklių nesuvienodinimas tarp valstybių, todėl jis ne visada atliepia poreikį optimaliai mokintis ir laiku aptikti nukrypimus nuo standartų (Hernandez, et al. 2017; Sadiku et. al., 2018). Pasikartojančios AI klaidos atvėrė kelią skaitmeninio intelekto įrankių atsiradimui (Namakkal, 2019).

Analizuojant literatūrą pateikiami šie skirtumai tarp AI ir CI (Khillar, 2023; Hernandez, et al. 2017; Sadiku et. al., 2018):

- AI ir CI apibrėžimo skirtumai:

Dirbtinis intelektas – tai kompiuterių mokslo sritis apimanti platų spektrą technologijų, kurios imituoja žmogaus intelektą. AI apima algoritmus, kurie leidžia kompiuteriams mokytis iš duomenų, atlikti sprendimus, spręsti problemas ir netgi įgyti įžvalgų.

Skaitmeninis intelektas yra AI pošakis, kuris apima praktines priemones, metodologijas ir sistemas, kurios padeda tvarkyti ir naudoti įmonės informaciją strateginiu ir efektyviu būdu.

- AI ir CI tikslų skirtumai:

AI yra vienas iš ypatingų XXI amžiaus technologinių laimėjimų, kuriais siekiama imituoti žmogaus intelektą mašinose, kad jie mąstytų ir elgtųsi kaip žmonės. Idėja yra sukurti protingas mašinas, kurios galėtų elgtis protingai ir mąstyti bei mokytis kaip žmonės. Tikslas – įdiegti žmogaus intelektą mašinose.

Pagrindinis CI tikslas yra suprasti skaičiavimo paradigmas, kurios įgalina protingą elgesį natūraliose ar dirbtinėse sistemose sudėtingose ir kintančiose aplinkose.

- AI ir CI taikymo skirtumai:

AI yra žmogaus intelekto, o ne natūralaus intelekto, modeliavimas mašinomis, ypač kompiuterinėmis sistemomis. Geriausias realus AI pavyzdys tikriausiai yra mašininis mokymasis, kuris reiškia automatizuotas sistemas, kurios gali apdoroti didelius duomenų kiekius, kad gautų naudingą informaciją. Kitos paplitusios AI programos yra kalbos atpažinimas, rašybos atpažinimas, optinis simbolių atpažinimas, mašininis matymas, natūralios kalbos apdorojimas, didelių duomenų sprendimai ir kt. Dirbtinis intelektas naudojamas beveik visuose sektoriuose, įskaitant finansus, sveikatos priežiūrą, socialinę žiniasklaidą, verslą, turizmą ir kt.

CI taikomi išmanieji buitiniai prietaisai, medicininė diagnostika, bankininkystė ir buitinė elektronika, optimizavimo programos, pramonė ir pan. .

Toliau pateiktoje lentelėje (žr. 3 lentelė) pateikiami skirtingi skaitmeninio intelekto apibrėžimai

3 lentelė. Skaitmeninio intelekto sampratos analizė

Autorius ir metai	Sąvokos paaiškinimas
Bazdek J. C., 1994	Sistema vadinama skaičiavimo požiūriu intelektualia, jei ji tvarko žemo lygio duomenis, pvz., skaitmeninius duomenis, turi modelio atpažinimas komponentus ir nenaudoja žinių dirbtinio intelekto prasme. Taip pat algoritmas rodo adaptyvius skaičiavimus.
Konar et al., 2006	Skaitmeninis intelektas gali būti apibrėžtas kaip protingų įrankių ir skaičiavimo priemonių modelis, kuris gali tiesiogiai priimti neapdorotus duomenis ir tiesiogiai juos apdoroti paskirstytu būdu bei toleruoja netikslumą, neapibrėžtumą, dalinę tiesą.
Siddique et al.,2013	Skaitmeninis intelektas, tai dirbtinio intelekto pogrupis, adaptacinių mechanizmų gebėjimas išmokti konkrečią užduotį iš besikeičiančių duomenų ar eksperimentinio stebėjimo.
Raj, 2019	Sąvoka „skaitmeninis intelektas“ paprastai reiškia kompiuterio gebėjimą išmokti konkrečią užduotį iš duomenų ar eksperimentinio stebėjimo. Skaitmeninis intelektas paprastai laikomas minkštojo skaičiavimo sinonimu.
Khillar, 2023	Skaičiavimo intelektas, taip pat vadinamas CI, reiškia skaičiavimo požiūriu intelektualią sistemą, kuriai būdinga skaičiavimo pritaikymo galimybė, atsparumas gedimams ir didelis skaičiavimo greitis.

Šaltinis: Konar et al. (2006), Bezdek (1994), Siddique et al. (2013), Raj (2019), Khillar (2023)

Apibendrinant galima, teigti, kad skaitmeninis intelektas yra naujai besivystanti tyrimų sritis. Jis laikomas dirbtinio intelekto šaka. Prie CI būtų galima priskirti tą metodą, kuris geba išmokti konkrečią užduotį iš duomenų ar eksperimentinio stebėjimo. Nors jis paprastai laikomas minkštojo skaičiavimo sinonimu, vis dar nėra visuotinai priimto skaitmeninio intelekto apibrėžimo.

1.6. Skaitmeninio intelekto modeliai

Mokslinėje klasikinėje literatūroje išskiriamos šios pagrindinės skaitmeninio intelekto technikos: dirbtiniai neuroniniai tinklai (angl. *Artificial neural network*), evoliucinis skaičiavimas (angl. *Evolutionary Computation ar genetic algorithm*), neaiškios logikos sistemos (angl. *Fuzzy logic*) (Bezdek, 1994; Kacprzyk, 2017; Chen, 2021).

Trumpai apie šiuos metodus:

Neuroniniai tinklai yra sudaryti iš neuronų, kurie vienas su kitu sujungti, ir pagal jungtis perduoda signalus (Johnston, 2018). Veikimas susideda iš to, kad tinklas priima įvesties signalą ir kelis kartus apdoroja, kas suteikia gilesnį ir naudingesnį mokymąsi (Rajawat, 2020).

Šio metodo taikymą galima suskirstyti į penkias grupes: duomenų analizė ir klasifikacija, asociatyvioji atmintis, grupių modelių generavimas ir valdymas. Paprastai šiuo metodu siekiama išanalizuoti ir klasifikuoti medicininius duomenis, veido atpažinimas ir sukčiavimo nustatymas (Berniukevičius, Kurilovas, 2017).

Neraiški logika (angl. *Fuzzy logic*) yra daugiareikšmės logikos forma, kurioje kintamųjų tiesos reikšmė gali būti bet koks realusis skaičius nuo 0 iki 1 (HaratiNik, et. al. 2012). Ji naudojama dalinės tiesos sąvokai nustatyti, kai tiesos reikšmė gali svyruoti nuo visiškai teisingos iki visiškai klaidingos (Majhi, 2019).

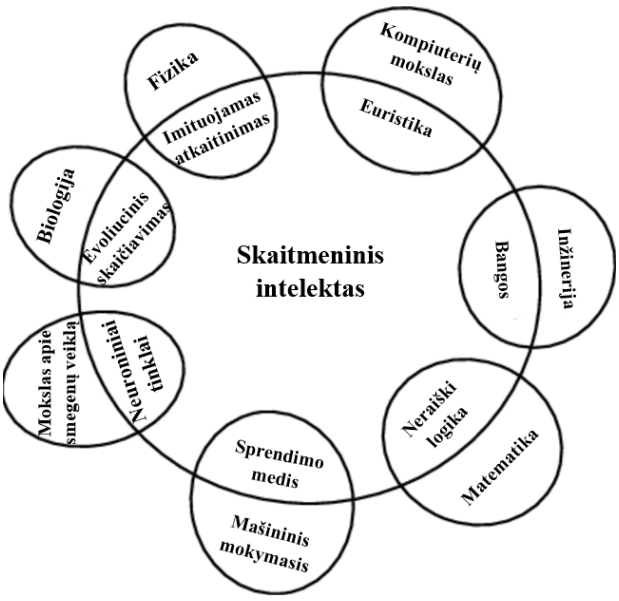
Analizuojant mokslinę literatūrą akcentuojama, jog neraiškioji logika yra vienas labiausiai pritaikomų skaitmeninio intelekto metodų (Majhi, 2019). Ši paradigma susideda iš matavimų ir procesų modeliavimo, sukurtų sudėtingiems realaus gyvenimo procesams (Heidarinia et. al 2014). Priešingai nei dirbtiniam intelektui, kuriam reikalingos tikslios žinios, skaitmenis intelektas ne ir susidurdamas su netiksliais duomenimis ar nepilnais duomenų rinkiniais gali mokintis.

Ši technika paprastai taikoma įvairioms sritims, kurioms reikalinga kontrolė kaip vaizdo apdorojimas ir sprendimų priėmimas, kai naudojatės vaizdo kamera (Siddique, 2013).

Evoliucinis algoritmas, tai algoritmas, imituojantis natūralios atrankos procesą. Jie padeda spręsti optimizavimo ir paieškos problemas. Genetiniai algoritmai priklauso didesnei evoliucinių algoritmų klasei. Genetiniai algoritmai imituoja natūralius biologinius procesus, tokius kaip paveldėjimas, mutacija, atranka ir kryžminimas (Rada, 2008). Genetiniai algoritmai, tai paieškos metodas, dažnai naudojamas informatikoje sudėtingiems, neakivaizdiems algoritminio optimizavimo ir paieškos problemų sprendimams rasti (Brabazon et al., 2008).

Pasak Rados (2008) pagrindinis šio principo taikymas apima tokias sritis kaip kelių tikslų optimizavimas, kuriems tradicinės matematinės technikos nebepakanka pritaikyti įvairioms problemoms, tokioms kaip DNR analizė.

11 paveikslas (žr. 11 pav.) iliustruoja ryšį tarp skaitmeninio intelekto ir kitų sričių. CI labai daug naudos suteikė fizika (imituojamas atkaitinimas), biologija (evoliucinis skaičiavimas ir spiečių intelektas), mokslas apie smegenų veiklą (dirbtiniai neuroniniai tinklai), matematinis kalbos modeliavimas (neaiškioji logika), kompiuterių mokslai (gramatika, sudėtingumo teorija ir euristika), inžinerija (bangos) ir mašininis mokymasis (sprendimų medžiai).



11 pav. Skaitmeninio intelekto ir kitų mokslų sinergija

Šaltinis: Raj (2019)

Taip pat svarbu paminėti, jog atsirado mokslininkai tokie kaip Pushpa (2017), Kacprzyk (2017) kurie prie CI papildomai priskiria šiuos modelius: Bajeso tinklas (angl. Bayesian/ Belief networks), Mokymosi teorija (angl. learning theory), Tikimybinė logika (angl. probabilistic reasoning).

Trumpai apie šiuos metodus:

Bajeso tinklas yra tikimybinis grafinis modelis, tai tam tikras grafas, naudojamas modeliuoti įvykiams, kurių negalima stebėti. Vėliau jį galima naudoti išvadoms daryti. Naudojamas kryptinis grafas, kuriame nėra ciklų. Grafo mazgai vaizduoja atsitiktinius kintamuosius. Jei du mazgus jungia briauna, su ja susijusi tikimybė, kad ji bus perduodama iš vieno mazgo į kitą.

Bajeso tinklai dažniausiai naudojami mašininio mokymosi srityje. Jie naudojami ten, kur reikia klasifikuoti informaciją. Pavyzdžiai - vaizdų, dokumentų ar kalbos atpažinimas ir informacijos paieška. (Kevin, Chen, 2011).

Mokymosi teorija aprašo, kaip mokiniai mokymosi metu gauna, apdoroja ir išlaiko žinias. Kognityvinė, emocinė ir aplinkos įtaka (Dunjko, Briegel 2018).

Tikimybinė logika apima tikimybės ir logikos naudojimą neapibrėžtose situacijose. Tikimybinė logika išplečia tradicines logikos tiesos lenteles tikimybinėmis išraiškomis. Tikimybinės logikos sunkumas yra jų polinkis padauginti tikimybinių ir loginių komponentų skaičiavimo sudėtingumą (Triepels, et al. 2018).

Žemiau pateikiama lentelė, kurioje lyginami skaitmeninio intelekto metodai (žr. 4 lentelė).

4 lentelė. Skaitmeninio intelekto technologijų palyginimas

	Technologija	Aprašymas	Taikymas	Privalumai	Trūkumai
Skaitmeninis intelektas	Neraiški logika	Naudojamos aibės nariams įvertinti iš daugybės rinkinių.	Patentų atpažinimas, kelių tikslų optimizavimas.	Užtikrina lengvą samprotavimą, įgyvendinimą ir gebėjimą valdyti neapibrėžtumus ir netiesiškumą.	Patikimumo stoka.
	Neuroniniai tinklai	Metodas kuria paprastą matematinę sistemą, panašią į smegenis, išmoksta ją ir įgyja įžvalgų spręsti problemas.	Netiesinio proceso modeliavimas, struktūros numatymas, anomalijų aptikimas, sprendinių sudarymas.	Atsparus klaidoms, mokosi iš pavyzdžių, smulkūs pokyčiai neturi didelės įtakos.	Reikalingas procesorius su lygiagretaus apdorojimo galimybe.
	Evoliucinis algoritmas	Remiasi natūraliu atrankos procesu.	Intelektuali paieška, kelių tikslų optimizavimas.	Lankstus, savaime prisitaikantis optimalus sprendimas.	Ankstyva konvergencija, lemianti vietinį optimalumą.
	Mokymosi teorija	Remiantis ankstesniais rezultatais, kad būtų galima numatyti būsimus rezultatus.	Diagnozė, aptikimas, nelaimių valdymas.	Problemos sprendimas taikoma bet kuriai nuspėjamai mokymosi sferai.	Trūksta detalumo.
	Tikimybinė logika	Nustato galimą problemos rezultatą, pagrįstą ankstesnėmis įžvalgomis.	Pramoninės diagnostikos prognozė.	Tvarko neapibrėžtumą ir rizikos skaičiavimus.	Brangus ir neefektyvus laiko požiūriu.

Šaltinis: Raj (2019)

Apibendrinant galime teigti, kad skaitmeninis intelektas yra vis dar nauja intelektualių sistemų vystymosi paradigma, kurią gali sudaryti neraiškių rinkinių technologija, dirbtiniai neuroniniai tinklai, evoliucinio skaičiavimo metodai, mašininio mokymosi algoritmai. Svarbu pabrėžti, kad CI šiai dienai turėjo įtakos beveik visose srityse pradedant biologijos, baigiant kompiuterių mokslu, fizika, inžinerijai, matematikai ir kizta.

1.7. skaitmeninio intelekto modelių taikymo fukcijos

Hurtas (2023) teigia, kad žmogaus smegenų ir dirbtinio intelekto veikimas turi panašumų, tokių kaip galimybę apdoroti duomenis, mokytis ir samprotauti.

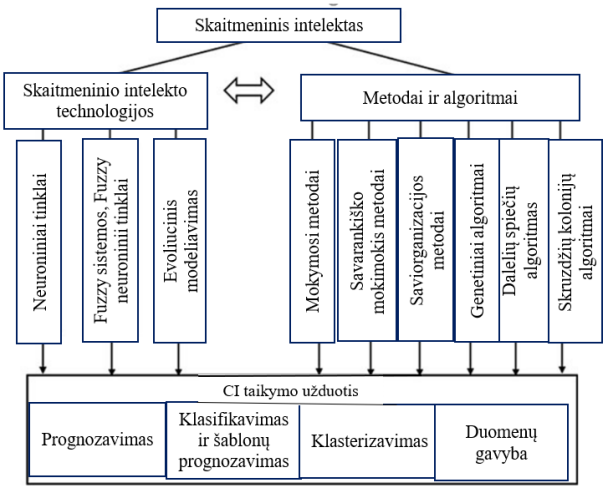
Mokslinėje literatūroje apie žmonių smegenų veiklą ir skaitmeninio intelekto metodų lyginimą išskiriami šie panašumai tarp skaitmeninio intelekto modelių ir žmonių smegenų veikimo:

1. Neuronų struktūra ir informacijos perdavimas: Skaitmeninis intelektas naudoja dirbtinius neuroninius tinklus, kurių veikimas atkartoja neuronų struktūrą smegenyse. Tinklai yra sudaryti iš daugelio dirbtinių neuronų, kurie gali perduoti ir apdoroti informacija, taip pat veikia ir žmonių smegenys. Skaitmeninio intelekto modeliai

perduoda informaciją per tinklą. Kiekvienas dirbtinis neuronas gauna įvesties signalus, apdoroja juos ir perduoda rezultatą kitam neuronui.

2. Mokymas ir prisitaikymas: Skaitmeniniai intelekto modeliai pagal praeities duomenis sukuria šabloną ir pritaikius apmokymo funkciją mokosi naudojant praeities duomenis bei pritaiko, kad geriau atliktų užduotis ateityje. Šis veiksmas panašu į žmonių smegenų gebėjimą mokytis iš patirties ir prisitaikyti prie naujų situacijų.
3. Prisitaikymas prie paklaidų: Skaitmeniniai intelekto modeliai, tokie kaip rekurentiniai neuroniniai tinklai, gali apdoroti sekas arba eilės duomenis. Jie turi gebėjimą prisitaikyti prie paklaidų ar praleistų informacijos fragmentų. Panašiai kaip ir žmonių smegenys žinodamos kontekstą gali spėti praleistą informaciją.

Skaitmeninis intelektas yra koncepcija, kuri apima skirtingas technologijas, algoritmus ir sistemas, leidžiančias kompiuteriams mokytis, analizuoti duomenis, atlikti užduotis, suprasti žmogišką kalbą ir net spėti bei spręsti problemas. Skaitmeninio intelekto struktūra gali būti sudaryta iš kelių pagrindinių komponentų (žr. 12 pav.).



12 pav. Skaitmeninio intelekto struktūra

Šaltinis: Kacprzyk (2017)

Kacprzykis (2017) išskiria šias skaitmeninio intelekto metodų taikymo sritis:

- Prognozavimas, atpažinimas. Sprendimų priėmimas ir veiksmų vykdymo etapas apima naudojimosi gautais rezultatais ar prognozėmis, kad būtų priimami sprendimai arba atliekami veiksmai. Tai gali būti automatinis sprendimų priėmimas, rekomendacijų sistema ar net autonominių sistemų veiksmų valdymas.
- Klasifikacijos modelio atpažinimas. Skirtingi algoritmai ir modeliai yra naudojami, kad sistema gebėtų atpažinti modelius, mokytųsi iš sukurtų šablonų ir tobulėtų.

- Grupavimas.
- Duomenų gavyba. Skaitmeninis intelektas gali apdoroti įvairius duomenis - nuo teksto ir vaizdų iki garso įrašų. Šie duomenys yra renkami, saugomi ir tvarkomi. Duomenų apdorojimo etape duomenys yra valomi, filtruojami ir apdorojami. Tai gali apimti duomenų normalizavimą, transformavimą į tinkamą formatą bei šių duomenų apdorojimą įvairiais būdais, pvz., statistiniais metodais arba mašininio mokymosi algoritmais.

Apibendrinant CI struktūros dalys gali būti sudėtingos ir skirtingos pagal konkrečią CI sistemą ar naudojamus algoritmus bei technologijas. Tačiau ši bendra struktūra padeda suprasti pagrindinius skaitmeninio intelekto veikimo principus.

1.7.1. Skaitmeninio intelekto modelio atpažinimo funkcija

Prieš dirbtinio intelekto sukūrimą žmonių smegenų veikla išsiskyrė ypatingu sugebėjimu atpažinti daiktus, vietas, elgesį (Warwick, 2012; Chen et al. 2022). Tačiau šiai dienai atpažinimo funkcija taip pat sutinkama ir skaitmeninio intelekto modeliuose.

Skaitmeninio intelekto modelis gali būti pritaikytas atpažinti įvairias nestandartines situacija, kai vartotojas elgiasi nebūdingai (prieš tai modelis apmokomas kaip vartotojas standartiškai elgiasi), kas gali rodyti pinigų plovimo veiklą (Kacprzyk, 2012). Čia yra kelios atpažinimo funkcijos, kurios gali būti naudojamos skaitmeninio intelekto modeliuose, siekiant atpažinti pinigų plovimą (Raj, 2019; Warwick, 2012):

1. Transakcijų šablonų atpažinimas: Skaitmeninio intelekto modeliai gali būti apmokyti atpažinti įprastus pinigų plovimo transakcijų šablonus. Tai apima dažnai pasikartojančius pinigų sumų ir laiko tarpo modelius, transakcijų tarpusavio ryšius arba transakcijų tipus, kurie dažnai susiję su pinigų plovimu. Atpažinus šiuos šablonus, modeliai gali įspėti apie galimus finansinio sukčiavimo atvejus.
2. Anomalijų paieška duomenyse: Skaitmeninio intelekto modeliai gali ieškoti anomalijų duomenyse, kurios neatitinka įprastos veiklos arba kurie yra neįprasti dėl savo dydžio, sandaros arba kitų savybių. Kaip pavyzdį galėtų būti padidėjęs mažų sumų srautas iš įvairių šaltinių arba neįprastai didelės pinigų įplaukos iš įvairių šaltinių, iš įvairių šalių. Tokie netipiniai duomenys gali būti indikacija finansinio sukčiavimo veiklai.
3. Socialinių tinklų analizė: finansinis sukčiavimas yra kompleksinė, sudėtinga veikla, kurioje dalyvauja daug asmenų ar organizacijų. Skaitmeninio intelekto modeliai gali analizuoti šiuos tinklus ir identifikuoti paslėptus ryšius arba veikėjus, kurie yra svarbūs.

4. Teksto analizė: Skaitmeninio intelekto modeliai gali būti apmokyti analizuoti tekstinę informaciją, pvz., el. pašto žinutes, finansinius dokumentus. Šie modeliai gali identifikuoti įtartinus žodžius, frazes ar sakinio struktūras, kurios gali rodyti finansinį sukčiavimą Pvz., specifiniai žodžiai ar frazės, susijusios su neįprastais pinigų perkėlimais arba slaptomis sandėrių schemomis, gali būti aptiktos ir panaudotos kaip signalai galimam finansiniam sukčiavimui.

Skaitmeninio intelekto modelio atpažinimo funkcija yra esminis komponentas, leidžiantis sistemai identifikuoti, klasifikuoti arba atpažinti duomenis, modelius arba reiškinius. Tai gali būti įvairių tipų funkcijos, priklausomai nuo taikomos technologijos ar algoritmo.

1.7.2. Skaitmeninio intelekto klasifikavimo funkcija

Klasifikavimo funkcija skaitmeninio intelekto modeliuose yra skirta identifikuoti, klasifikuoti konkrečias pinigų transakcijas, klientus arba veiksmus kaip potencialiai įtartinus arba potencialiai susijusius su finansiniu sukčiavimu.

Klasifikavimo funkcija remiasi mokymosi algoritmais, kurie gali išmokti atpažinti šablonus ir taisykles, kurie dažnai pasireiškia finansinio sukčiavimo operacijose. Čia yra keletas pavyzdžių klasifikavimo funkcijų, kurios gali būti naudojamos skaitmeninio intelekto modeliuose, siekiant klasifikuoti šią veiklą:

1. Klasifikavimas pagal rizikos lygį: Skaitmeninio intelekto modeliai gali būti apmokyti priskirti transakcijoms ar klientams rizikos lygius, kurie rodo galimą ryšį su nelegalia vykdoma veikla. Modeliai gali mokytis iš istorinių duomenų, kuriuose yra žinomi finansinio sukčiavimo atvejai, ir taikyti taisykles arba mašininio mokymosi metodus nustatyti, ar nauja transakcija ar klientas atitinka įprastą elgesio modelį arba turi įtartinų bruožų.
2. Klasifikavimas pagal transakcijų tipus: Skaitmeninio intelekto modeliai gali būti apmokyti atpažinti specifinius transakcijų tipus, kurie yra dažnai susiję su finansiniu sukčiavimu. Pavyzdžiui, tai gali būti tarptautinės pinigų pervedimo operacijos, įplaukos iš įtartinos šalies arba įtartinos sandėriai tarp klientų, kurie neturi įprastų veiksmų tarpusavio ryšių.
3. Klientų rizikos profilio klasifikavimas: Skaitmeninio intelekto modeliai gali būti apmokyti sukurti klientų rizikos profilius, kurie leidžia klasifikuoti klientus pagal jų įtartinas veiklas. Modeliai gali analizuoti kliento transakcijų istoriją, asmens duomenis,

finansinius rodiklius ir kitus faktorius, kad nustatytų, ar klientas yra potencialiai įtartinas arba turėtų būti nagrinėjamas išsamia.

Šios atpažinimo funkcijos gali būti naudojamos individualiai arba kombinuojamos, siekiant sukurti skaitmeninio intelekto modelius, kurie geba efektyviai aptikti ir atpažinti finansinio sukčiavimo veiklą.

1.8. Skaitmeninio intelekto modelių taikymas siekiant atpažinti finansinį sukčiavimą

Kadangi, skaitmeninio intelekto metodų įvairovę yra labai didelė šiame skyriuje analizuojant mokslinę literatūrą siekiama išsiaiškinti, kokie metodai daugiausiai naudojami moksliniuose tyrimuose siekiant atpažinti finansinį sukčiavimą bankų ir finansinių technologijų įmonių veikloje.

5 lentelėje (žr. 5 lentelė) pateikiami literatūroje nagrinėti CI metodų taikymo tyrimai bankų veikloje siekiant aptikti finansinį sukčiavimą. Matome, kad labai dažnai naudojami šie metodai: neuroninis tinklas ir sprendimo medis ir jo pagrindu sukurti algoritmai.

5 lentelė. Skaitmeninio intelekto modelių taikymas finansinio sukčiavimo atpažinimo tyrime

Finansinio sukčiavimo tipas	**Aprašymas**	**Technologijos taikymas**	**Autoriai**
Kreditinių kortelių sukčiavimas	neteisėtas kortelės naudojimas be tinkamo savininko leidimo	Neuroniniai tinklai;	Özçelik, et. al. 2010; Srivastava, et. al. 2016; Ghobadi, Rohani, 2016; Malini, N.; Pushpa, M, 2017; Desrousseaux, Bernard, Mariage 2021; Rocha-Salazar, Segovia-Vargas, Camacho-Miñano, 2021
		Sprendimų medis; Atsitiktiniai miškai; Gradiento didinimo algoritmai	Awoyemi, 2017; Randhawa, et. al. 2018; Salehi, Ghazanfari, Fathian, 2017; Chen, Wang, Kuo, 2017; Badawi, 2021; Barua et al. 2021
		Naïve Bayes	Ngai et. al., 2011; Song et. al. 2020
		Atraminių vektorių klasifikatorius (Support vector machine)	Gyamf, 2018; Mareeswari, Gunasekaran, 2016 Kho, Vea, 2017
		Neraiški logika	HaratiNik, et. al. 2012

5 lentelės tęsinys

Finansinio sukčiavimo tipas	**Aprašymas**	**Technologijos taikymas**	**Autoriai**
P2P skolinimo sukčiavimo aptikimas: didelių duomenų metodas.		Neuroniniai tinklai	Xu, Lu, Chau. 2015
		Sprendimų medis	Xu, Lu, Chau. 2015

Šaltinis: sudaryta autorės

Kombinuojant šiuos metodus galime sukurti neuroninį tinklą, kuris mokysis iš finansinių operacijų duomenų ir pagrįstų gautus rezultatus, naudojant sprendimo medžio algoritmą atliekant išvadas ar klasifikavimą, remiantis neuroninio tinklo prognozėmis.

Neuroniniai tinklų ir sprendimų medžiai yra plačiai naudojami finansinio sukčiavimo atpažinimo uždaviniuose dėl keleto svarbių priežasčių (Desrousseaux, Bernard, Mariage 2021); Rocha-Salazar, Segovia-Vargas, Camacho-Miñano 2021); Salehi, Ghazanfari, Fathian 2017); Badawi 2021):

6 lentelė. Neuroninių tinklų ir sprendimo medžių modelių pritaikymo naudos

Savybė	**Metodas**	**Komentaras**	**Autorius**
Gebėjimas išmokti kompleksinius neaiškius ryšius	Neuroniniai tinklų	Neuroniniai tinklai yra priemonė, kuri gali išmokti sudėtingus neaiškius ryšius tarp duomenų elementų. Finansinio sukčiavimo atveju dažnai yra daugybė sąryšių ir šablonų, kurie gali būti sunkiai aptinkami arba suprantami žmogui, tačiau neuroninis tinklas gali išmokti juos atpažinti.	Desrousseaux, Bernard, Mariage (2021)
Duomenų modeliavimas	Neuroniniai tinklų	Neuroniniai tinklai gali būti naudojami atpažinti sudėtingus duomenų modelius, kuriuose yra daug kintamųjų ir sąveikų. Finansinio sukčiavimo atveju gali būti naudinga modeliuoti finansinius ryšius, mokėjimų srautus, pavedimus ir kitus duomenis, kad būtų nustatyti netipiniai ir įtartiną elgesį rodantys požymiai.	Rocha-Salazar, Segovia-Vargas, Camacho-Miñano (2021)
Automatinis didelių duomenų kiekių apdorojimas	Neuroniniai tinklų, Sprendimo medžių	Gali efektyviai analizuoti didelius duomenų rinkinius, kurie gali būti susiję su finansinio sukčiavimo atvejais. Finansinės transakcijos, sąskaitų duomenys ir kiti finansiniai įrašai gali sudaryti didelį duomenų srautą, kurį šie modeliai gali apdoroti ir atrasti įtartinus veiksmus.	Desrousseaux, Bernard, Mariage (2021); Rocha-Salazar, Segovia-Vargas, Camacho-Miñano (2021); Badawi (2021)

6 lentelės tęsinys

Savybė	**Metodas**	**Komentaras**	**Autorius**
Adaptacija ir atnaujinimas	Neuroniniai tinklų	Neuroniniai tinklai gali būti pritaikomi ir atnaujinami pagal naujus duomenis. Tai svarbu, nes sukčiavimo schemos ir metodai nuolat kinta. Naudojant neuroninius tinklus, galima atnaujinti modelius ir apmokyti juos atpažinti naujas sukčiavimo formas ir tendencijas.	Desrousseaux, Bernard, Mariage (2021)
Trūkstamų duomenų tvarkymas	Sprendimo medžių	Sprendimų medžio metodas gali efektyviai tvarkyti trūkstamus duomenis. Tai yra naudinga, kai finansinio sukčiavimo atveju gali būti ne visada turimi visi duomenys arba kai trūksta informacijos apie tam tikrus veiksnius. Sprendimų medžiai gali taisyti medžio šakas, priklausomai nuo turimos informacijos, ir toliau klasifikuoti įtartinas transakcijas.	Badawi (2021)
Požymių svarbos nustatymas	Sprendimo medžių	Sprendimų medžio metodas gali nustatyti svarbiausius požymius, kurie prisideda prie finansinio sukčiavimo atpažinimo. Medžio konstravimo metu požymiai yra pasirinkti pagal jų informacijos laimėjimą arba kitus kriterijus. Tai padeda identifikuoti klasių skirtumus ir svarbius požymius, kurie gali būti susiję su neteisėtomis finansinėmis veiklomis.	Salehi, Ghazanfari, Fathian (2017)

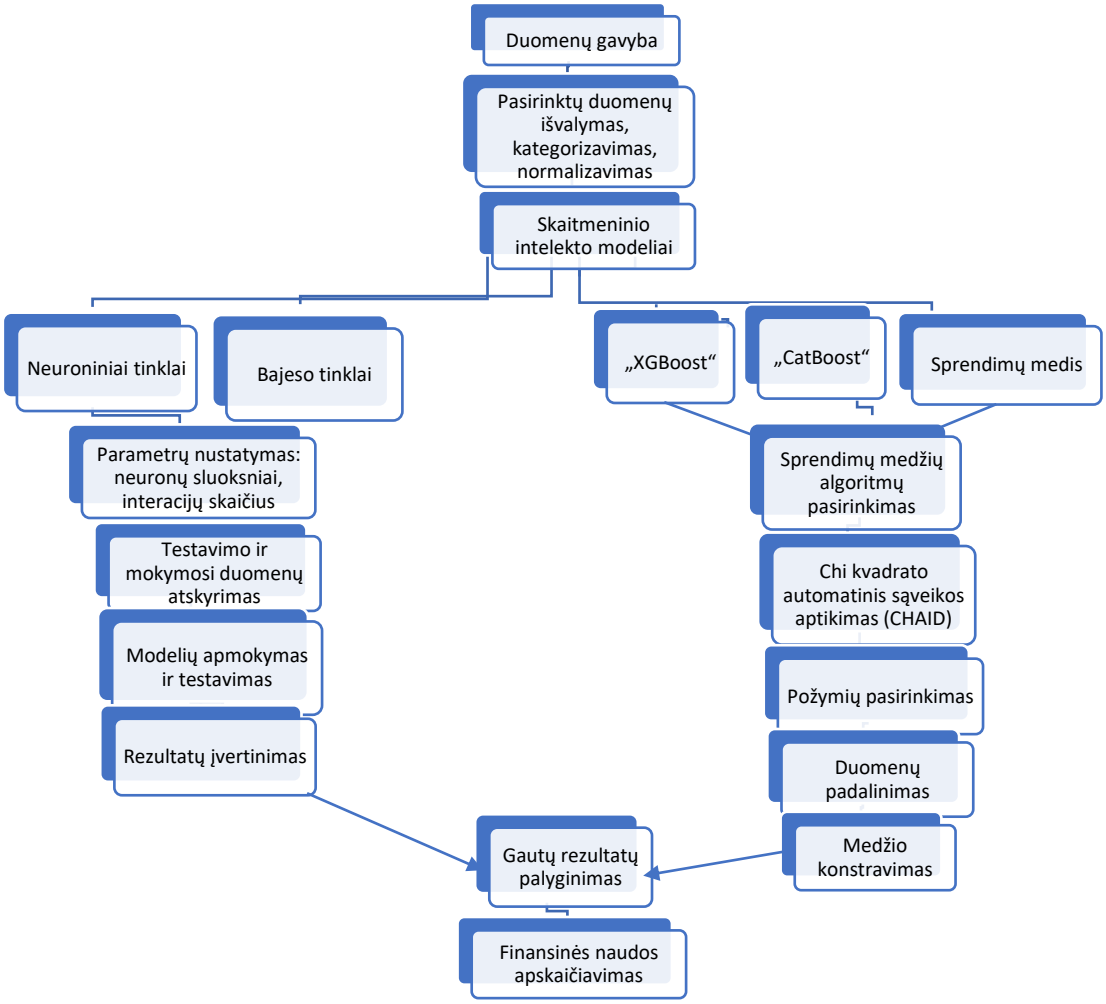
Šaltinis: Desrousseaux et al. (2021), Rocha-Salazar et al. (2021), Salehi et al. (2017), Badawi (2021)

Apibendrinant išanalizuotus mokslinėje literatūroje aprašytus finansinio sukčiavimo bankų ir finansinių technologijų įmonių atvejų tyrimus matome, kad labai dažnai naudojami CI metodai yra neuroniniai tinklai, sprendimų medis ir kiti iš jo sukurti algoritmai (atsitiktiniai miškai, „„XGBoost““, „„CatBoost““), atraminių vektorių algoritmai. Dažnas šių metodų naudojimas yra grindžiamas tuo, kad jie pasižymi šiomis savybėmis: sugeba apdoroti didžiulius duomenų kiekius, padeda aptikti netipinį finansinį elgesį ir nustatyti įtartinas transakcijas ar veiksmus, kurie gali būti susiję su neteisėtomis finansinėmis veiklomis. Dėl šių išvardintų metodų savybių būtent šie metodai bus pritaikyti tyrimo dalyje.

2. FINANSINIO SUKČIAVIMO ATPAŽINIMO TYRIMO METODŲ VERTINIMAS

Didėjant finansinių priemonių ir produktų įvairovei, rinka darosi vis sudėtingesnė. Be to, finansų įstaigoms nustatomi vis griežtesni reikalavimai, susiję su jų pareiga aptikti, tirti ir pranešti apie įtartiną elgesį. Norint visiškai atitikti reikalavimus, nebeužtenka tradicinių įrankių tokių kaip duomenų lyginimas, statistinių parametrų skaičiavimas, duomenų vizualizavimas. Veiksmingas, proaktyvus finansinio sukčiavimo atpažinimas tapo pažangios stebėjimo sistemos, išsamių lyginamųjų duomenų technologijos, kurių veikimo pagrindas yra dirbtinis intelektas. Ši technologija gali apdoroti didžiuosius duomenis ir tuo pačiu mokintis, nuspėti galimus piktnaudžiavimo atvejus. Toliau pateigtoje schemoje pavaizduota tyrimo darbų eiga (žr. 13 pav.).

.



13 pav. Tyrimo darbo eiga

Šaltinis: sudaryta autorės

2.1. Duomenų apdorojimas

2.1.1. Duomenų anomalijos ir jų tikrinimas

Anomalijų aptikimas yra toks metodas, kuris yra naudojamas siekiant nustatyti neįprastus modelius, kurie yra neatitinkantys tikėtino kliento elgesio (Chandola, Banerjee, Kumar, 2009). Pačios anomalijos tai yra reti įvykiai, kurie kelia įtarimus, nes tam tikri rodikliai ženkliai išsiskiria nuo kitų duomenų. Anomalijos yra skirstomos į (Shikha, Agrawal, 2015):

• Taško anomalijas (angl. Point Anomalies), kai tam tikras rodmuo yra atitolęs nuo kitų rodmenų. Sukčiavimų atveju galima atsižvelgti į mokėtinas sumas, kurios gali būti stipriai didesnės lyginant su kitomis.

• Kontekstines anomalijas (angl. Contextual Anomalies), kurios priklauso nuo konteksto ir yra įprastas reiškinys laiko eilučių duomenyse. Vienas iš pavyzdžių galėtų būti tam tikru laiko momentu išleidžiama vienoda suma ar vienoda operacija tam tikrame regione.

• Kolektyvines anomalijas (angl. Collective Anomalies), kai duomenų rinkinys padeda nustatyti anomalijas ir vaizduoja netikėtus derinius. Pavyzdžiui, kai yra perkama daugybė brangių daiktų.

Apibendrinant, anomalijos yra nustatomos pagal aprašytas taisykles. Jei iškelta hipotezė yra teisinga pagal tam tikras taisykles, galima interpretuoti, kad tai yra anomalija. Tačiau paprastieji statistiniai modeliai ne visada suveikia. Duomenys gali vaizduoti anomalinį reiškinį, bet riba tarp elgsenos nėra tiksliai nusakyta arba vyraujantis sezoniškumas yra neįvertintas, todėl rezultatai yra iškreipiami. Mašininio mokymosi algoritmai anomalijų nagrinėjime naudojami vis plačiau ir tolimesniame tyrime bus palyginami keli metodai ir jų tikslumai mokėjimų sukčiavimuose.

2.1.2. Duomenų balansavimas

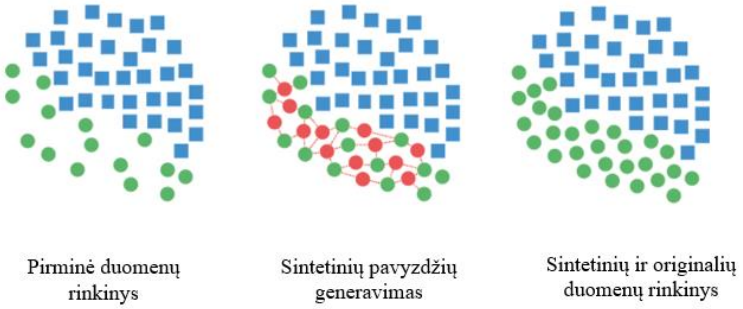
Nesubalansuoti duomenys yra dažna mašininio mokymosi problema, dėl kurios kyla problemų skaičiuojant koreliaciją, su klasių atskyrimu ir vertinimu, o visa tai turi neigiamos įtakos modelio veikimui (Hassan, Kadir, Hussein 2020).

Finansinio sukčiavimo tyrimuose dažnu atveju duomenys yra nesubalansuotas, nes didžioji dauguma transakcijų nėra apgaulingos. Šiame darbe taikytas SMOTE (angl. Synthetic Minority Over-sampling Technique) metodas.

Norėdami išspręsti klasių disbalanso problemą, pasirinkimas paprastai yra susijęs su nepakankamu (angl. Undersampling) arba per didele (angl. Oversampling) atranka (Hassan, Kadir, Hussein 2020):

- Nepakankama atranka (daugumos klasės skaičiaus mažinimas);

- Per didelė atranka (mažumos klasės skaičiaus didinimas).



14 pav. SMOTE metodo veikimas

Šaltinis: Analyttica Datalab's Medium (2021)

SMOTE duomenų rinkinyje sukuria sintetinius (ne pasikartojančius) pavyzdžius. Šis veiksmas labai svarbus, nes po jo mūsų duomenys tampa subalansuoti. Šis metodas veikia kurdama sintetinius mažumos klasės pavyzdžius, interpoliuodama[2] esamus mažumos klasės egzempliorius. Pasak Hasano ir jo kolegų (2020) metodas norėdamas sukurti sintetinį pavyzdį atlieka šiuos veiksmus:

- Pirmas žingnis SMOTE identifikuoja mažumos klasės egzempliorių ir pasirenka vieną ar daugiau artimiausių mažumos klasės kaimynų;
- Antras žingsnis, jia sukuria naują egzempliorių, atsitiktinai pasirinkdamas vieną iš kaimynų. Apskaičiuoja skirtumą tarp dviejų egzempliorių savybių verčių;
- Trečias žingnis SMOTE padaugina šį skirtumą iš atsitiktinio skaičiaus nuo 0 iki 1 ir prideda rezultatą prie pradinio mažumos egzemplioriaus savybių reikšmių, kad sukurtų naują sintetinį egzempliorių.
- Ketvirtas žingsnis, kartodama šį procesą keliems mažumos klasės egzemplioriams, SMOTE sukuria didesnį ir įvairesnį sintetinių mažumos klasės egzempliorių rinkinį, kurį galima įtraukti į pradinį duomenų rinkinį.

Gautas duomenų rinkinys yra labiau subalansuotas, o tai gali pagerinti mašininio mokymosi modelių našumą sumažinant šališkumą daugumos klasės atžvilgiu.

Apibendrinant, SMOTE yra paprastas ir efektyvus išbalansuotų duomenų rinkinių tvarkymo būdas. Svarbu paminėti, kad kartais jis gali sugeneruoti triukšmingus stebinius, ypač, jeigu mažumos klasės duomenys yra panašūs į daugumos klasės, bet reiktų pastebėti, kad su

[2] Interpoliacija (angl. Interpolation). Prognozavimo metodas, kai kintamojo rodiklio reikšmė apytiksliai nustatoma remiantis žinomomis reikšmėmis.

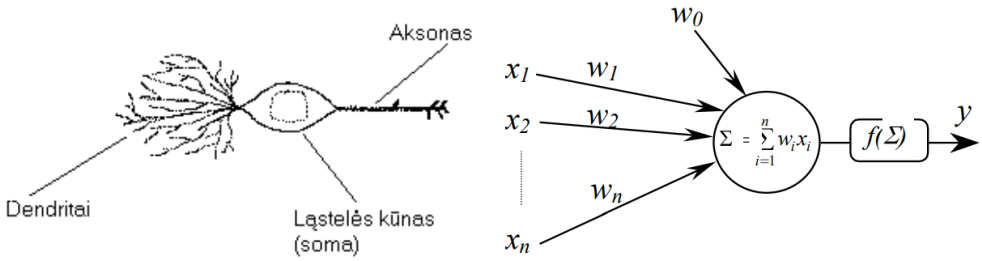
tokiais duomenimis, kai skirtingų klasių stebiniai yra panašūs, bet kokiam metodui mašininio mokymo metodui būtų sunkiau prognozuoti teisingą atsakymą.

2.2. Skaitmeninio intelekto modelių aprašymas

Šiame darbe bus analizuojami šie skaitmeninio intelekto modeliai: neuroninis tinklas, Bajeso tinklas, Sprendimo medis, „„CatBoost““, „„XGBoost““, kurie toliau bus aprašomi.

2.2.1. Neuroninių tinklų metodo aprašymas

Neuroninis tinklas (taip pat vadinamas dirbtiniu neuroniniu tinklu arba ANN) yra prisitaikanti sistema, kuri mokosi naudodama tarpusavyje sujungtus mazgus arba neuronus sluoksniuotoje struktūroje, panašioje į žmogaus smegenis (Johnston, 2018) . Neuroninių tinklų ir žmogaus smegenų neuronų sandara pavaizduota 15 paveiksle (žr. 15 pav.).



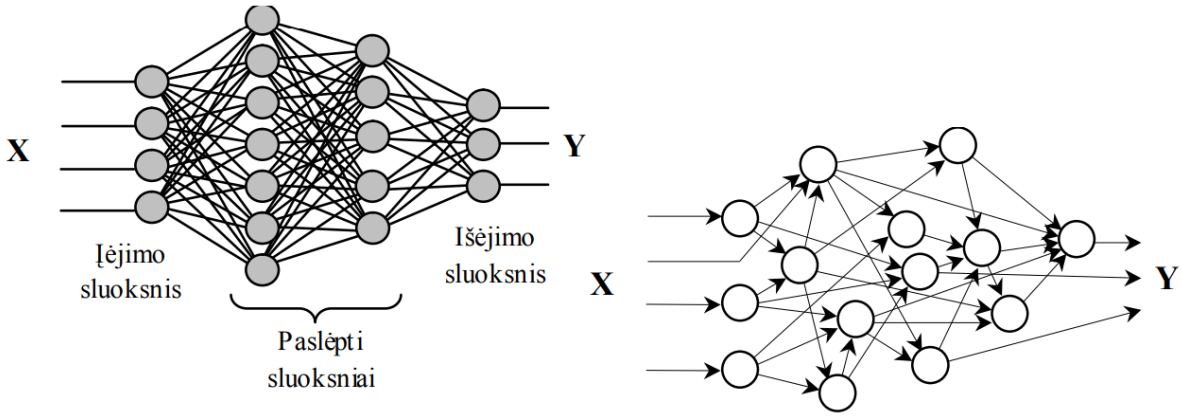
15 pav. Žmogaus smegenų neurono sandara (kairėje) ir dirbtinio neurono sandara (dešinėje)

Šaltinis: Johnston (2018)

Neuroninis tinklas gali mokytis iš duomenų, todėl jį galima išmokyti atpažinti modelius, klasifikuoti duomenis ir numatyti būsimus įvykius (Rajawat, 2020). Jis gali būti išmokytas naudojant daugybę pavyzdžių, kad atpažintų kalbos ar vaizdų modelius, kaip tai daro žmogaus smegenys. Šiame darbe neuroninių tinklų modelis pasirinktas, nes jis gali būti panaudotas klasifikavimui ir regresinėms užduotims atlikti.

Dirbtinio neuroninio tinklo architektūra gali apibrėžti, kaip veikia keli neuronai, kurie yra išdėstomi vienas kito atžvilgiu (Desrousseaux, Bernard, Mariage 2021). Tokios architektūros yra struktūrizuotos kompozicijos. Tam tikros architektūros mokymas apima taikymą suderinti veiksmus siekiant sureguliuoti neuronų svorius ir slenksčius. Neuroninių tinklų architektūros pagal disponavimo funkcija gali būti skirstomos į vieno sluoksnio, daugiasluoksnius, pasikartojančius ir apraizgytus tinklus (Rocha-Salazar, Segovia-Vargas, Camacho-Miñano 2021).

Įkvėptas biologinės nervų sistemos, neuronų tinklas sujungia kelis apdorojimo sluoksnius, naudodamas paprastus elementus, veikiančius lygiagrečiai. Tinklas susideda iš įvesties sluoksnio, vieno ar daugiau paslėptų sluoksnių ir išvesties sluoksnio.



16 pav. Neuronų tinklo architektūra. Sluoksniuotas (kairėje) ir nesluoksniuotas neuroniniai tinklai

Šaltinis: Math Word, 2023

16 paveiksle (kairėje) parodytas įėjimo sluoksnis. Šio sluoksnio paskirtis įvesti į tinklą įėjimo kintamųjų reikšmes. Toks pirmojo sluoksnio tinkle vaizdavimas yra įprastas, kadangi prieš duomenų apdorojimą neuroniniu tinklu dažnai seka pirminio duomenų apdorojimo etapas (pavyzdžiui, normavimas, centravimas). Pavyzdyje (žr. 16 pav.) paslėptų ir išėjimo sluoksnių neuronai sujungti su visais prieš tai esančio sluoksnio neuronais, bet neuronai gali būti sujungiami ne su visais gretimų sluoksnių neuronais. Vieno sluoksnio neuronai dažniausiai turi tą pačią perdavimo funkciją. Kiekviename sluoksnyje yra keli mazgai arba neuronai, o kiekvieno sluoksnio mazgai kaip įvestis naudoja visų ankstesnio sluoksnio mazgų išvestis, taip visi neuronai jungiasi vienas su kitu per skirtingus sluoksnius. Kiekvienam neuronui paprastai priskiriamas svoris, kuris koreguojamas mokymosi proceso metu. Svorio sumažėjimas arba padidėjimas keičia to neurono signalo stiprumą (Rocha-Salazar, Segovia-Vargas, Camacho-Miñano 2021).

Neuroninių tinklų sluoksnių atsakomybės (Desrousseaux, Bernard, Mariage 2021):

1. Įvesties sluoksnis (angl. Input layer) atsakingas už informacijos, signalų, funkcijų arba informacijos priėmimą.

2. Paslėpti, tarpiniai arba nematomi sluoksniai (angl. Hidden layer) atsakingi už duomenų ištraukimą, susiję su nagrinėjamu procesu ar konkrečia sistema. Šie sluoksniai atlieka daugybę vidinių procesų.

3. Išvesties sluoksnis (angl. Output layer) atsakingas už gamybą ir juo duomenys pateikiami galutiniai tinklo išėjimai.

Daugiasluoksnis neuroninis tinklas

Daugiasluoksnis neuroninis tinklas (*angl. Multi input Layer Neural Network*) dirbtinio neuroninio tinklo tipas, turintis daugiau nei vieną įvesties sluoksnį (Math Word, 2023). Tradiciniame neuroniniame tinkle yra tik vienas įvesties sluoksnis, kuris gauna įvestį iš išorinio pasaulio ir perduoda jį apdoroti į tinklą. Kelių įvesties sluoksnių neuroniniame tinkle yra keli įvesties sluoksniai, kurie gali priimti skirtingų tipų įvesties duomenis. Šis veiksmas leidžia tinklui tvarkyti sudėtingus įvesties duomenis, kuriuos gali sudaryti kelių tipų informacija arba šaltiniai. Pavyzdžiui, kompiuterinio matymo programoje kelių įvesties sluoksnių neuroniniame tinkle gali būti vienas įvesties sluoksnis vaizdo duomenims ir kitas teksto duomenims, apibūdinantiems vaizdą. Sujungus šiuos skirtingus įvesties duomenų tipus, tinklas gali atlikti tikslesnes prognozes ar klasifikacijas. Tada kiekvieno įvesties sluoksnio išvestis įvedama į paslėptą sluoksnį, kur tinklas apdoroja duomenis, kad sužinotų modelius ir ryšius tarp įvesties duomenų ir norimos išvesties. Galiausiai išvesties sluoksnis pateikia galutinį rezultatą.

Kelių įvesties sluoksnių neuroniniai tinklai yra galingas įrankis sudėtingiems įvesties duomenims tvarkyti ir gali būti naudojamas įvairiose srityse, įskaitant natūralios kalbos apdorojimą, vaizdo analizę bei kalbos atpažinimą (Malini, Pushpa, 2017; Ghobadi, Rohani, 2016).

Apibendrinant, neuroninis tinklas gali mokytis iš duomenų, todėl jį galima išmokyti atpažinti modelius, klasifikuoti duomenis ir numatyti būsimus įvykius. Neuroninis tinklas suskaido įvestį į abstrakcijos sluoksnius. Jis gali būti išmokytas naudojant daugybę pavyzdžių, kad atpažintų kalbos ar vaizdų modelius, kaip tai daro žmogaus smegenys. Neuroninio tinklo elgesį apibrėžia atskirų jo elementų prijungimo būdas ir tų ryšių stiprumas arba svoris. Šie svoriai treniruočių metu automatiškai koreguojami pagal nurodytą mokymosi taisyklę, kol dirbtinis neuroninis tinklas teisingai atlieka norimą užduotį.

2.2.2. Bajeso tinklų metodo aprašymas

Prieš pereinant prie Bajeso tinklo metodo būtina aptarti Bajeso metodą, iš kurio šis metodas išsivystė. Pasak Heckermano (1996) pirmą kartą Bajeso metodo sąvoka paminėta 1985 metais, nors Bajeso teorema jau žinoma nuo 1763. Bajeso teoriją aprašė Thomas Bayes, kuris ir yra šios teorijos išradėjas (Buzius, 2010). Dabar Bajeso metodas yra labai plačiai naudojamas įvairiems sprendimo priėmimo uždaviniams spręsti, tarp jų ir finansinio sektoriaus problemoms spręsti.

Bajeso metodas tampa vis svarbesniu metod. Heckermanas (1996) teigia, jog pirmiausia Bajeso tinklai atsirado ir išsivystė siekiant įvesti tikimybes į ekspertines sistemas ir tai vis dar yra vienas iš labiausiai naudojamų metodų. Specializuoti Bajeso tinklų variantai nepriklausomai naudojami daugelyje sričių, kaip pvz., genetikoje (ryšių analizėse), atpažinti

kalbai (angl., speech recognition), sekime, stebėjime, duomenų suspaudimui ir kodavimui, žaidimuose, kur įtraukiama tikimybė, įvairių duomenų bazių naršymui, robotikos kūrimui ir t.t. (Wiley, 2006.). Šio tiriamojo mokslinio darbo tikslas – panaudoti Bajeso metodą finansų srityje.

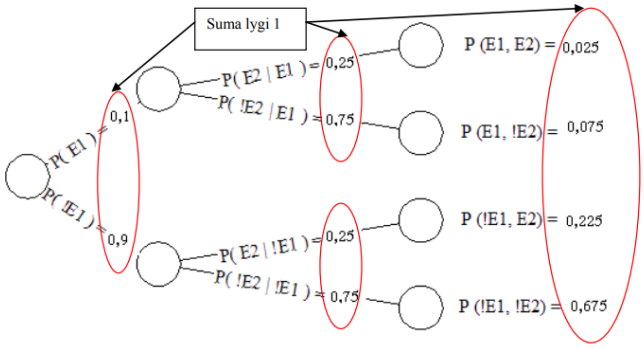
Toliau pateikiamas Bajeso teoremos pavyzdys (Bazius, 2010):

Analogiškai galima šią teoremą pritaikyti finansinio sukčiavimo atpažinimo tyrime: tikslas nustatyti ar operacija legali ar nelegali. tai tikimybė, kad operacija bus finansinis sukčiavimas maža, bet jeigu vartotojo operacija atitinka kelis požymius pavyzdžiui pervedimo suma yra neįprastai didelė, tai tikimybė didėja. Jei dar pridėsime, kad klientas paprastai neperveda labai didelių sumų, tai dar labiau padidina tikimybę, kad tai gali būti finansinis sukčiavimas (Song, et. al., 2020).

Pavyzdys, jog turime du įvykius – A ir B. Šie įvykiai yra tarpusavyje susiję, ir sakykime, jog B daro įtaką įvykio A atsiradimui ar buvimui. Norime sužinoti įvykio A tikimybę (P), žinant, jog įvykis B įvyko ar vyksta (B egzistavimas, buvimas) yra tiesa. Tai žymima taip – P(A|B). P(A|B) yra lygi bendrai įvykių B ir A įvykimo tikimybei padalintai iš įvykio B tikimybės. Ši formulė šiuo atveju yra labai paprasta, ji gali būti kur kas sudėtingesnės formos, pvz., įvykį A gali turėti įtaką ne vien tik įvykis B, bet ir C, D, E ir t.t. Kitas variantas: įvykiai, kurie turi įtaką, yra ne tik A, bet ir daugiau.

Skirtingai nuo naivaus Bajeso klasifikatoriaus, Bajeso tinklai atsižvelgia į atributų tarpusavio priklausomybę ir tam kuria kryptinius grafus su mazgų įtakojimu kitų buvimui (Buzius, 2010). Taigi naivaus Bajeso metodas yra paprastesnė Bajeso tinklų atmaina.

Pavyzdys du įvykiai E1 ir E2, atitinkamai pagal požymį, kad smarkiai viršija greitį blaivus ir kitas požymis - vairuoja dieną. Tarkime, kad vairuotojas jau padarė avariją, tada tikimybė, jog padaryta avarija todėl, kad smarkiai viršijo greitį blaivus yra 1 iš 10, t.y, 10%, analogiškai, jog neblaivus bus 90 % ir šį įvykį žymėsime priešingu įvykiui E1,!E1. Toliau, jei tikimybė, kad jis važiavo naktį, žinant, kad padarė avariją, viršydamas greitį neblaivus, yra 75% ir žymima P(!E2|E1), tai Bajeso tinklas atrodys kaip pavaizduota 17 paveiksle (žr. 17 pav.).



17 pav. Bajeso tinklas

Šaltinis: Buzius (2010)

Iš 17 paveikslo matome, kad didžiausia tikimybė, jog avarija padaryta yra dėl to, kad vairuotojas viršijo greitį išgėręs naktį, o mažiausia, jog viršijo greitį būdamas blaivus dieną, atitinkamai 67,5% ir 2,5%. Žinoma, šis pavyzdys yra labai elementarus, tačiau priežastinių, darančių įtaką įvykių gali būti žymiai daugiau. Vieni gali daryti įtaką arba visus likusius įvykius ar tik dalį jų, taip susidaro didelis ir painus tinklas su daug priežastinių ryšių.

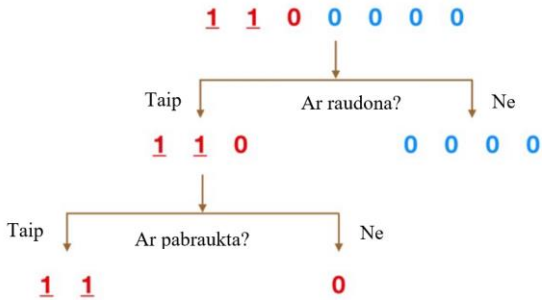
Apibendrinant, Bajeso tinklai leidžia modeliuoti įvykių tarpusavio priklausomybę naudojant tikimybes. Ši savybė suteikia galimybę nustatyti, kaip tikimybiniai įvykiai gali paveikti kitus įvykius finansiniame procese. Taip pat svarbu, tai jog naudojant Bajeso tinklus, galima nustatyti, kada transakcijos yra nukrypę nuo įprastų finansinių operacijų modelio. Tai padeda identifikuoti potencialias sukčiavimo atvejus, nes anomalijos gali rodyti, kad vyksta neįprastos ar nepatvirtintos transakcijos. Bajeso tinklai gali būti nuolat atnaujinami su naujais duomenimis ir įvykiais, todėl jie gali prisitaikyti prie besikeičiančių sąlygų. Šis modelis gali efektyviai tvarkyti ir analizuoti didelius duomenų rinkinius, kad būtų galima atpažinti potencialius sukčiavimo atvejus.

2.2.3. Sprendimo medžio metodo aprašymas

Sprendimo medis (angl. *Random forest*) algoritmas buvo pasirinktas, nes jis veikia kaip dvejetainis klasifikatorius, todėl jis tinkamas sukčiavimo aptikimui, kai bet kuri operacija priskiriama vienai iš dviejų (0 arba 1) sukčiavimo klasių (Badawi, 2021).

Pasak Salehi, Ghazanfari, Fathian (2017) ir Randhawa, et al. (2018) Sprendimo medžio algoritmo tikslas – sukurti modelį, numatantį tikslinio kintamojo reikšmę, kuriam sprendimų medis naudoja medžio atvaizdavimą, kad išspręstų problemą, kai lapo mazgas atitinka klasės etiketę, o atributai yra pavaizduoti vidiniame medžio mazge.

Pavyzdžiui, mūsų duomenų rinkinį sudaro skaičiai, esantys 18 paveiksle. Turime du 1 ir penkis 0 (1 ir 0 yra mūsų klasės) ir norime atskirti klases pagal jų savybes. Ypatybės yra spalvos (raudona arba mėlyna) ir tai, ar stebėjimas pabrauktas, ar ne.



18 pav. Sprendimų medis

Šaltinis: Kaggle duomenų bazė

Spalva atrodo kaip gana akivaizdi savybė, kurią reikia padalyti, nes visi 0, išskyrus vieną, yra mėlyni. Taigi galime naudoti klausimą -„Ar raudona?”. Galite manyti, kad mazgas medyje yra taškas, kuriame kelias dalijasi į dvi dalis – kriterijų atitinkantys stebėjimai eina žemyn šaka „Taip”, o tie, kurie nesileidžia šaka „Ne”.

Šaka „Ne” turi tik vieną atsakymą - 0, todėl šioje vietoje tyrimas baigiasi, bet mūsų „Taip”šaka vis tiek gali būti padalinta toliau. Dabar galime naudoti antrąją funkciją ir paklausti: „Ar ji pabraukta? ir padaryti antrą padalijimą.

Sprendimo medžio parametrai (Salehi, Ghazanfari, Fathian, 2017):

- **Entropija**: Entropija valdo, kaip sprendimų medis nusprendžia padalyti duomenis. Tai turi įtakos tam, kaip sprendimų medis nubrėžia savo ribas. Entropijos reikšmės svyruoja nuo 0 iki 1. Kuo mažesnė entropijos vertė, tuo labiau metodu galima pasitikėti.

  Entropijos formulė:

  $$H(s) = probablity\ of\ log_2\,(p+) - probability\ of\ log_2\,(p-) \qquad (1)$$

  kai, p+ yra pozityvi klasė ir p- yra negalyti klasė.

- **Informacijos gavimas**: Informacijos gavimas naudojamas norint nuspręsti, kurią savybę padalyti kiekviename medžio kūrimo etape. Geriausias yra paprastumas, todėl norime, kad mūsų medis būtų mažas. Norėdami tai padaryti, kiekviename žingsnyje turėtume pasirinkti padalijimą, dėl kurio gaunami gryniausi dukteriniai mazgai. Dažniausiai naudojamas grynumo matas vadinamas informacija.

  Kiekvienam medžio mazgui informacijos reikšmė parodo, kiek informacijos apie klasę mums suteikia ypatybė. Padalijimas su didžiausiu informacijos padidėjimu bus laikomas pirmuoju padalijimu ir procesas tęsis tol, kol visi antriniai mazgai bus švarūs arba tol, kol informacijos padidėjimas bus 0.

  $$\mathrm{Gain(S, A)} = H(s) \sum_{k=0}^{n} \binom{n}{k} \frac{|Sv|}{|s|} H(Sv) \qquad (2)$$

- ***Gini Inpurity*** yra matavimas, naudojamas sprendimų medžiams kurti, siekiant nustatyti, kaip duomenų rinkinio ypatybės turėtų padalinti mazgus, kad sudarytų medį. Tiksliau, duomenų rinkinio Gini priemaiša yra skaičius nuo 0 iki 0,5, o tai rodo tikimybę, kad nauji atsitiktiniai duomenys nebus klasifikuojami, jei jiems būtų suteikta atsitiktinė klasės etiketė pagal klasės pasiskirstymą duomenų rinkinyje.

Apibedrinant, sprendimo medžio metodas pasižymi aukštu atpažinimo tikslumu. Jis gali efektyviai atskirti sukčiavimo atvejus nuo įprastų finansinių transakcijų, nes jis apmokamas iš

gausios duomenų rinkinio įvairovės. Taip pat šis metodas gali nustatyti, kurie duomenų požymiai ar veiksniai yra svarbiausi sukčiavimo atpažinimui, tai leidžia įmonėms efektyviai koncentruotis į svarbiausius rizikos veiksnius.

2.2.4. „Catboos“t metodo aprašymas

„CatBoost“ yra sprendimų medžių gradiento didinimo algoritmas, kuris pasižymi didesniu tikslumu už logistinės regresijos, gradiento didinimo bei šiek tiek didesniu tikslumu už atsitiktinių miškų algoritmus (Chen, Han, 2021; Barua et. al. 2021).

Chen, Han (2021) ir Barua el. al. (2021) išskyria „CatBoost“ algoritmo privalumus:

- Gerai tvarkosi su kategoriniais kintamaisiais.
- Greitesnis lyginant su kitais gradiento didinimo algoritmais.
- Puikiai tinka didelėms duomenų imtims.

Pasak Nguyen et al. (2022) „CatBoost“ patenka į populiariausių prižiūrimos klasifikacijos algoritmų sąrašą, nes sėkmingai sprendžia statistines problemas, su kuriomis susiduria kiti esami pažangiausi gradiento didinimo algoritmai tokie kaip „XGBoost“, Lightboost.

Prokhorenkova et al. (2018) paaiškina „CatBoost“ algoritmo veikimą:

- „CatBoost“, prognozavimo modelį F, gautą po kelių padidinimo žingsnių, greičiausiai patirs reiškinį, vadinamą „prognozavimo poslinkiu“, kuris yra F(xk) pasiskirstymo poslinkis | xk mokymo pavyzdžiui xk iš F(x) | skirstinio x bandymo pavyzdžiui x . Autorius atrado šią problemą remdamasis hipoteze, kad egzistuoja duomenų rinkinys D={(xk,yk)}k=1..n, , kur xk=(x1k,...,xmk) yra atsitiktinis m elementų vektorius. ir yk∈ R yra dvejetainis tikslinis kintamasis. Mėginiai (xk , yk) yra pasiskirstę nepriklausomai ir identiškai pagal skirstinį P(·,·) . Mokymosi užduoties tikslas – išmokyti funkciją H: Rm → R, kuri sumažina tikėtinus nuostolius: L(F):=EL(y,F(x)) kur L(·,·) yra sklandaus praradimo funkcija ir (x, y) yra testavimo duomenys, paimti iš mokymo duomenų D.
- Gradiento didinimo procedūra iteratyviai sukuria aproksimacijų seką Ft:Rm→R,t=0,1,… godžiai. Remiantis ankstesniu aproksimavimu Ft−1 , Ft gaunamas adityviu būdu, kad Ft=Ft−1+αht kur α yra žingsnio dydis ir funkcija ht:Rm→R (bazinė prognozė), kad sumažintų nuostolių funkciją:

Be to, paskirstymo poslinkis taip pat gali įvykti iš anksto apdorojant kategorines savybes konvertuojant jas į tikslinę statistiką. Tikslinė statistika yra paprastas statistinis modelis, kuris taip pat gali sukelti tikslo nutekėjimą ir prognozės. Autoriai sukūrė naują didinimo algoritmą, vadinamą užsakytu didinimu, kuris primena užsakyto tikslo statistikos metodą. „„„CatBoost““

taip pat turi kitą padidinimo režimą, vadinamą „paprastu“, kuris yra standartinis GBDT algoritmas su integruota užsakyta tikslinės statistikos statistika. Medžio kūrimo „CatBoost“ procedūra aprašyta pseudokode.

Užsakytame didinimo režime mokymosi proceso metu palaikome pagalbinius modelius Mr,j , kur Mr,j(i) yra dabartinė i-ojo pavyzdžio prognozė, pagrįsta pirmaisiais j pavyzdžiais permutacijoje σr . Kiekvienoje algoritmo iteracijoje t atrenkame atsitiktinę permutaciją σr iš {σ1 ,… , σs } ir pagal šią permutaciją sukuriame medį Tt (Awoyemi, 2017). Pirma, kategoriškiems požymiams visa tikslinė statistika apskaičiuojama pagal šią permutaciją. Antra, permutacija turi įtakos mokymosi medžiui procedūrai. Paprastame režime, jei yra kategorinių požymių, jis palaiko pagalbinius modelius Mr, atitinkančius tikslinę statistiką, pagrįstą σ1 ,… , σs .

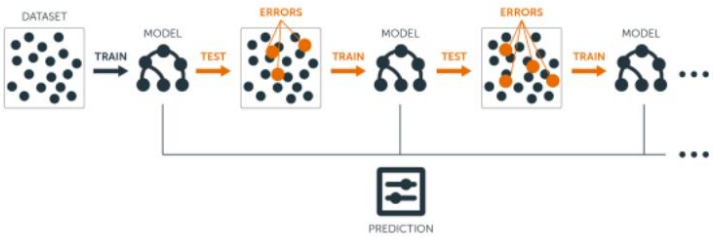
Apibendrinant, „CatBoost“ pagrindinės prognozės yra nepastebimi sprendimų medžiai, kurie yra medžiai, suskirstyti pagal nuoseklius kriterijus visame lygyje. Tokie medžiai yra subalansuoti, leidžia greitai atlikti bandymą. Norėdami įrodyti „CatBoost“ efektyvumą, autoriai palygino jį su kitais gradiento didinimo algoritmais, įskaitant „XGBoost“ ir „LightGBM“ (Nguyen, et. al. 2022). Rezultatai parodė, kad panašaus dydžio ansambliuose „CatBoost“ galima įvertinti maždaug 25 kartus greičiau nei „XGBoost“ ir maždaug 60 kartų greičiau nei LightGBM.

2.2.5 „XGBoost“ metodo aprašymas

„XGBoost“ metodas yra pastaraisiais metais praktikoje vienas iš labiausiai pritaikomų algoritmų mašininiame mokyme (Chen, Guestrin, 2016).

Šis metodas veikia treniruojant didelį skaičių sprendimo medžių, kiekvienas medis yra treniruojamas su duomenų imtimi, o galutinė prognozė yra daroma atsižvelgiant visų medžių prognozes (XGBboost developer, 2020). Taigi, principas yra toks pats, kaip ir atsitiktinio miško, skirtumas yra tas, kad šiame metode yra naudojamas gradientinis medžių didinimas, o kad jį suprasti, reiktų trumpai susipažinti, kas yra medžių didinimas (ang. *Boosting*) ir gradientinis didinimas (ang. *Gradient boosting*). Medžių didinimas yra iteracinis metodas, kuris remiasi ansamblio technika, nes daugelis medžių yra sujungiami kartu, kad būtų gaunamas galutinis rezultatas. Tačiau medžių didinimas skiriasi tuo, kad užuot mokę visus medžius atskirai vienas nuo kito, didinimas treniruoja modelius iš eilės, o kiekvienas naujas modelis mokomas, siekiant ištaisyti ankstesnių medžių padarytas klaidas. Medžiai yra pridedami atskirai, kol nebeišeina atlikti patobulinimų (Friedman, 2001).

„XGBoost“ naudoja gradientinį didinimą, kurio principas yra toks, kad nauji modeliai yra mokomi numatant ankstesnių modelių klaidas, 19 paveiksle (žr. 19 pav.) yra matomas metodo procesas.



19 pav. Medžių mokymas gradientiniu metodu

Šaltinis: XGBboost developer (2020)

O nueinant labiau į detales, šį metodą galima aprašyti per kelis žingsnius(XGBboost developer, 2020):

1. Suskaičiuoti tikslo kintamojo vidurkį $\bar{y}$.

2. Apskaičiuoti liekanas $\epsilon i = \bar{y} - \hat{y}i$, kur $\hat{y}i$ šiame žingsnyje atitinka tikras reikšmes.

3. Sukonstruojamas sprendimų medis, kuriame kiekvienas lapas, t.y. prognozuojama reikšmė atitinka liekanos reikšmes, o jeigu į lapą papuola daugiau negu viena reikšmė, imamas tų reikšmių vidurkis.

4. Atliekama prognozė, naudojant medžius iš ansamblio, t.y. (po pirmos iteracijos imamas vienas medis, po antros pirmas ir antras ir t.t.), nauja prognozė yra $\hat{y}i = \bar{y} + \alpha\epsilon i$, kur α yra hiperparametras, nurodantis mokymosi greitį.

5. Perskaičiuojamos naujos klaidų reikšmės, kurios yra naudojamos naujame medyje kaip reikšmės, esančios lapuose.

6. Kartojami 3-5 žingsniai tiek kartų, kiek nurodyta hiperparametre.

7. Apskaičiuoti galutinę prognozę, naudojant visą medžių ansamblį.

Taigi, tiek „XGBoost“, tiek gradientinis medžių didinimas remiasi gradientinio didinimo principu, tačiau skirtumas yra jau modeliavimo detalėse, tiksliau, „XGBoost“ metodas naudoja labiau reguliarizuotą modelio formulizavimą, kad išvengtų modelio persotinimo, o tas užtikrina didesnį modelio našumą (Chen & Guestrin, 2016).

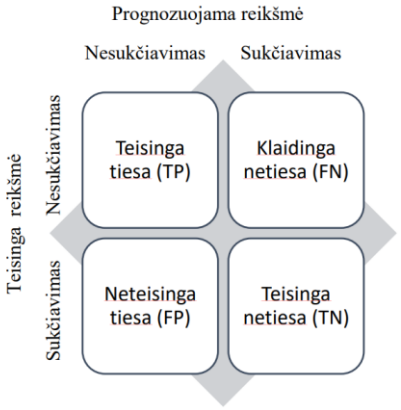
2.3. Modelio vertinimo parametrai

2.3.1. Klaidų matrica

Siekiant įvertinti pasirinkto modelio tikslumą yra lyginamos tarpusavyje prognozuojamos ir tikrosios reikšmės pagal klaidų matrica (angl. Confusion matrix).

Ši matrica vizualiai pavaizduoja gautus duomenis pagal 4 skirtingas prognozuojamas ir faktines (teisingas) verčių kombinacijas (Narkhede, 2018):

- neteisinga tiesa (FP) (angl. False Positive): naudojama apibūdinti teigiamas prognozes, kurios iš tikrųjų yra neigiamos;
- teisinga tiesa (TP) (angl. True Positive): naudojama apibūdinti teigiamas prognozes, kurios iš tikrųjų yra teigiamos;
- klaidinga netiesa (FN) (angl. False Negative): naudojama norint apibūdinti neigiamas prognozes, kurios iš tikrųjų yra teigiamos;
- teisinga netiesa (TN) (angl. True Negative): naudojama apibūdinti neigiamas prognozes, kurios iš tikrųjų yra neigiamos.



20 pav. Klaidų matrica

Šaltinis: Narkhede (2018)

Pateikta klaidų matrica parodo, kaip galėtų būti interpretuojami skaičiai, jei matricos duomenys būtų susiję su pinigų plovimu pagal prognozuojamas ir faktines reikšmes (žr. 21 pav.)

2.3.2. Duomenų klasifikavimo tikslumas

Siekiant nustatyti klasifikavimo tikslumą yra naudojami ir kiti matai (Tichonov, 2018):

- bendras klasifikavimo tikslumas (angl. Accuracy) rodo teisingai suklasifikuotų ir visų duomenų santykį, ir gali būti naudojamas tada, kai esančios klasės subalansuotos;

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \quad (4)$$

- tikslumas (angl. Precision) - gali parodyti, kiek prognozuotų reikšmių yra teisingi rezultatai;

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

- jautrumas (angl. Recall) - gali parodyti, kokia dalis buvo teisingai prognozuota teigiamų reikšmių;

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

- F1 rodiklis (angl. F1 score) yra reikalingas, kai norime rasti pusiausvyrą tarp tikslumo ir jautrumo, ir kai klasių paskirstymas yra netolygus - didelis skaičius faktinių neigiamų reikšmių (Flach et al., 2015).

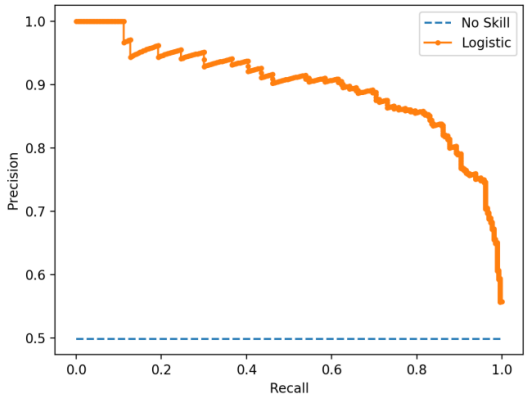
$$F1 = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (7)$$

Apibendrinant, norint įvertinti metodą būtina įvertinti jo tiklsumą, efektyvumą.

2.3.3. Modelio našumo vertinimas

Modelio našumą gali padėti įvertinti precision-recall (AUC) kreivė, ypač tais atvejais, kai vienos iš klasių yra daugiau nei kitos (Ling, Huang, Zhang, 2006).

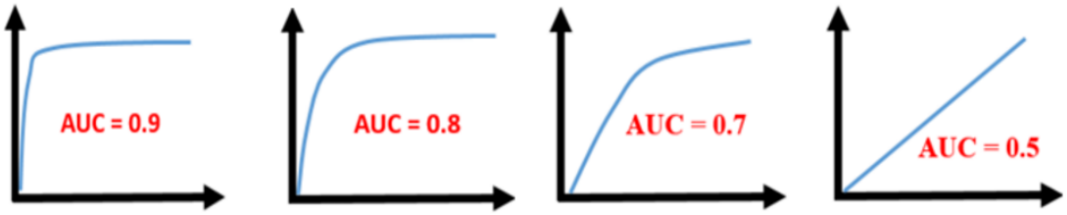
Ši kreivė atvaizduoja santykį tarp tikslumo ir atšaukimo (angl. recall), kai keičiasi klasifikatoriaus sprendimo riba (arba slenkstis). Precision yra santykis tarp tikrų teigiamų atvejų (TP) ir visų teigiamų atvejų (TP + FP), o recall yra santykis tarp tikrų teigiamų atvejų (TP) ir visų tikrų teigiamų atvejų (TP + FN) (žr. 21 pav.).



21 pav. Preciziškumo - atkūrimo kreivė

Šaltinis: Kaggle duomenų bazė

Pasak Linko ir jo kolegų (2006) Precision-recall kreivė padeda vizualiai įvertinti, kaip keičiasi klasifikatoriaus tikslumas ir atšaukimas priklausomai nuo sprendimo ribos (slenksčio) (žr. 22 pav.).



22 pav. AUC kreivės

Šaltinis: D'Souza (2018)

Siekiant įvertinti sukurto modelio rezultatus, blogiausias AUC egzistuoja, kai siekia 0, o geriausias AUC, kai yra arti 1. Kuo didesnė AUC vertė, tuo geresnis modelio našumas ir tuo tiksliau jis atlieka klasifikavimą.

2.4. Išlaidų taupymas pritaikius finansinio sukčiavimo atpažinimo metodus veikloje

Kiekvienas neatpažintas finansinio sukčiavimo atvejis finansų sektoriuje veikiančiai įmonei atneša nuostolį, nes teisingas finansinio sukčiavimo sandorio nustatymas leidžia sutaupyti nurodytas išlaidas (formulė 8). Šiame darbe bus naudojamas išlaidų taupymo metodika, kuris remiasi pelnu pagrįstos paskolos įsipareigojimų nevykdymo prognozavimo metodu (Hajek, Abedin, Sivarajah, 2022).

Išlaidų taupymas teisinga tiesa apskaičiuojama pagal formulę:

$$CS_{TP} = \sum_{k=0}^{n} \binom{n}{k} (TP_i * A_i * 3{,}36) - (TP_B * A_F * 3{,}36) \qquad (8)$$

Formulėje:

- CS_{TP} yra išlaidų sutaupymas iš teisingos tiesos (true positive) transakcijų;
- TP_i yra i-oji TP operacija;
- A_i yra i-osios operacijos suma;
- TP_B yra TP operacijų, sukčiavimo aptikimo sistemos skaičius;
- A_F yra vidurkis nesąžiningų sandorių suma.

Taip pat šios formulės autorius siūlo viską dauginti iš 3,36, nes 2021 m. duomenimis US įstaigoms sukčiavimas kainavo už kiekvieną dėl sukčiavimo prarastą dolerį 3,36 $. TP_B įvardija, jog vidutinis sėkmingų bandymų sukčiauti procentas yra 48 % (t. y. TPB = 0,52).

Iš kitos pusės pagal US duomenis pinigų pervedimai per mobiliąsias programėles sugeneruoja metinę maržą, kuri siekia 3,5 proc. (Bansal et al., 2019), todėl turėtų būti atsižvelgta ir neteisingos tiesos sandorio kainą, apskaičiuotą kaip šių operacijų maržos sumažėjimą.

Legalios transakcijos įvertinimo kaip finansinis sukčiavimas (Neteisingos tiesos) kaina apskaičiuojama pagal formulę:

$$Cost_{FP} = (TN * A_L * 0{,}035) - \sum_{k=0}^{n} \binom{n}{k} (FP_i * AT_i * 0{,}035) \quad (9)$$

Formulėje:

- $Cost_{FP}$ yra neteisingos tiesos (FP) operacijų kaina;
- TN yra teisingai aptiktas finansinio sukčiavimo sandoris
- FP_j yra j-oji FP operacija;
- A_L yra vidutinė teisėtų operacijų suma;
- AT_j yra j-osios operacijos suma.

Bendra sutaupytų išlaidų suma apskaičiuojama pagal formulę:

$$CS_{total} = CS_{TP} - Cost_{FP} \quad (10)$$

Šis metodas bus taikomas tyrimo dalyje siekiant apskaičiuoti finansinę naudą įmonei įdiegus į sako finansinio sukčiavimo aptikimo sistemą.

3. FINANSINIO SUKČIAVIMO ATPAŽINIMO TYRIMAS

Finansinis sukčiavimas yra įsisenėjusi ir labai aktuali tema finansų sektoriuje. Dauguma automatizuotų algoritmų turi didelį klaidingų teigiamų rezultatų rodiklį (angl. *false positive rate*), kai teisėtos operacijos neteisingai pažymėtos kaip finansinis sukčiavimas. Taip pat didelė problema – klaidingi neigiami duomenys (angl. *false negatives*) t. y. neaptikti ne legalūs sandoriai. Šiai problemai išspręsti yra naudojama daug dirbtinio bei skaitmeninio intelekto algoritmų, kurie iš istorinių duomenų gali atpažinti galimą finansinį sukčiavimą.

Finansinio sukčiavimo atpažinimo tyrime naudojami sintetiniai mobilios programėlės „Paysim“ pinigų pervedimai iš duomenų bazės Kaggle. „PaySim“ duomenų rinkinys imituoja pinigų pervedimą mobiliojoje programėlėje. Duomenys buvo sugeneruoti pagal tarptautinės įmonės „Paysim“, kuri teikia finansinių paslaugų atsiskaitymą per mobiliąją programėlę, realų operacijų pavyzdį.

7 lentelė. Analizuojamų duomenų aprašymas

Duomenys	Duomenų aprašymas
step	Atvaizduoja laiko vienetą realiame pasaulyje. Šiuo atveju 1 žingsnis yra 1 valanda laiko. Iš viso žingsnių 744 (30 dienų modeliavimas).
type	Pinigų pervedimo tipas: grynųjų įnešimas, grynųjų išsigryninimas, debetas, mokėjimas su prekes ar paslaugas, pinigų pervedimas.
amount	Pinigų suma
nameOrig	Kliento kodas sistemoje.
oldbalanceOrg	Pradinis lėšų balansas prieš pinigų pervedimo.
oldbalanceDest	Pradinis pinigų gavėjo lėšų balansas pinigų gavėjo prieš operaciją.
newbalanceOrg	Lėšų likutis po pinigų pervedimo.
newbalanceDest	Naujas pinigų gavėjo balansas po pinigų pervedimo.
nameDest	Klientas, kuris yra pervedamų pinigų gavėjas.
isFraud	Sistema fiksuoja finansinį sukčiavimą, kai klientas visus turėtus pinigus perveda į kitą sąskaitą arba išsigrynina visus sąskaitoje turėtus pinigus.
isFlaggedFraud	Sistema siekia kontroliuoti didžiulius pervedimus iš vienos sąskaitos į kitą, todėl bandymas pervesti daugiau nei 200 000 per vieną operaciją žymima, kaip galim nusikalstama veikla.

Šaltinis: sudaryta autorės remiantis Kaggle duomenų baze

Prieš pereinant prie skaitmeninio intelekto modelių skaičiavimo kitame skyriuje išanalizuosime duomenis.

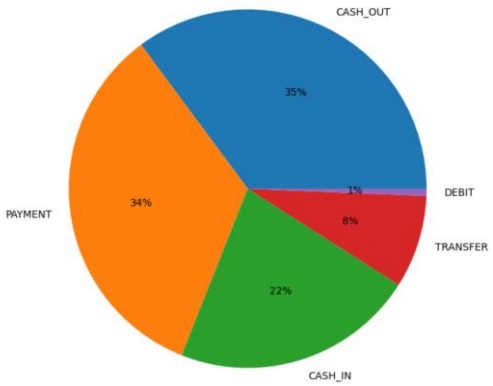
3.1. Finansinio sukčiavimo atpažinimo tyrimo duomenų pristatymas ir anomalijų tikrinimas

Šiame skyriuje bus vizualiai analizuojami pasirinkti mobiliosios atsiskaitymų programėlės „Paysim“ duomenys. Duomenų rinkinį sudaro 6362620 eilučių, todėl duomenų vizualiam atvaizdavimui naudojama Python programavimo kalba.

Pirmieji žingsniai paruošiant duomenis:

1. Patikrinama ar nėra nulinių reikšmių;
2. Patikrinama ar nėra dublikatų;
3. Patikrinama ar nėra išbalansavimo duomenyse.

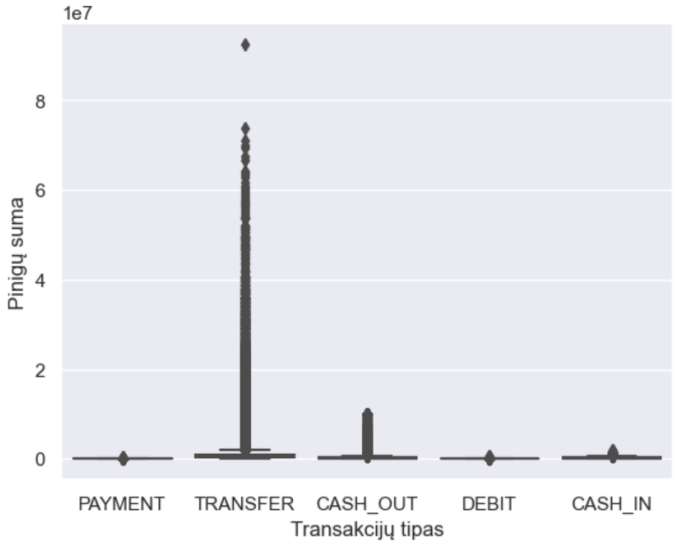
Pradedame žvalgomąją duomenų analizę (angl. EDA - Exploratory Data Analysis), kad geriau suprastume duomenų rinkinį nuo transakcijų tipų pasiskirstymo (žr. 23 pav.).



23 pav. Transakcijų tipas

Šaltinis: sudaryta autorės

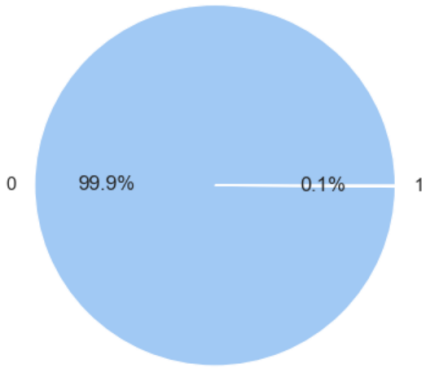
Analizuojant transakcijų kiekį matome, jog daugiau nei trečdalį t. y. 34 proc. visų transakcijų sudarė atsiskaitymas už prekes ir paslaugas, 35 proc. pinigų išsigryninimas. Tuo tarpu pervedimų kiekis sudarė 8 proc. nuo viso kiekio, tačiau 24 paveikle (žr. 24 pav.) matome, jog pervedimai sudaro didžiausią pinigų sumą.



24 pav. Transakcijų tipai vertinant pagal pinigų sumą

Šaltinis: sudaryta autorės

Transakcijų tipų grafikas leidžia daryti prielaidą, jog pervedimo į kitą sąskaitą sumos yra ženkliai didesnės nei kitų transakcijų tipų.

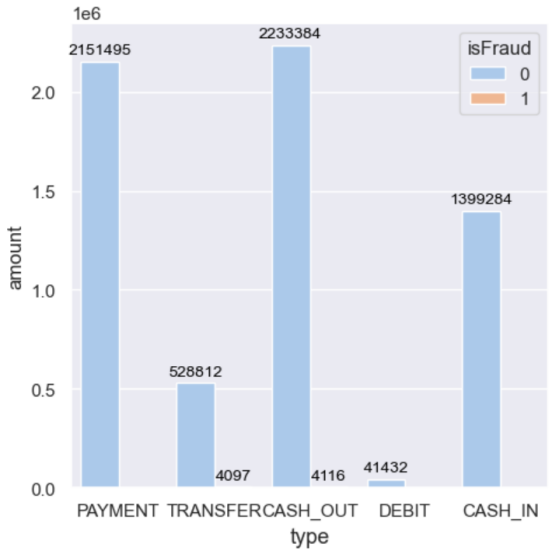


25 pav. Neteisėtų ir teisėtų pervedimų pasiskirstymas.

Šaltinis: sudaryta autorės

Analizuojamame istorinius duomenis neteisėta veikla buvo fiksuojama 0,1 proc. nuo visų pervedimų. Laiku neaptikus finansinio sukčiavimo, tai tampa įmonės išlaidomis. Taip pat 25 paveikslas rodo, jog duomenys yra nesubalansuoti. Plačiau apie šio duomenų rinkinio duomenų subalansavimą buvo rašome 2.1.2 dalyje.

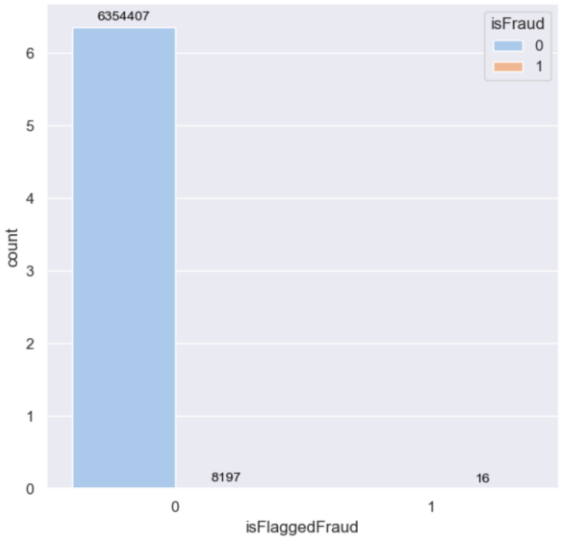
27 pav. matome pinigų pervedimo tipo teisėto ir neteisėto operacijų skaičių. Šiame grafike išsiskiria du operacijų tipai – pervedimas į kitą sąskaitą ir pinigų išsigryninimas, kuriuose buvo fiksuota nelegali veikla.



26 pav. Pinigų pervedimo tipo teisėtos ir neteisėtos operacijos

Šaltinis: sudaryta autorės

Taip pat iš 26 paveikslo matome, jog finansinio sukčiavimo atvejų skaičius tiek pinigų pervedimas į kitą sąskaitą, tiek pinigų išsigryninimas sudaro labai panašų skaičių.

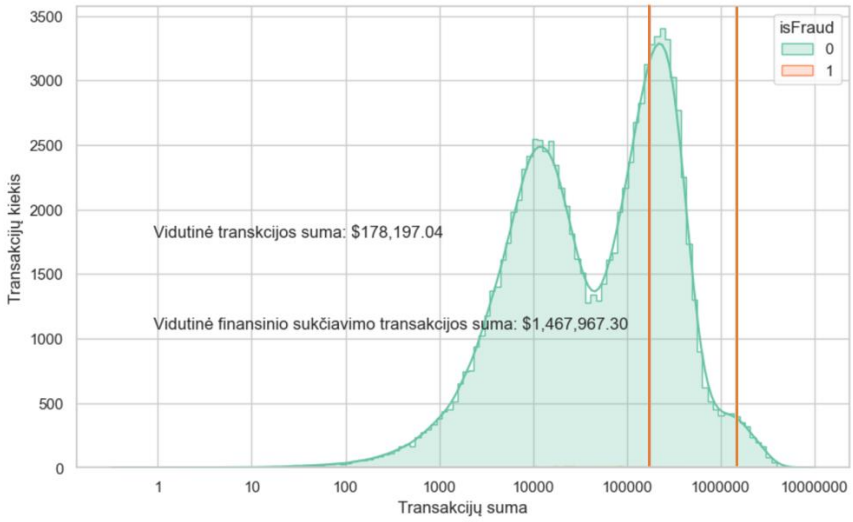


27 pav. Rizikingų transakcijų palyginimas su finansiniu sukčiavimu

Šaltinis: sudaryta autorės

Aptariant rizikingų transakcijų palyginimą su finansinio sukčiavimo atvejais (žr. 27 pav.) iš visų operacijų tik 16 buvo pažymėtos kaip sukčiavimas. Toks mažas skaičius rodo, kad sistema retai aptinka visas nesąžiningas operacijas. Tačiau, rizikingų transakcijų palyginimo su finansiniu sukčiavimu grafike matome, jog visos identifikuotos kaip rizikingos transakcijos, kurių pinigų suma buvo didesnė nei 200 000 vietos valiuta buvo finansinis sukčiavimas.

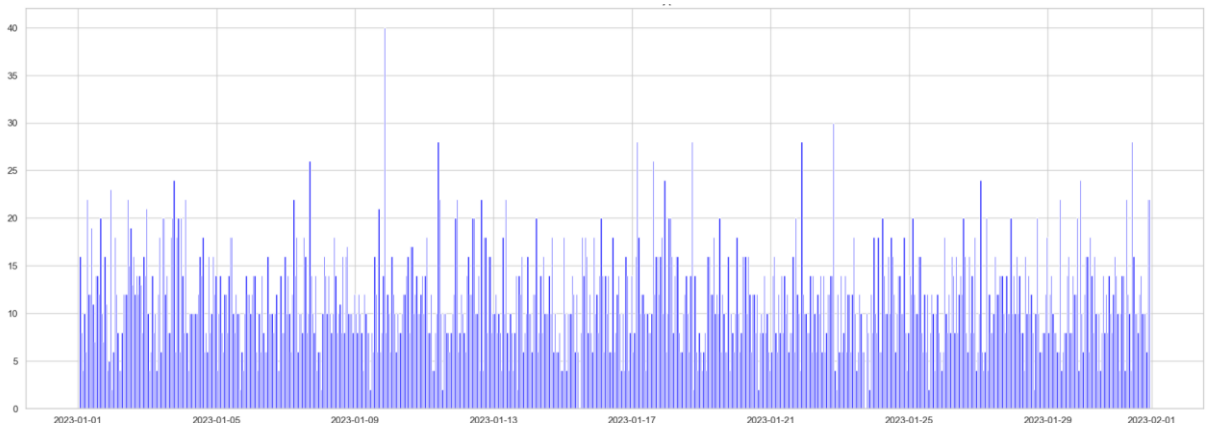
Taip pat svarbu suprasti ar transakcijos, kurios buvo identifikuotos kaip finansinis sukčiavimas išsiskiria nuo legalių pervedimų. 29 pav. matome, jog vidutinė transakcija yra 178197 vietos valiuta. Tuo tarp finansinio sukčiavimo vidutinė suma yra beveik dešimt kartų didesnė t. y. 1467967 vietine valiuta.



28 pav. Transakcijų sumų pasiskirstymas

Šaltinis: sudaryta autorės

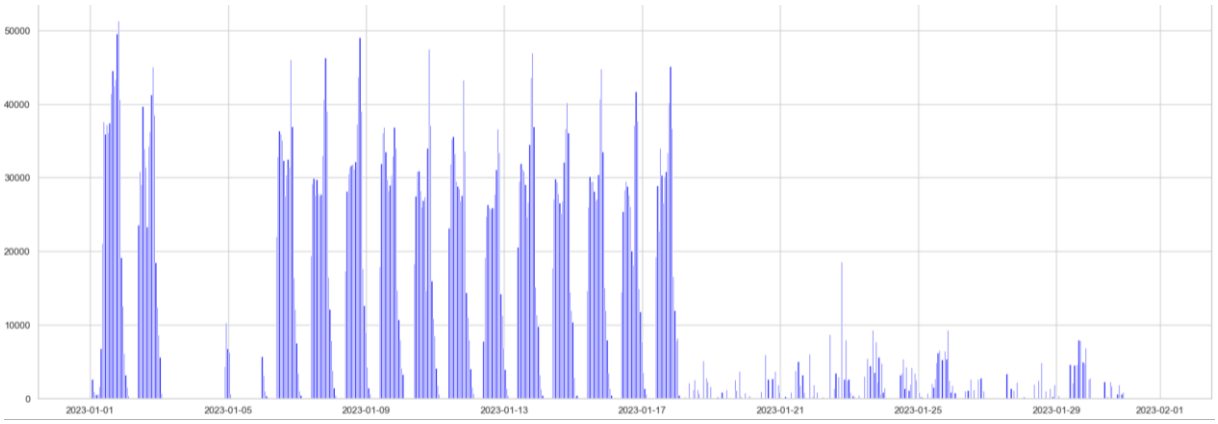
Pristatant duomenis buvo minėta, kad 1 žingsnis (angl. step) yra lygus 1 val., o iš viso yra 744 žingsniai, kas sudaro 30 d. laikotarpį.



29 pav. Finansinio sukčiavimo transakcijų kiekis tiriamu periodu

Šaltinis: sudaryta autorės

29 paveiksle (žr. 29 pav.) pavaizduotas finansinių transakcijų kiekis nuo 2023.01.01 iki 2023.02.01 matome, kad lėšų pervedimai yra pasiskirstę tolygiai, nes vykdant nelegalius veiksmus pervedant pinigus į kitą sąskaitą ar išsigryninant juos, tai daroma nenuosekliai, didesnėmis sumomis, kad greičiau būtų pasisavinti svetimos lėšos.



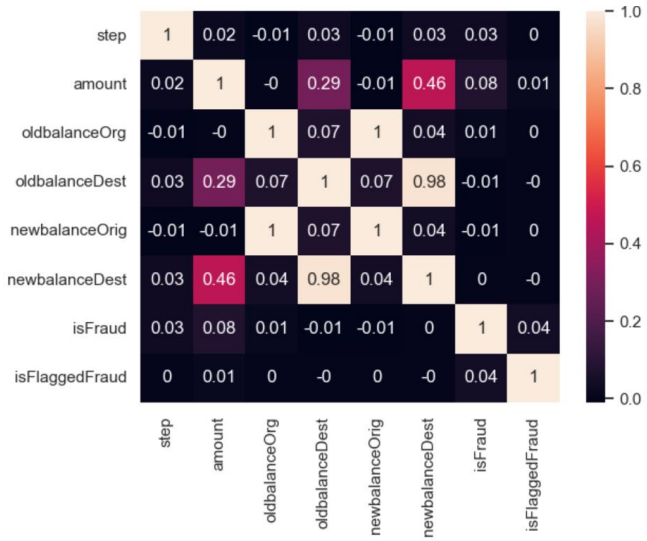
30 pav. Legalių transakcijų kiekis tiriamu periodu

Šaltinis: sudaryta autorės

Tuo tarpu legalių transakcijų grafike (žr. 30 pav.) pastebimas transakcijų šablonas, t. y. įprastai žmonės tam tikru metu gauna įplaukas (darbo užmokestis, pensija, stipendija ir kita) į sąskaitą. Taip pat panašiu metu apmoka sąskaitas, perka prekes.

Analizuojant duomenis svarbu patikrinti ryšį tarp kintamųjų. Tačiau, prieš tai turime patikrinti ar nėra duomenyse susidubliavusių reikšmių bei nulinių reikšmių.

Jei yra stipri koreliacija tarp dviejų kintamųjų, tai gali reikšti, kad vieno kintamojo reikšmė gali būti naudojama prognozuoti kitą.



31 pav. Ryšys tarp kintamųjų

Šaltinis: sudaryta autorės

Sudarius koreliacinę matricą (žr. 31 pav.) matome, jog vidutinė koreliacija yra tarp sumos (angl. amount) ir pradinio pinigų gavėjo lėšų balanso. (angl. oldbalanceDest) ir sumos

(angl. amount) ir naujo pinigų gavėjo lėšų likučio po pinigų pervedimo balansas (angl. newbalanceDest).

Abibendrinant galime teigti, kad finansinis sukčiavimas buvo transakcijų pervedime ir pinigų išsigryninime. Taip pat visos pažymėtos rizikingos transakcijos, kurių pinigų suma buvo didesnė nei 200 000 vietos valiuta buvo užfiksuoti kaip finansinis sukčiavimas. Iš duomenų matoma, kad finansinio sukčiavimo transakcijų sumos kelis kartus didesnės nei legalios transakcijos. Šis duomenų atvaizdavimo metodas tik dar kartą patvirtino, kad transakcijos pažymėtos, kaip finansinis sukčiavimas buvo didesnių sumų, tai galima aiškinti, jog nusikaltėliai užgrobę sąskaitą, nori kuo greičiau pasisavinti visus pinigus. Apskaičiavus koreliaciją matome, jog vidutinis ryšys yra tarp sumos ir pinigų gavėjo balanso, tai būdinga sukčiavimui, nes nusikaltėlių tikslas pervesti arba išgryninti visus sąskaitoje esančius pinigus.

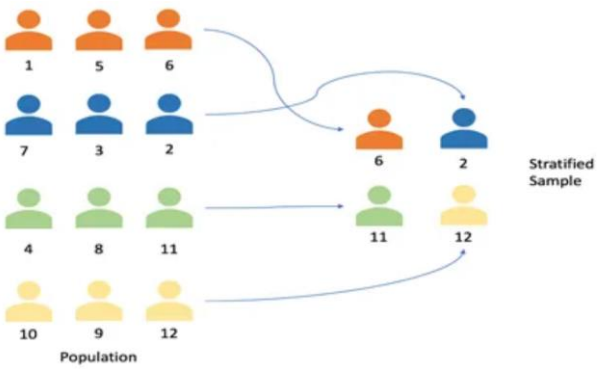
3.2. Skaitmeninio intelekto metodų apskaičiavimas finansinio sukčiavimo atveju

3.2.1. Duomenų apdorojimas

Kitas žingsnis, prieš pradėdami taikyti modelių mokymą yra apdoroti duomenis, kad įsitikintumėte, jog jie yra tinkamo formato. pasirinktiems skaitmeninio intelekto metodams ir mašininio mokymosi algoritmams.

Duomenų apdorojimui atliekami šie žingsniai:

- Nenaudojami stulpeliai ištrinami. Iš duomenų modelio ištriname šiuos stulpelius - *step, nameOrig, nameDest*;
- Stulpeliai *isFraud* ir *isFlaggedFraud* duomenų tipai konvertuojami į loginį tipą – 0 ir 1, kai 1 reiškia finansinį sukčiavimą.
- Užkoduojami kategoriniai stulpeliai: *amount, oldbalanceOrg, newbalanceOrg, oldbalanceDest, newbalanceDest.*
- Suskirstomi duomenys į mokymo ir testavimo rinkinius. Stratifikuota atranka naudojama siekiant užtikrinti, kad abiejų rinkinių (mokymo ir testavimo) tikslinio kintamojo pasiskirstymas būtų panašus (žr. 32 pav.). Testavimo dydis 0,20.



32 pav. Stratifikavimo pavyzdys

Šaltinis: Analyttica Datalab's Medium (2021)

Kaip matome iš aukščiau pateikto paveikslėlio šis metodas padeda padalyti į mokymo ir testavimo rinkinius, užtikrinant, kad klasių pasiskirstymas testų rinkinyje tiksliai atitiktų mokymo rinkinio klases.

3.2.2. Skaitmeninio intelekto metodų vertinimas

Daugiasluoksnio Neuroninio tinklo metodo vertinimas

Kelių įvesties sluoksnių neuroninis tinklas yra dirbtinio neuroninio tinklo tipas, turintis daugiau nei vieną įvesties sluoksnį, tai leidžia tinklui tvarkyti sudėtingus įvesties duomenis, kuriuos gali sudaryti kelių tipų informacija arba šaltiniai. Šioje tyrimo dalyje analizuojama 5 įvesčių neuroninis tinklas.

Phyton programoje buvo naudojamas *MLPClassifier* funkcijų paketas.

Naudoti derinimo parametrai:

• Paslėptasis sluoksnis (hiddenSize) – 5;

• Maksimalus iteracijų skaičius (maxit) – 149124.

```
model.compile(optimizer='adam',
              loss='binary_crossentropy',
              metrics=['accuracy'])

model.fit(X_train, y_train, epochs=5)
```

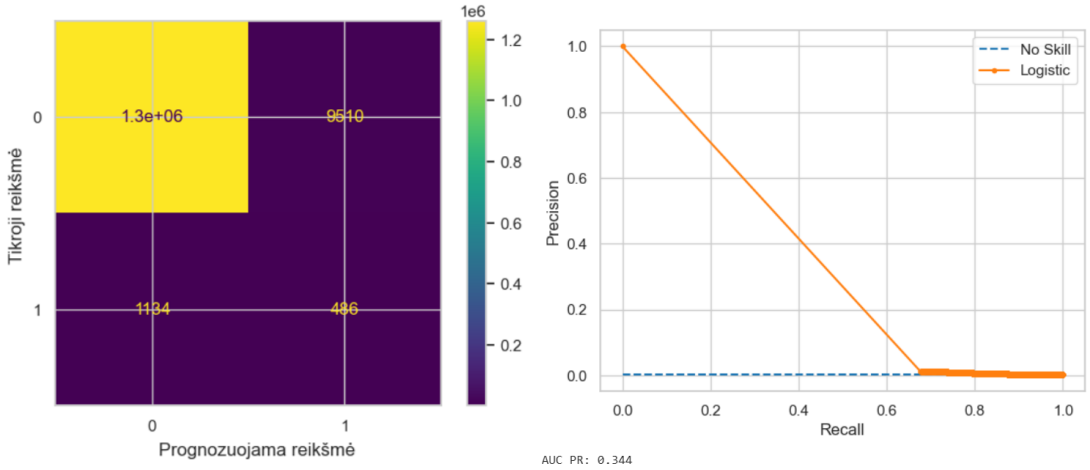
```
2022-09-11 12:54:11.394237: I tensorflow/compiler/mlir/mlir_graph_optimization_pass.cc:185] None of the MLIR Optimization
Passes are enabled (registered 2)

Epoch 1/5
149124/149124 [==============================] - 219s 1ms/step - loss: 0.0065 - accuracy: 0.9987
Epoch 2/5
149124/149124 [==============================] - 218s 1ms/step - loss: 0.0043 - accuracy: 0.9992
Epoch 3/5
149124/149124 [==============================] - 214s 1ms/step - loss: 0.0041 - accuracy: 0.9992
Epoch 4/5
149124/149124 [==============================] - 214s 1ms/step - loss: 0.0040 - accuracy: 0.9992
Epoch 5/5
149124/149124 [==============================] - 211s 1ms/step - loss: 0.0039 - accuracy: 0.9993
```

33 pav. Daugiasluoksnio neuroninio tinklo modelis

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Finansinio sukčiavimo tyrime labai svarbu subalansuoti duomenis tam, kad metodai teisingai klasifikuotų duomenis. Įvertiname neuroninių tinklų metodą su nesubalansuotais duomenimis.



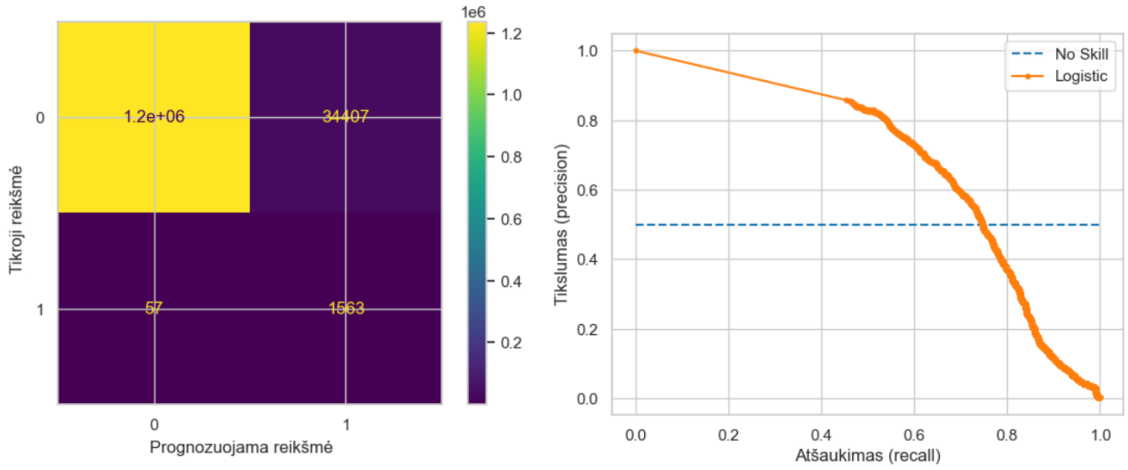
34 pav. Neuroninio tinklo klaidų matrica (kairėje) ir precision- recall kreivė (dešinėje) su nesubalansuotais duomenimis

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Su nesubalansuotais duomenimis matome, jog neuroninių tinklų modelis teisingai suklasifikavo, jog buvo finansinis sukčiavimas iš 1620 tik 486. Taip pat apie žemą klasifikavimo kokybę rodo ir AUC rodiklis 0,34.

Toliau šiome tyrimo eigoje bus naudojami duomenys subalansuoti pagal SMOTE metodą.

Neuroninių tinklų modelio vertinimą pradėsime nuo klaidų matricos aptarimo (žr. 35 pav.). Matome, kad teisingai modelis atspėjo, kad transakcija buvo neteisėta 1482 iš 1620, o 138 neidentifikavo, jog tai buvo sukčiavimo atvėjis. Iš klaidų matricos matome gana didelį kiekį neteisingai priskirtų legalių transakcijų prie sukčiavimo – net 34938. Praktikoje pritaikytas šis modelis sukeltų daug papildomo darbo, nes transakcija būtų sustabdyta ir turėtų būti patikrinta. Aukštas klaidingai pažymėtų legalių transakcijų kiekis gali rodyti, kad šio modelio slenkstis (angl. threshold) yra labai aukštas, tačiau mažinant slenkstį mažinamas jautrumas (angl. recall) ir taip galima praleisti finansinio sukčiavimo atvėjį.



35 pav. Neuroninio tinklo klaidų matrica (kairėje) ir preciziškumo - atkūrimo kreivė (dešinėje)

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Analizuojant modelio vertinimą (žr. lentelė 8) Neuroninio tinklo modelio bendras tikslumas 0,97, kuris yra ganėtinai aukštas, nes iš visų tikrų nelegalių transakcijų t. y. 1620, teisingai atspėjo, jog buvo vykdoma nelegali veikla net 1482 kartus. Modelio atšaukimo (jautrumas) vertinimas siekia 0,70, o tai rodo, kad modelis gali aptikti 70 proc. nesąžiningos veiklos. Tuo tarpu tikslumas (angl. precision) - 0,6. Mažesnis tikslumas reiškia, kad didesnis skaičius teisėtų operacijų yra neteisingai pažymėtas.

8 lentelė. Neuroninio tinklo modelio vertinimas

Modelio vertinimo rodiklis	Bendrasis tikslumas	Tikslumas (precision)	Atšaukimas (recall)	F1	AUC
Vertinimas	0,97	0,60	0,70	0,08	0,68

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

F1 balas yra labai mažas t. y. 0,08. Rodiklis atspindi apgaulės nustatymo kompromisą tarp tikros nesąžiningos veiklos fiksavimo ir klaidingų pavojaus signalų.

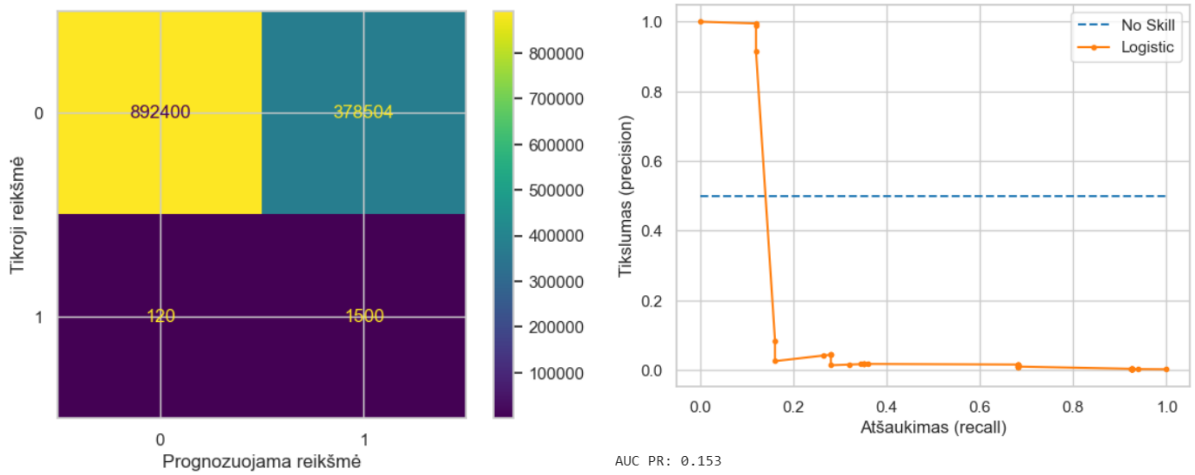
Apibendrinant, nors bendras modelio tikslumas yra didelis t. y. 0,97, tačiau AUC rodiklis, kuris parodo modelio našumą yra 0,68 dėl didelio kiekio legalių transakcijų interpretavimo kaip nelegalių. Modelyje svarbu ne tik laiku nustatyti sukčiavimą ir jį sustabdyti, bet ir turėti kuo mažesnį skaičių klaidingos netiesos, nes kiekvieną sustabdytą transakciją turi patikrinti žmogiškasis resursas, tuo pačiu gali kilti ir pasipiktinimas ir iš vartotojo dėl sustabdyto pavedimo.

Bajeso tinklo metodo vertinimas

Bajeso tinklai remiasi tikimybės teorija, leisdami modeliuoti įvykių tarpusavio priklausomybę ir naudoti tikimybes, kad būtų numatyti ateities įvykiai. Tai leidžia efektyviai įtraukti tikimybės sąvokas į modelius ir įvertinti neapibrėžtumą.

Phyton programoje buvo naudojamas *BernoulliNB* funkcijų paketas.

Analizuojant Bajeso metodo klaidų matricą (žr. 36 pav.) matome, kad teisingai modelis suklasifikavo, kad transakcija buvo neteisėta 1500, 120 neįidentifikavo, todėl modelio tikslumas siekia 0,7. Taip pat, iš klaidų matricos matome, jog labai aukštas neteisingai suklasifikuotų transakcijų, kurios klaidingai buvo priskirtos prie nelegalios veiklos.



36 pav. Bajeso tinklo klaidų matrica (kairėje) ir preciziškumo - atkūrimo kreivė (dešinėje)

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Bendras modelio vertinimas yra žemas, nes atšaukimo (jautrumas) vertinimas siekia 0,12, o tai rodo, kad modelis gali aptikti 12 proc. nesąžiningos veiklos. Tuo tarpu tikslumas (angl. precision) - 0,1. Mažesnis tikslumas reiškia, kad didesnis skaičius teisėtų operacijų yra neteisingai pažymėtas.

9 lentelė. Bajeso tinklo modelio vertinimas

Modelio vertinimo rodiklis	Tikslumas (precision)	Atšaukimas (recall)	F1	AUC	Tikslumas (accurasy)
Vertinimas	0,1	0,12	0,01	0,15	0,7

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Išanalizavus Bajeso tinklų modelio vertintus kriterijus galima teigti, kad modelis nėra tinkamas naudoti praktikoje, nes nors ir teisingai aptiko didžiąją dalį finansinio sukčiavimo atvejų, tačiau tuo pačiu labai daug klaidingai priskyrė legalių sandorių prie nelegalios veiklos. Toks netinkamas klasifikavimas įmonei kainuotų labai daug laiko tikrinant sustabdytas

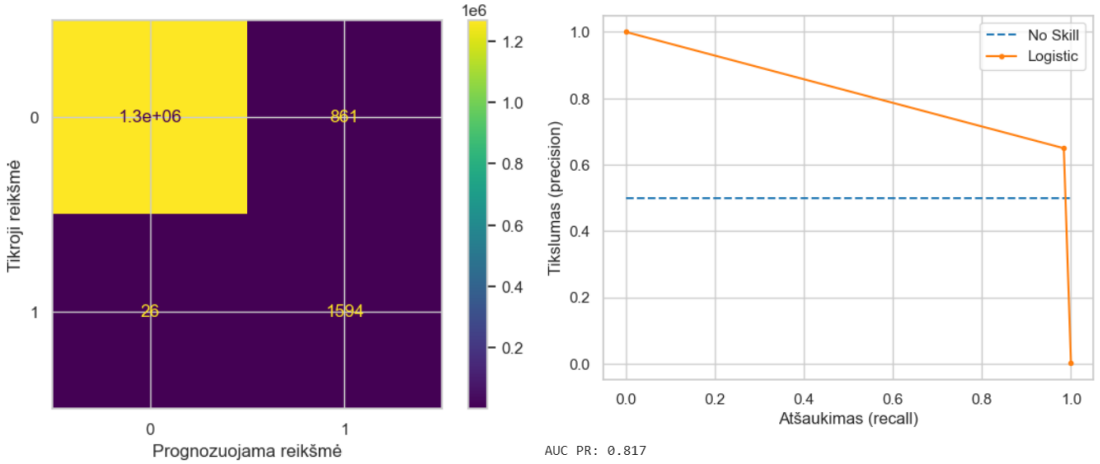
transakcijas. Taip pat tai kelia ir klientų pasipiktinimą, kai teisėtas pavedimas būtų sustabdytas dėl sistemos klaidos.

Sprendimų medžio metodo vertinimas

Sprendimo medžio (angl. *desition tree*) algoritmas buvo pasirinktas, nes jis veikia kaip dvejetainis klasifikatorius, todėl jis tinkamas sukčiavimo aptikimui, kai operacija priskiriama vienai iš dviejų (0 arba 1) sukčiavimo klasių.

Phyton programoje buvo naudojamas *DecisionTreeClassifier* funkcijų paketas.

Analizuojant sprendimo medžio klaidų matricą (žr. 37 pav.) matome, kad teisingai modelis suklasifikavo, kad transakcija buvo neteisėta 1594 ir tik 26 neįidentifikavo, todėl modelio tikslumas siekia 0,98. Taip pat, iš klaidų matricos matome, jog palyginus su prieš tai vertintais metodais klaidingai suklasifikuotų neteisingai legalių transakcijų priskirtų prie sukčiavimo yra tik 861.



37 pav. Sprendimų medžio klaidų matrica (kairėje) ir preciziškumo - atkūrimo kreivė (dešinėje)

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Aptariant modelio vertinimą (žr. lentelė 10) atšaukimo (jautrumas) vertinimas siekia 0,99, o tai rodo, kad modelis gali aptikti 99 proc. nesąžiningos veiklos. Taip pat F1 balas yra gana aukštas t. y. 0,82. Rodiklis atspindi apgaulės nustatymo kompromisą tarp tikros nesąžiningos veiklos fiksavimo ir klaidingų pavojaus signalų.

10 lentelė. Sprendimo medžio modelio vertinimas

Modelio vertinimo rodiklis	Tikslumas (presicion)	Atšaukimas (recall)	F1	AUC	Tikslumas (accurasy)
Vertinimas	0,65	0,98	0,82	0,82	0,99

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

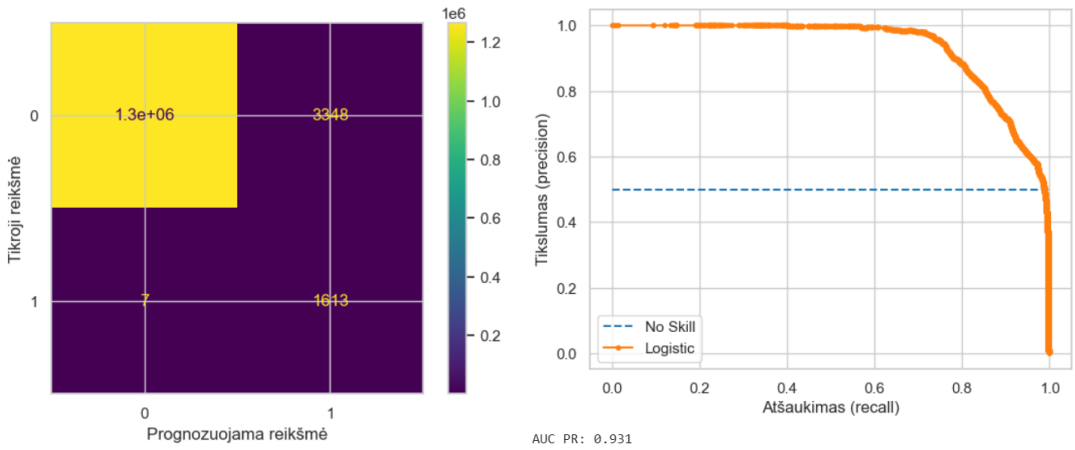
Išanalizavus sprendimo medžio modelio vertinimo kriterijus galima teigti, kad modelis būtų tinkamas įdiegti į finansinio sukčiavimo aptikimo įrankį ar sistemą, nes AUC rodiklis siekia 0,82, tai reiškia, kad modelis teisingai identifikuoja sukčiavimo atvejus net 82 proc. iš 100 proc. Taip pat klaidingai priskirtų prie nelegalios veiklos transakcijų skaičius yra priimtinas atsižvelgiant į visa transakcijų kiekį.

„CatBoost“ metodo vertinimas

„CatBoost“ algoritmo privalumai – gerai tvarkosi su kategoriniais kintamaisiais, yra greitesnis lyginant su kitais gradiento didinimo algoritmais, puikiai tinka didelėms duomenų imtims.

Phyton programoje buvo naudojamas *CatBoost Classifier* funkcijų paketas.

„CatBoost“ modelio klaidų matrica (žr. 38 pav.). rodo, kad teisingai finansinis sukčiavimas buvo aptiktas 1613 iš 1620 atvejų, t. y. beveik visi. Taip pat neteisingai legalių transakcijų priskirtų prie sukčiavimo – 3348 beveik 10 kartų mažiau nei Neuroninio tinklo metode.



38 pav. „CatBoost“ klaidų matrica (kairėje) ir preciziškumo - atkūrimo kreivė (dešinėje)

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Kalbant apie kitus modelio vertinimo rodiklius tikslumas (angl. precision) 0,6. Atšaukimo (angl. recall) vertinimas siekia 0,96, o tai rodo, kad modelis gali aptikti 96 proc. nesąžiningos veiklos. Tačiau F1 balas siekia tik 0,49 dėl klaidingų pavojaus signalų, kai sandoriai buvo teisėti.

11 lentelė. „CatBoost“ modelio vertinimas

Modelio vertinimo rodiklis	Tikslumas (precision)	Atšaukimas	F1	AUC	Tikslumas (accurasy)
Vertinimas	0,6	0,96	0,49	0,93	0,99

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

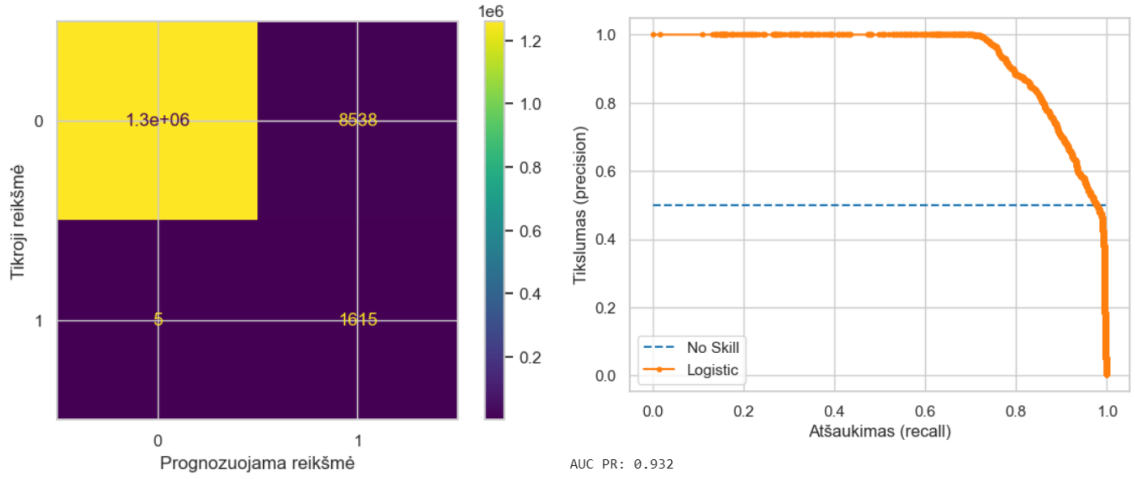
Apibendrinant, bendras modelio tikslumas yra aukštas t. y. 0,99. Taip pat ir AUC rodiklis, kuris parodo modelio našumą siekia 0,93. Tačiau, klaidingai priskirtų prie nelegalios veiklos transankcijų skaičius jau galėtų būti diskutuojamas ir būtent šioje vietoje reikėtų peržiūrėti modelio parametrus.

„XGBoost“ metodo vertinimas

„XGBoost“ yra dar vienas galingas mašininio mokymosi algoritmas, kuris yra plačiai naudojamas klasifikavimo uždaviniams, įskaitant finansinio sukčiavimo aptikimą.

Phyton programoje buvo naudojamas funkcijų *XGBClassifier* paketas.

„XGBoost“ modelio vertinimą pradėsime nuo klaidų matricos aptarimo (žr. 39 pav.). Klaidų matricoje matome, kad modelis teisingai suklasifikavo, kad tai yra neteisėta veikla 1615 iš 1620. Taip pat matome, kad neteisingai legalių transakcijų priskirtų prie sukčiavimo – net 8538. Palyginus su visu transakcijų skaičiumi klaidingai sustabdytų transakcijų kiekis yra vidutinis, tačiau tai visgi svarbus rodiklis, nes kiekviena sustabdyta transakcija yra patikrinama, o tai kainuoja papildomą žmogišką resursą.



39 pav. „„XGBoost““ klaidų matrica (kairėje) ir preciziškumo - atkūrimo kreivė (dešinėje)

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Analizuojant „XGBoost“ modelio vertinimą (žr. lentelė 17) bendras tikslumas 0,99, kuris yra ganėtinai aukštas, nes iš visų tikrų nelegalių transakcijų t. y. 1620 praleido tik 5 atvejus, kurie buvo neteisėti. Modelio atšaukimo (angl. recall) vertinimas siekia 0,95, o tai rodo, kad modelis gali aptikti 95 proc. nesąžiningos veiklos. Tuo tarpu tikslumas (angl. precision) - 0,6. F1 balas yra gana mažas t. y. 0,31. Rodiklis atspindi apgaulės nustatymo kompromisą tarp tikros nesąžiningos veiklos fiksavimo ir klaidingų pavojaus signalų.

12 lentelė. „XGBoost“ modelio vertinimas

Modelio vertinimo rodiklis	Tikslumas	Atšaukimas	F1	AUC	Tikslumas (accurasy)
Vertinimas	0,6	0,95	0,31	0,93	0,99

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Apibendrinant, nors bendras modelio tikslumas yra didelis t. y. 0,97, tačiau AUC rodiklis, kuris parodo modelio našumą yra 0,63 dėl didelio kiekio legalių transakcijų interpretavimo kaip nelegalių. Modelyje svarbu ne tik laiku nustatyti sukčiavimą ir jį sustabdyti, bet ir turėti kuo mažesnį skaičių klaidingos netiesos, nes kiekvieną sustabdytą transakciją turi patikrinti žmogiškasis resursas, tuo pačiu gali kilti ir pasipiktinimas ir iš vartotojo dėl sustabdyto pavedimo.

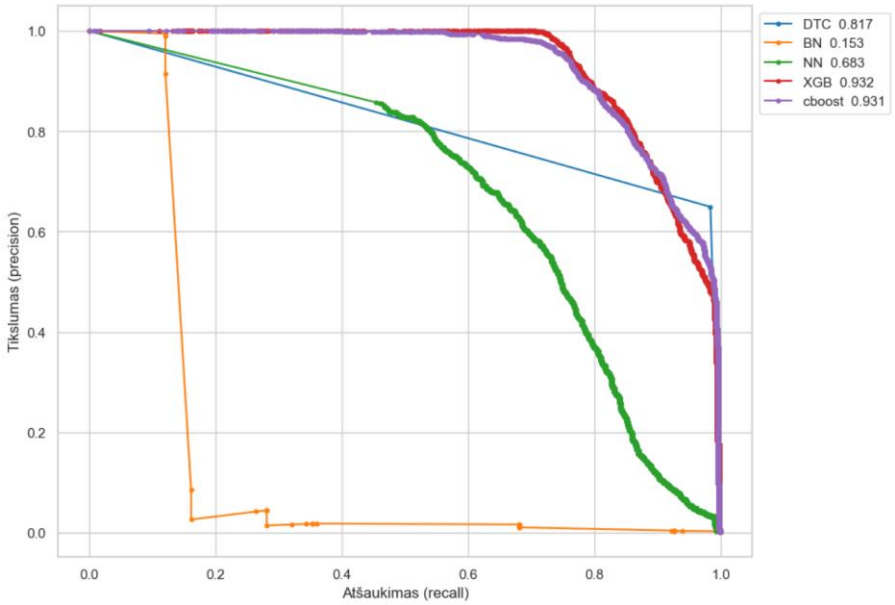
3.3. Skaitmeninio intelekto metodų palyginimas

Šioje darbo dalyje bus palyginami vertinti metodai.

Bendrą modelių palyginimo vertinimą pradėsime nuo AUC kreivės, kuri parodo modelio našumą. 40 paveiksle matome, jog „XGBoost“ ir „CatBoost“ metodų AUC kreivės yra 0,93. Šis rodiklis yra labai aukštas, todėl vertinant tik pagal šį rodiklį būtų galima teigti, kad būtent šie gradiento didinimo algoritmai turėtų būti pasirinkti diegiant finansinio sukčiavimo atpažinimo įrankyje.

Analizuojant kitus metodus, kaip sprendimų medį, jo AUC vertinimas yra 0,82. Šis vertinimas taip pat aukštas. Kiek mažesnė AUC rodiklis t. .y. 0,68 neuroninio tinklo. Žemiausias tikslumo ir atšaukimo santykio vertinimas yra Bajeso tiklo, kuris siekia tik 0,15.

Visgi AUC rodiklis yra tik vienas iš modelio vertinimo parametrų, todėl siekiant objektyviai įvertinti bendrą modelio veikimą ir jo pertaikymą finansinio sukčiavimo aptikimo sistemoje būtina atsižvelgti ir į kitus kriterijus, tokius kaip klaidų skaičius ir kita.



40 pav. Metodų: neuroniniai tinklai, bajeso tinklas, sprendimo medis, „XGBoost“, „CatBoost“ preciziškumo - atkūrimo kreivė

Šaltinis: sudarytas autorės remiantis Kaggle duomenų baze

Analizuojant žemiau pateiktą lentelę (žr. 14 lentelė) jeigu lygintume visus kriterijus geriausias metodas būtų sprendimo medis, nes iš 1620 finansinio sukčiavimo atvejų buvo teisingai suklasifikuoti atvejai 1594. Lyginant bendrai šis rezultatas tik 3 iš 5, nes „XGBoost“ teisingai atspėjo, kad tai neteisėta veikla 1613, o „CatBoost“ 1615. Sprendimų medžio metodą išskiria iš kitų lyginamų technikų, žemas klaidingai legalių transakcijų priskyrimas prie finansinio sukčiavimo. Šis kriterijus svarbus, nes klaidingai sustabdytos transakcijos kainuoja laiko ir pinigų įmonei, nes turi būti skiriamas žmogiškasis resursas kiekvienai transakcijai patikrinti. Taip pat gali kilti nepasitenkinimas iš kliento, kurio transakcija buvo sustabdyta.

13 lentelė. CI metodų modelių vertinimo kriterijų palyginimas

Metodas / Vertinimo kriterijai	Neuroniniai tinklai	Bajeso tinklai	Sprendimų medis	„XGBoost“	„CatBoost“
Teisingai teigiama(angl. true possitive)	1236497	892400	1270043	1267556	1262366
Klaidingai neigiama (angl. false negative)	34407	378504	861	3348	8538
Klaidingai teigiama (angl. false positive)	57	120	26	7	5
Teisingai neigiama (angl. true negative)	1563	1500	1594	1613	1615
Tikslumas (angl. accuracy)	0,97	0,7	0,99	0,99	0,99
Tkslumas (angl. precision)	0,6	0,1	0,65	0,6	0,6
Atšaukimas (angl. recall)	0,7	0,12	0,98	0,96	0,95

13 lentelės tęsinys

Metodas / Vertinimo kriterijai	Neuroniniai tinklai	Bajeso tinklai	Sprendimų medis	„XGBoost“	„CatBoost“
F1	0,08	0,01	0,82	0,49	0,31
AUC	0,68	0,15	0,82	0,93	0,93

Šaltinis: sudaryta autorės

„XGBoost“ ir „CatBoost“ metodų vertinimas yra labai panašus. Šių metodų vertinimai išsiskiria klaidingai priskirtų legalių transakcijų prie finansinio sukčiavimo atvejų. „CatBoost“ nors praleido tik 5 finansinio sukčiavimo atvejus, visgi labai daug klaidų sugeneravo neteisingai priskirdama legalias transakcijas prie nelegalių 8538, kai tuo tarpu „XGBoost“ praleido 7 finansinio sukčiavimo atvejus ir klaidingai priskyrė daugiau nei du kartus mažiau 3348. Abu šie metodai yra potencialūs, tačiau reikėtų skirti dėmesį taisyklių rinkiniui dėl klaidingai priskirti nelegaliai veiklai legalių transakcijų.

Neuroninių tinklų metodo patikimumo vertinimas būtų 4 iš 5. Ši technologija iš 1620 finansinio sukčiavimo atvejų teisingai suklasifikavo 1563 atvejus, praleido 57 nelegalios veiklos atvejus. Tačiau, šis metodas palyginus su prieš taip aptartais metodais padarė labai daug klaidų priskirdama legalias transakcijas prie nelegalių. Šio metodo sugeneruotos klaidos yra per didelės net ir vertinant tai, kad klasifikuotų transakcijų kiekis taip pat yra didelis t. .y. virš 1,6 mln. transakcijų.

Bajeso tinklo metodas prasčiausiai klasifikavo legalias ir nelegalias transakcijas. Ši technologija iš 1620 finansinio sukčiavimo atvejų teisingai suklasifikavo 1500 atvejus, praleido 120 nelegalios veiklos atvejus, tai reiškia, kad 92 proc. teisingai suklasifikavimo nelegalius atvejus. Tačiau labai daug klaidų buvo padaryta legalias transakcijas priskiriant prie nelegalių. Šis metodas pagal skaičiuotus parametrus sugeneruotų labai daug laiko ir išlaidų įmonei, nes kiekvieną sustabdytą operaciją turėtų patikrinti banko ar įmonės darbuotojas, todėl šio metodo išlaidų taupymas nebus vertinamas.

14 lentelė. Skaitmeninio intelekto metodų išlaidų taupymo apskaičiavimas

	Neuroninių tinklų	Sprendimų medis	„XGBoost“	„CatBoost“
TP	1236497	1270043	1267556	1262366
TP_B	35970	2455	4961	10153
TF	1563	1594	1613	1615
A	178197			
A_F	1467967			
CS_{total}	167 258 761 008,00	226 033 806 453,00	225 587 244 771,00	212 128 543 701,00

Šaltinis: sudarytas autorės remiantis išlaidų taupymo apskaičiavimu

Iš pateiktos 14 lentelės (žr. 14 lentelė) matome, kad sprendimų medžio metodas pritaikytas Paysim finansinio sukčiavimo aptikimo sistemoje įmonei galėtų padėti sutaupyti 226 milijardus. Afrikos valiutos (tiksli valiuta nebuvo įvardinta). Tačiau labai ne daug atsilieka ir „XGBoost“ metodas, kuris pritaikytas veikloje padėtų sutaupyti 225 milijardus. „CatBoost“ pritaikymas sutaupytų 212 milijardus, o neuroninių tinklų 167 milijardus.

Apibendrinant, galime teigti, kad metodais, kurie paremti sprendimo medžio technika davė geresnius rezultatus nei neuroninio tinklo veikimu parenti metodai. Išanalizavus metodo vertinimo kriterijus ir pritaikius išlaidų taupymo paskaičiavimą finansų srityje dirbanti įmonė savo veikloje pritaikiusi sprendimų medžio metodą finansinio sukčiavimo įrankyje galėtų sutaupyti 226 milijardus. Taip pat daug žadantys algoritmai yra „XGBoost“ ir „CatBoost“, nes jie geriausiai suklasifikavo finansinio sukčiavimo atvejus, tačiau šiems metodams reikia padirbėti ties klaidingų legalių transakcijų priskyrimu prie nelegalių.

Nors skaitmeninis intelektas turi didelį potencialą, tačiau svarbu identifikuoti ir iššūkius, su kuriais susiduriame taikant šiuos metodus praktiškai. Šio tyrimo darbo apribojimai yra duomenys, adaptavimas kintančioje aplinkoje, techniniai resursai. Kiekvieną iš apribojimų aptarsime plaučiau.

Taigi, finansiniai duomenys yra įvardijami kaip vienas dažniausiai pasitaikančių apribojimų tokio tipo tyrimuose vien dėl to įvairovės trūkumo. Šiame tyrime panaudoti sintetiniai duomenys sukurti pagal tikrų duomenų pavyzdį. Tyrime naudotų duomenų rinkinys yra ribotas. Ar transakcija yra finansinis sukčiavimas ar ne buvo galima spėti iš ne didelio skirtingų duomenų kiekio – pervedamos sumos dydžio, transakcijos tipo ir kita. Kuo duomenys įvairesni, tuo tiksliau galima sudaryti vartojo elgesio šabloną. Norint išnaudoti visą skaitmeninio intelekto metodų potencialą būtina įvairi duomenų bazę, nes šie metodai ir mokosi iš istorinių duomenų.

Kitas šio tyrimo ribojimas yra labai glaudžiai susijęs su jau aptartu ribojimu, t. y. adaptyvumas. Jeigu duomenys yra riboti tada net ir nedideli pokyčiai, ar naujos situacijos gali sukelti sunkumų prisitaikant prie nežinomų situacijų, kai istoriniuose duomenyse nėra pakankamai informacijos. Tokiu atveju metodai gali generuoti daug klaidingų reikšmių, kurias turi papildomai patikrinti žmogus, o visa tai įmonei generuoja papildomus kaštus.

Techninius resursus taip pat galime priskirti prie ribojančių veiksnių. Norint kokybiškai ir efektyviai dirbti su dideliais duomenų rinkiniais atitinkamai yra reikalingi ir didesni resursai, naujos technologijos. Šio tyrimo duomenų eilučių skaičius 1,6 mln. jau reikalavo specifinių programų tokių kaip Phyton, kad būtų galima apdoroti duomenis, nes tradiciniai įrankiai tokie kaip excel programa, SPSS yra nepritaikyti apdoroti tokio dydžio duomenis.

Šio darbo rezultatai įrodė, jog skaitmeninio intelekto metodų panaudojimas finansinio sukčiavimo atpažinimo tyrime turi didžiulį potencialą dėl tikslios klasifikacijos atpažįstant finansinį sukčiavimą. Šis tyrimas galėtų būti naudingas įmonėms, kurios kuria sistemas skirtas pinigų plovimo ir teroristų finansavimo prevencijos sustiprinimui ar finansinių operacijų stebėsenos įrankius. Tęsiant šį tyrimą būtų naudinga pagalvoti apie galimybę įtraukti į duomenų rinkinį įvairesnius duomenų šaltinius tokius kaip pažink savo vartotoją (angl. Know your customer), tai leistų jau ankstesniame žingsnyje geriau identifikuoti ar klientas yra rizikingas ar ne.

IŠVADOS

1. Atlikus mokslinės literatūros analizę apie finansinio sukčiavimo reiškinį galima teigti, kad didėjant finansinių paslaugų ir produktų įvairovei, finansų įmonėms tampa vis sudėtingiau atitikti keliamus reikalavimus dėl finansinio sukčiavimo kontrolės, nes šio reiškinio mastas - didžiulis. Išleidžiamos milžiniškos sumos kovai su šia nelegalia veikla, tačiau sėkmingų atvejų rodiklis, kai identifikuojamas finansinis sukčiavimas labai žemas. Taip pat svarbu paminėti, jog visuotinai sutariama, kad tradicinių statistikos metodų nebeužtenka ir dirbtinio intelekto pagrindu sukurti sprendimai yra būdas suvaldyti finansinę nusikalstamą veiklą.
2. Šiame darbe pasirinkta analizuoti dirbtinio intelekto pošakio skaitmeninio intelekto metodų pritaikymą finansinio sukčiavimo atpažinimo tyrime. Skaitmeninio intelekto įrankiai yra lankstesnis ir jų technika remiasi „minkštaisiais skaičiavimo metodais", kurie leidžia prisitaikyti prie daugelio situacijų. Taip pat šie metodai yra artimi žmogaus samprotavimo būdui, t. y. naudoja netikslią ir neišsamią informaciją, gali atlikti kontrolės veiksmus, ko priešingai dirbtinio intelekto metodai nesugeba.
3. Išanalizavus mokslinėje literatūroje aprašytus finansinio sukčiavimo bankų ir finansinių technologijų įmonių tyrimų atvejus matome, kad labai dažnai naudojami CI metodai yra neuroniniai tinklai, sprendimo medis ir kiti pagal jo veikimo logika sukurti algoritmai tokie kaip atsitiktiniai miškai, „XGBoost", „CatBoost". Dažnas šių metodų naudojimas yra grindžiamas tuo, kad jie pasižymi šiomis savybėmis: sugeba apdoroti didžiulius duomenų kiekius, padeda aptikti netipinę finansinę elgseną ir nustatyti įtartinas transakcijas ar veiksmus, kurie gali būti susiję su neteisėtomis finansinėmis veiklomis.
4. Atlikus metodų vertinimo tyrimą galime teigti, kad metodai, kurie veikia sprendimo medžio veikimo logika davė geresnius rezultatus nei neuroninio tinklo veikimu paremti metodai. Įvertinus pasirinktus CI metodus ir pritaikius išlaidų taupymo atskaičiavimą finansų srityje dirbanti įmonė savo veikloje pritaikiusi sprendimų medžio metodą finansinio sukčiavimo įrankyje galėtų sutaupyti 226 milijardus, pritaikius „XGBoost" algoritmo metodą 225 milijardus, „CatBoost" algoritmo metodą 212 milijardus. Taip pat svarbu pabrėžti, kad puikiai pripažįstami kaip efektyvus algoritmai yra „XGBoost" ir „CatBoost", nes jie tiksliausiai suklasifikavo finansinio sukčiavimo atvejus. Visgi šiems metodams reikia korekcijų dėl klaidingų legalių transakcijų priskyrimo prie nelegalių, Kiekviena sustabdyta transakcija įmonei kainuoja pinigus ir laiką bei gali neigiamai paveikti įmonės įvaizdį, nes klientas gali jausti nepasitenkinimą, kai jo pervedamų pinigų operacija yra sustabdoma ir jis laiku negali atsiskaityti už prekes ar paslaugas.

LITERATŪROS SĄRAŠAS

Analyttica Datalab's Medium, (2021). What is meant by ‘Stratified Split’? | by Analyttica Datalab | Medium

Analyttica Datalab's Medium, (2021). SMOTE(Synthetic Minority Over-sampling Technique) | by parth dholakiya | Medium

Awoyemi, J.O.; Adetunmbi, A.O.; Oluwadare, S.A. (2017).Credit card fraud detection using machine learning techniques: A comparative analysis. In Proceedings of the international conference on computing networking and informatics (ICCNI), Ota, Nigeria, 29–31 October 2017; pp. 1–9

Anti- money Laudering Market.(2022). https://www.marketsandmarkets.com/Market-Reports/anti-money-laundering-solutions-market-95490454.html?gclid=CjwKCAjwm8WZBhBUEiwA178UnEPu4IQINJ45Df5rS12j.GtpVAsafr7CCMOsyZ-ZYOWInX60-Bbj9mBoCUwkQAvD_BwE

Baesens, B. (2017). Network-based fraud detection for social security fraud. Management Science, 63(9), 3090–3110. https://doi.org/10.1287/mnsc.2016.2489

Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection.

Barua, Sujoy, Divya Gavandi, Pooja Sangle, Leena Shinde, and Jyoti Ramteke. (2021). “Swindle: Predicting the Probability of Loan Defaults Using „CatBoost“ Algorithm.” Pp. 1710–15 in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). Erode, India: IEEE

Badawi, A. (2021). Detection of money laundering in bitcoin transactions. https://ieeexplore.ieee.org/abstract/document/9770661

Berniukevičius, A., Kurilovas, E.,(2017). Dirbtiniai neuroniniai tinklai personalizuotam mokymuisi. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiqya_sm7D6AhVnkIsKHaI2AXoQFnoECC4QAQ&url=https%3A%2F%2Fwww.zurnalai.vu.lt%2FLMR%2Farticle%2Fdownload%2F17755%2F16919%2F&usg=AOvVaw3mk3w2ve7B0CEmK6f2urge

Bezdek, J. S., (1994) What is Computational Intelligence? https://www.researchgate.net/publication/220045330_What_is_Computational_Intelligence

Brabazon, A., O’neill, M., & Dempsey, I. (2008). An Introduction to Evolutionary Computation in Finance. https://doi.org/10.1109/MCI.2008.929841

Buzius, G. (2010). Bajeso metodo taikymas kredito rizikos valdyme. https://epublications.vu.lt/object/elaba:2080843/2080843.pdf

Chen, Wang, Kuo, (2017). Computational Intelligence in Economics and Finance. https://www.researchgate.net/publication/279133764_Computational_Intelligence_in_Economics_and_Finance

Chen, T., Guestrin, C. (2016).“„XGBoost“: A scalable tree boosting system”. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data mining. 2016, pp. 785–794.

Chen, Y., Han, X. (2021). „CatBoost“ for Fraud Detection in Financial Transactions. https://ieeexplore.ieee.org/document/9342475

Ch, R., Gadekallu, T. R., Abidi, M. H., 3, (2022). Computational System to Classify Cyber Crime Offenses using Machine Learning. https://www.mdpi.com/2071-1050/12/10/4087

Chandola, V., Banerjee, A., Kumar, V., (2009) Anomaly Detection: A Survey. https://dl.acm.org/doi/10.1145/1541880.1541882

Chen, Y., Wei, Z., Gou, H., Liu, H., Gao, Li., He, X., (2022). How far is brain-inspired artificial intelligence away from brain? https://www.frontiersin.org/articles/10.3389/fnins.2022.1096737/full
Deloitte. (2022). The new physics of financial services. How artificial intelligence is transforming the financial ecosystem. Prieiga per internetą: https://www2.deloitte.com/bd/en/pages/financial-services/articles/artificial-intelligence-transforming-financial-ecosystem-deloitte-fsi.html
Desrousseaux,R., Bernard, G., Mariage, J. (2021). Profiling Money Laundering with Neural Networks: a Case Study on Environmental Crime Detection. https://ieeexplore.ieee.org/document/9643405
D'Souza, S. (2018). Let's learn about AUC ROC Curve! Prieiga per internetą: https://medium.com/greyatom/lets-learn-about-auc-roc-curve-4a94b4d88152
Duhart, B., Hernández-Gress, N., (2016). Review of the Principal Indicators and Data Science Techniques used for the Detection of Financial Fraud and Money Laundering. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7881560
Dunjko, V., Briegel H. (2018). Machine learning & artificial intelligence in the quantum domain: a review of recent progress. https://iopscience.iop.org/article/10.1088/1361-6633/aab406/meta
Finteth Gloabal. (2023). Ancient scams to modern threats: How financial fraud costs us $3.7tn annually. https://fintech.global/2023/09/13/ancient-scams-to-modern-threats-how-financial-fraud-costs-us-3-7tn-annually/
Ghobadi, F. Rohani, M. (2016) Cost sensitive modeling of credit card fraud using neural network strategy. In Proceedings of the 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), Tehran, Iran, 14–15 December 2016; pp. 1–5
Gyamfi, N.K.; Abdulai, J. Bank Fraud Detection Using Support Vector Machine. In Proceedings of the 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 1–3 November 2018; pp. 37–41.
Gupka, S., (2021). Data Mining-based Financial Statement Fraud Detection: Systematic Literature Review and Meta-analysis to Estimate Data Sample Mapping of Fraudulent Companies Against Non-fraudulent Companies https://journals.sagepub.com/doi/10.1177/0972150920984857
Hajek, P., Abedin, M. Z., Sivarajah, U. (2022). Fraud Detection in Mobile Payment Systems using an „XGBoost"-based Framework. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9560719/
HaratiNik, M.R.; Akrami, M.; Khadivi, S.; Shajari, M. FUZZGY: A hybrid model for credit card fraud detection. In Proceedings of the 6th International Symposium on Telecommunications (IST), Tehran, Iran, 6–8 November 2012; pp. 1088–1093
Hassan, M. A. J., Kadir, M. M., Hussein, D. (2020) Improving Classification Performance for a Novel Imbalanced Medical Dataset Using SMOTE Method. http://eprints.cihanuniversity.edu.iq/1302/
Heckerman, A. (1996) tutorial on learning with Bayesian networks, Microsoft Research tech. report, MSR-TR-95-06, ftp://ftp.research.microsoft.com/pub/tr/TR-95-06.PS
Heidarinia, N., Harounabadi A., Sadeghzadeh M. (2014) An Intelligent Anti-Money Laundering Method for Detecting Risky Users in the Banking Systems https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.662.6190&rep=rep1&type=pdf
Hernandez-Julio, Y.F., et al. (2017). Framework for the development of business intelligence using computational intelligence and serviceoriented architecture. Proceedings of the 12th Iberian Conference on Information Systems and Technologies (CISTI), 1–7.

Hurt, A. (2023) AI and the Human Brain: How Similar Are They? https://www.discovermagazine.com/technology/ai-and-the-human-brain-how-similar-are-they

Jain, A., Shinde, S., (2019) A Comprehensive Study of Data Mining-based Financial Fraud Detection Research. https://ieeexplore.ieee.org/document/9033767
Johnston, M. (2018). Seizing the moment for artificial intelligence. Prieiga per internetą: https://blog.evergreengavekal.com/seizing-the-moment-for-artificial-intelligence/
Kacprzyk, J. (2017) Studies in Computational Intelligence. Polish Academy of Sciences, Warsaw
Kadar, T. (2023) Global Banking Fraud Index 2023. https://seon.io/resources/global-banking-fraud-index/
Kaggle. Synthetic Financial Datasets For Fraud Detection. Synthetic Financial Datasets For Fraud Detection (kaggle.com)
Kanade, V. (2021). What Is Fraud Detection? Definition, Types, Applications, and Best Practices | Toolbox It-security. https://www.toolbox.com/it-security/vulnerability-management/articles/what-is-fraud-detection
Kaspersky. (2020) Share of account takeover incidents increased by 20 percentage points compared to 2019. https://www.kaspersky.com/about/press-releases/2021_share-of-account-takeover-incidents-increased-by-20-percentage-points
Kevin, L. J Chen, J. (2011). Prediction and assessment of student learning outcomes in calculus a decision support of integrating data mining and Bayesian belief networks https://ieeexplore.ieee.org/abstract/document/5764024
Kho, J.R.D.; Vea, L.A. (2017) Credit card fraud detection based on transaction behavior. In Proceedings of the TENCON 2017-2017 IEEE Region 10 Conference, Penang, Malaysia, 5–8 November 2017; pp. 1880–1884.
Khillar, 2023. Difference Between AI and CI. Difference Between AI and CI | Difference Between
Kobadilov, B., Omarov, G., & Yermekbayeva, D. (2020). Financial technologies trends and how they will shape financial markets [Article]. The Economy: Strategy and Practice, 15(2), 151–157. https://doi.org/10.51176/JESP/issue_2_T13
Konar, A. (2006). Computational intelligence: principles, techniques and applications. Springer Science & Business Media,
Kroll. (2023) Fraud and Financial Crime Report Can technology stop the threat of economic, crypto and ESG crimes? https://www.kroll.com/-/media/kroll-images/pdfs/2023-fraud-and-financial-crime-report.pdf
Ling, C. X., Huang, J., Zhang, H. (2006) AUC: a Statistically Consistent and more Discriminating Measure than Accuracy. https://www.csd.uwo.ca/~xling/papers/ijcai03.pdf
Narkhede, S. (2018). Understanding Confusion Matrix. Prieiga per internetą: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62
Nguyen, N., Duong, T., Chau, T, Nguyen, V., Trinh, T.,Tran, D. (2022) A Proposed Model for Card Fraud Detection Based on „CatBoost“ and Deep Neural Network. https://ieeexplore.ieee.org/abstract/document/9882043
Ngai, E.W.T.; Hu, Y.; Wong, Y.H.; Chen, Y.; Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decis. Support Syst. 2011, 50, 559–569
Majhi, S. K. (2019). Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection [Article]. Evolutionary Intelligence, 14(1), 35–46. https://doi.org/10.1007/s12065-019-00260-3
Mareeswari, V.; Gunasekaran, G. Prevention of credit card fraud detection based on HSVM. In Proceedings of the 2016 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 25–26 February 2016; pp. 1–4.
Mantegna, R. N., & Giudici, P. (2018). Article 1 Citation: Giudici P (2018) Fintech Risk Management: A Research Challenge for. Frontiers in Artificial Intelligence | Www.Frontiersin.Org, 1, 1. https://doi.org/10.3389/frai.2018.00001

Malini, N., Pushpa, M. (2017) Analysis on credit card fraud identification techniques based on KNN and outlier detection. https://www.semanticscholar.org/paper/Analysis-on-credit-card-fraud-identification-based-Malini-Pushpa/697b5eadbf49039151ebea954fd220faccf04d22

Math Word. (2023). What Is a Neural Network?. https://ch.mathworks.com/discovery/neural-network.html?gclid=Cj0KCQiA3uGqBhDdARIsAFeJ5r2im4Mxjlw2BbzK2IG8PxhFZsGgHg9RMt2BDG1epVPljhgKa9r_YYMaAo4fEALw_wcB&ef_id=Cj0KCQiA3uGqBhDdARIsAFeJ5r2im4Mxjlw2BbzK2IG8PxhFZsGgHg9RMt2BDG1epVPljhgKa9r_YYMaAo4fEALw_wcB:G:s&s_kwcid=AL!8664!3!604196280133!p!!g!!neural%20network&s_eid=psn_40629618653&q=neural+network&gad_source=1

Özçelik, M.H.; Duman, E.; I ̧sik, M.; Çevik, T. Improving a credit card fraud detection system using genetic algorithm. In Proceedings of the 2010 International Conference on Networking and Information Technology, Manila, Philippines, 11–12 June 2010; pp. 436–440

Pol, R. F., (2020), Anti-money laundering: The world's least effective policy experiment? Together, we can fix it, https://www.tandfonline.com/doi/pdf/10.1080/25741292.2020.1725366?needAccess=true

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. (2018) "„CatBoost“: Unbiased boosting with categorical features", *Proc. 32nd Conf. Workshop Neural Inf. Process. Syst.*, pp. 1-11, Dec. 2018, [online]
Available: https://proceedings. neurips.cc/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf.

Rada R., (2008). Expert systems and evolutionary computing for financial investing: A review https://www.sciencedirect.com/science/article/pii/S0957417407001911

Rajawat A. S., Jain S., (2020). Fusion Deep Learning Based on Back Propagation Neural Network for Personalization. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9170693

Raj J. S. (2019). A comprehensive survey on the computational intelligence techniques and its applications. https://irojournals.com/iroismac/V1/I3/02.pdf

Randhawa, K.; Loo, C.K.; Seera, M.; Lim, C.P.; Nandi, A.K. Credit Card Fraud Detection Using AdaBoost and Majority Voting. IEEE Access 2018, 6, 14277–14284

Rocha-Salazar, J., Segovia-Vargas, M., Camacho-Miñano, M. (2021). Money laundering and terrorism financing detection using neural networks and an abnormality indicator. https://www.sciencedirect.com/science/article/pii/S0957417420311209

Sabau, A., (2012) Survey of Clustering based Financial Fraud Detection Research https://www.researchgate.net/publication/267686837_Survey_of_Clustering_based_Financial_Fraud_Detection_Research

Sadiku, Matthew N. O., Foreman, P.E. Justin, Musa Sarhan M., (2018) Computational Intelligence. Prairie View A&M University/Prairie View, TX, United States of America Doi: 10.19044/esj.2018.v14n21p56 URL:http://dx.doi.org/10.19044/esj.2018.v14n21p56

Salehi, A., Ghazanfari, M., Fathian, M. (2017). Data Mining Techniques for Anti Money Laundering http://www.ripublication.com/ijaer17/ijaerv12n20_120.pdf

Shikha, A., Agrawal, J. (2018). Survey on Anomaly Detection using Data Mining Techniques, https://www.sciencedirect.com/science/article/pii/S1877050915023479

Song, R., Huang L., Cui, W., Óskarsdóttir, M., Vanthienen, J. (2020) Fraud Detection of Bulk Cargo Theft in Port Using Bayesian Network Models. https://www.mdpi.com/2076-3417/10/3/1056

Srivastava, A.; Yadav, M.; Basu, S.; Salunkhe, S.; Shabad, M. Credit card fraud detection at merchant side using neural networks. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 667–670.

Siddique, A. (2013). Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing. John Wiley & Sons. https://download.e-bookshelf.de/download/0003/7273/32/L-G-0003727332-0002367307.pdf
Teichmann, F. (2019). Recent trends in money laundering. Crime, Law and Social Change 73, 237–247. https://doi.org/10.1007/s10611-019-09859-0
Tichonov, J. (2018). Klasifikavimo metodais grindžiami skaitmeninių vaizdų glaudinimo sprendimai: Daktaro disertacija. Vilnius: Vilniaus universitetas
Triepels, R., Daniels, H., Feelders, A. (2018) Data-driven fraud detection in international shipping. https://www.sciencedirect.com/science/article/abs/pii/S0957417418300083
The State of Fraud And Financial Crime In the U.S. (2022). https://www.pymnts.com/study/state-of-fraud-and-financial-crime-in-the-united-states-aml/
UK Finace. (2022). ANNUAL FRAUD REPORT The definitive overview of payment industry fraud in 2022. https://www.ukfinance.org.uk/system/files/2023-05/Annual%20Fraud%20Report%202023_0.pdf
Warwick, K. (2012) Artificial Intelligence The Basic. Artificial Intelligence: The Basics - Kevin Warwick - Google knygos
What Is a Neural Network?. https://se.mathworks.com/discovery/neural-network.html?gclid=Cj0KCQjwpompBhDZARIsAFD_Fp-jBegVRcrpQ00s936FrTS90777YABx7FQXV2HOaO9MOoKlNUSqrHEaAgz4EALw_wcB&ef_id=Cj0KCQjwpompBhDZARIsAFD_Fp-jBegVRcrpQ00s936FrTS90777YABx7FQXV2HOaO9MOoKlNUSqrHEaAgz4EALw_wcB:G:s&s_kwcid=AL!8664!3!604196280133!p!!g!!neural%20network&s_eid=psn_40629618653&q=neural%20network
West, J., Bhattacharya, B. (2015). Intelligent financial fraud detection: A comprehensive review. https://www.sciencedirect.com/science/article/pii/S0167404815001261
XGBboost developer. „XGBoost“ Documentation. 2020. url: https://“XGBoost“.readthedocs.io/en/latest/
Wiley, J. (2006). Naive Bayes Estimation And Bayesian Networks, Data Mining Methods and Models By Daniel T. Larose.
Xu, J., Lu,Y., Chau. (2015) P2P Lending Fraud Detection: A Big Data Approach. https://link.springer.com/chapter/10.1007/978-3-319-18455-5_5

1 PRIEDAS. Neuroninių tinkle metodo vertinimo skaičiavimo kodas Phyton programoje

Neural Networks

```
nn_model = MLPClassifier(solver='lbfgs', alpha=1e-5,
                         hidden_layer_sizes=(5, 2), random_state=1)
nn_model.fit(Xsc_train,y_train_sm)
nn_prediction = nn_model.predict(Xsc_test)
nn_prob = nn_model.predict_proba(Xsc_test)
```

```
list_metrics_nn = metrics_estimation(nn_model, Xsc_train, Xsc_test, y_train_sm, y_test, nn_prediction, nn_prob)
```

```
Precison score:  0.6001064962726305
Recall score:    0.695679012345679
Threshold:       0.9999150133621805
F1 score: 0.083160
Log Loss score: 0.091233
Brier score: 0.022985
```

```
# retrieve the probabilities for the positive class
nn_prob_positive = nn_prob[:, 1]

# calculate the no skill line as the proportion of the positive class
no_skill = len((y_train_sm[y_train_sm==1])+(y_test[y_test==1])) / len(y_train_sm + y_test)
# plot the no skill precision-recall curve
plt.plot([0, 1], [no_skill, no_skill], linestyle='--', label='No Skill')

# calculate inputs for the PR curve
precision, recall, thresholds = precision_recall_curve(y_test, nn_prob_positive)

# plot PR curve
plt.plot(recall, precision, marker='.', label='Logistic')
# axis labels
plt.xlabel('Atšaukimas (recall)')
plt.ylabel('Tikslumas (precision)')
# show the legend
plt.legend()
plt.show()

# calculate and print PR AUC
auc_pr = auc(recall, precision)
print('AUC PR: %.3f' % auc_pr)
```

```
cm_nn = confusion_matrix(y_test, nn_prediction, labels=nn_model.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm_nn, display_labels=nn_model.classes_)
disp.plot()
plt.xlabel ('Prognozuojama reikšmė')
plt.ylabel ('Tikroji reikšmė')
plt.show()
print(classification_report(nn_prediction, y_test))
```

```
              precision    recall  f1-score   support

           0       0.97      1.00      0.99   1236554
           1       0.96      0.04      0.08     35970

    accuracy                           0.97   1272524
   macro avg       0.97      0.52      0.53   1272524
weighted avg       0.97      0.97      0.96   1272524
```

2 PRIEDAS. Bajeso tinklų metodo vertinimo skaičiavimo kodas Phyton programoje

```
# Bajeso tinklas
from sklearn.naive_bayes import BernoulliNB, MultinomialNB
```

```
BN = BernoulliNB()
BN.fit(Xsc_train, y_train_sm)
BN_prediction = BN.predict(Xsc_test)
BN_prob = BN.predict_proba(Xsc_test)
```

```
list_metrics_BN = metrics_estimation(BN, Xsc_train, Xsc_test, y_train_sm, y_test, BN_prediction, BN_prob)
```

```
Precison score:  0.9948979591836735
Recall score:    0.12037037037037036
Threshold:       0.9982118177925006
F1 score: 0.007861
Log Loss score: 0.499835
Brier score: 0.158836
```

```
# retrieve the probabilities for the positive class
BN_prob_positive = BN_prob[:, 1]

# calculate the no skill line as the proportion of the positive class
no_skill = len((y_train_sm[y_train_sm==1])+(y_test[y_test==1])) / len(y_train_sm + y_test)
# plot the no skill precision-recall curve
plt.plot([0, 1], [no_skill, no_skill], linestyle='--', label='No Skill')

# calculate inputs for the PR curve
precision, recall, thresholds = precision_recall_curve(y_test, BN_prob_positive)

# plot PR curve
plt.plot(recall, precision, marker='.', label='Logistic')
# axis labels
plt.xlabel('Atšaukimas (recall)')
plt.ylabel('Tikslumas (precision)')
# show the legend
plt.legend()
plt.show()

# calculate and print PR AUC
auc_pr = auc(recall, precision)
print('AUC PR: %.3f' % auc_pr)
```

```
cm_BN = confusion_matrix(y_test, BN_prediction, labels=BN.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm_BN, display_labels=dtc_model.classes_)
disp.plot()
plt.xlabel ('Prognozuojama reikšmė')
plt.ylabel ('Tikroji reikšmė')
plt.show()
```

```
print(classification_report(BN_prediction, y_test))
```

```
              precision    recall  f1-score   support

           0       0.70      1.00      0.82    892520
           1       0.93      0.00      0.01    380004

    accuracy                           0.70   1272524
   macro avg       0.81      0.50      0.42   1272524
weighted avg       0.77      0.70      0.58   1272524
```

3 PRIEDAS. Sprendimo medžio metodo vertinimo skaičiavimo kodas Phyton programoje

Decition tree be smote

```
dtc_model=DecisionTreeClassifier(random_state=42)
dtc_model.fit(X_train,y_train)
dtc_prediction=dtc_model.predict(X_test)
dtc_prob = dtc_model.predict_proba(X_test)
```

```
list_metrics_dtc = metrics_estimation(dtc_model, X_train, X_test, y_train, y_test, dtc_prediction, dtc_prob)
```

```
Precison score:  0.9094881398252185
Recall score:    0.8993827160493827
Threshold:       1.0
F1 score: 0.904407
Log Loss score: 0.008360
Brier score: 0.000242
```

```
# retrieve the probabilities for the positive class
dtc_prob_positive = dtc_prob[:, 1]

# calculate the no skill line as the proportion of the positive class
no_skill = len((y_train[y_train==1])+(y_test[y_test==1])) / len(y_train + y_test)
# plot the no skill precision-recall curve
plt.plot([0, 1], [no_skill, no_skill], linestyle='--', label='No Skill')

# calculate inputs for the PR curve
precision, recall, thresholds = precision_recall_curve(y_test, dtc_prob_positive)

# plot PR curve
plt.plot(recall, precision, marker='.', label='Logistic')
# axis labels
plt.xlabel('Recall')
plt.ylabel('Precision')
# show the legend
plt.legend()
plt.xlabel('Atšaukimas (recall)')
plt.ylabel('Tikslumas (precision)')
plt.show()

# calculate and print PR AUC
auc_pr = auc(recall, precision)
print('AUC PR: %.3f' % auc_pr)
```

```
cm_dtc = confusion_matrix(y_test, dtc_prediction, labels=dtc_model.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm_dtc, display_labels=dtc_model.classes_)
disp.plot()
plt.xlabel('Tikroji reikšmė')
plt.ylabel('Prognozuojama reikšmė')
plt.show()
```

4 PRIEDAS. „XGBoost“ metodo vertinimo skaičiavimo kodas Phyton programoje

XGBoost

```
xgb_model = XGBClassifier(learning_rate=0.1,random_state=42)
xgb_model.fit(Xsc_train, y_train_sm)
xgb_prediction = xgb_model.predict(Xsc_test)
xgb_prob = xgb_model.predict_proba(Xsc_test)
```

```
list_metrics_xgb = metrics_estimation(xgb_model, Xsc_train, Xsc_test, y_train_sm, y_test, xgb_prediction, xgb_prob)
```

```
Precison score:  0.6006339144215531
Recall score:    0.9358024691358025
Threshold:       0.9610462188720703
F1 score: 0.274357
Log Loss score: 0.018038
Brier score: 0.005130
```

```
# retrieve the probabilities for the positive class
xgb_prob_positive = xgb_prob[:, 1]

# calculate the no skill line as the proportion of the positive class
no_skill = len((y_train_sm[y_train_sm==1])+(y_test[y_test==1])) / len(y_train_sm + y_test)
# plot the no skill precision-recall curve
plt.plot([0, 1], [no_skill, no_skill], linestyle='--', label='No Skill')

# calculate inputs for the PR curve
precision, recall, thresholds = precision_recall_curve(y_test, xgb_prob_positive)

# plot PR curve
plt.plot(recall, precision, marker='.', label='Logistic')
# axis labels
plt.xlabel('Atšaukimas (recall)')
plt.ylabel('Tikslumas (precision)')
# show the legend
plt.legend()
plt.show()

# calculate and print PR AUC
auc_pr = auc(recall, precision)
print('AUC PR: %.3f' % auc_pr)
```

```
#confusion matrix
cm = confusion_matrix(y_test, xgb_prediction, labels=xgb_model.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=xgb_model.classes_)
disp.plot()
plt.xlabel ('Prognozuojama reikšmė')
plt.ylabel ('Tikroji reikšmė')
plt.show()
```

```
print(classification_report(xgb_prediction, y_test))
```

```
              precision    recall  f1-score   support

           0       0.99      1.00      1.00   1262371
           1       1.00      0.16      0.27     10153

    accuracy                           0.99   1272524
   macro avg       1.00      0.58      0.64   1272524
weighted avg       0.99      0.99      0.99   1272524
```

```
#The XGBoost library provides a built-in function to plot features ordered by their importance.
from xgboost import plot_importance
# plot feature importance
plot_importance(xgb_model)
```

5 PRIEDAS. „CatBoost“ metodo vertinimo skaičiavimo kodas Phyton programoje

```
# Catboost
cbc_model = CatBoostClassifier(iterations=20,
                               loss_function='Logloss',
                               verbose=True)
cbc_model.fit(X_train, y_train)
cbc_prediction = cbc_model.predict(X_test)

cm_cbc = confusion_matrix(y_test, cbc_prediction, labels=cbc_model.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm_cbc, display_labels=cbc_model.classes_)
disp.plot()
plt.xlabel ('Prognozuojama reikšmė')
plt.ylabel ('Tikroji reikšmė')
plt.show()
```

```
Learning rate set to 0.5
0:      learn: 0.3145070        total: 604ms    remaining: 11.5s
1:      learn: 0.1723959        total: 1.14s    remaining: 10.2s
2:      learn: 0.0993267        total: 1.68s    remaining: 9.54s
3:      learn: 0.0591998        total: 2.22s    remaining: 8.89s
4:      learn: 0.0363624        total: 2.76s    remaining: 8.28s
5:      learn: 0.0225653        total: 3.31s    remaining: 7.72s
6:      learn: 0.0144196        total: 3.88s    remaining: 7.21s
7:      learn: 0.0095353        total: 4.51s    remaining: 6.77s
8:      learn: 0.0065552        total: 5.16s    remaining: 6.3s
9:      learn: 0.0048189        total: 5.83s    remaining: 5.83s
10:     learn: 0.0038126        total: 6.43s    remaining: 5.26s
11:     learn: 0.0030878        total: 7.06s    remaining: 4.71s
12:     learn: 0.0026459        total: 7.69s    remaining: 4.14s
13:     learn: 0.0023559        total: 8.33s    remaining: 3.57s
14:     learn: 0.0021842        total: 8.92s    remaining: 2.97s
15:     learn: 0.0020837        total: 9.52s    remaining: 2.38s
16:     learn: 0.0019283        total: 10s      remaining: 1.77s
17:     learn: 0.0017785        total: 10.6s    remaining: 1.18s
18:     learn: 0.0017418        total: 11.1s    remaining: 587ms
19:     learn: 0.0017083        total: 11.7s    remaining: 0us
```

```
cboost_model = CatBoostClassifier(verbose=0)
cboost_model.fit(Xsc_train, y_train_sm)
cboost_prediction = cboost_model.predict(Xsc_test)
cboost_prob = cboost_model.predict_proba(Xsc_test)
```

```
list_metrics_cboost = metrics_estimation(cboost_model, Xsc_train, Xsc_test, y_train_sm, y_test, cboost_prediction, cboost_prob)
```

```
Precison score:  0.6007751937984496
Recall score:    0.9567901234567902
Threshold:       0.9851168054567936
F1 score: 0.490199
Log Loss score: 0.009892
Brier score: 0.002280
```

```
# retrieve the probabilities for the positive class
cboost_prob_positive = cboost_prob[:, 1]

# calculate the no skill line as the proportion of the positive class
no_skill = len((y_train_sm[y_train_sm==1])+(y_test[y_test==1])) / len(y_train_sm + y_test)
# plot the no skill precision-recall curve
plt.plot([0, 1], [no_skill, no_skill], linestyle='--', label='No Skill')

# calculate inputs for the PR curve
precision, recall, thresholds = precision_recall_curve(y_test, cboost_prob_positive)

# plot PR curve
plt.plot(recall, precision, marker='.', label='Logistic')
# axis labels
plt.xlabel('Atšaukimas (recall)')
plt.ylabel('Tikslumas (precision)')
# show the legend
plt.legend()
plt.show()

# calculate and print PR AUC
auc_pr = auc(recall, precision)
print('AUC PR: %.3f' % auc_pr)
```

```
#confusion matrix
cm = confusion_matrix(y_test, cboost_prediction, labels=cboost_model.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=cboost_model.classes_)
disp.plot()
plt.xlabel ('Prognozuojama reikšmė')
plt.ylabel ('Tikroji reikšmė')
plt.show()
```

6 PRIEDAS. Straipsnis – Pinigų plovimo atpažinimo tyrimas naudojant skaitmeninį intelektą

VILNIUS TECH
Vilnius Gediminas Technical University

26-osios Lietuvos jaunųjų mokslininkų konferencijos „Mokslas – Lietuvos ateitis" teminė konferencija
Proceedings of the 26th Conference for Junior Researchers "Science – Future of Lithuania"

EKONOMIKA IR VADYBA / ECONOMICS AND MANAGEMENT

eISSN 2029-7149
2023 m. spalio 17 d., Vilnius, Lietuva
17 October 2023, Vilnius, Lithuania
eISBN 978-609-476-343-4
Article Number: vvf.2023.1417

Finansų valdymas / Financial Management
https://vilniustech.lt/jmk-vvf

PINIGŲ PLOVIMO ATPAŽINIMO TYRIMAS NAUDOJANT SKAITMENINĮ INTELEKTĄ

Akvilė EREMINĖ*, Nijolė MAKNICKIENĖ

[1] *Vilniaus Gedimino technikos universitetas, Verslo vadybos fakultetas, Saulėtekio al. 11, LT-10223 Vilnius, Lietuva*
* *El. paštas akvile.eremine@stud.vilniustech.lt*

Gauta 2023-02-20; priimta 2023-06-15

Santrauka. Pinigų plovimo prevencija, siekiant atpažinti ir laiku sustabdyti šią nusikalstamą veiklą, vis dar išlieka viena aktualiausių temų globaliu mastu. Didelis duomenų kiekis, susijęs su pinigų plovimo reikalavimų laikymusi, ir didėjantis nusikalstamų metodų sudėtingumas verčia finansų įmones ieškoti naujų priemonių, kurios padėtų įvykdyti reguliavimo įsipareigojimus. Siekiant šio tikslo technologijos, kurių veikimas paremtas dirbtiniu intelektu, tampa neatsiejamos nuo pinigų plovimo prevencijos. Šiame straipsnyje pagrindžiama pinigų plovimo tikroji kaina pasaulinei ekonomikai. Tyrimo tikslas – įvertinti skaitmeninio intelekto pritaikymo svarbą pinigų plovimo prevencijoje. Šiam tikslui pasiekti buvo lyginamos skirtingos technologijos, taikomos finansiniam sukčiavimui atpažinti. Ypatingas dėmesys šiame straipsnyje skiriamas dirbtinio intelekto pogrupiui skaitmeninio intelekto ir mašininio mokymosi metodams. Remiantis naujausiais moksliniais tyrimais, šių technologijų derinys leidžia numatyti ateities įvykius, sukurti tinkamas taisykles, priimti tikslingus sprendimus, komunikuoti su tiksline auditorija. Dirbtinio intelekto svarbą pinigų plovimo prevencijoje pabrėžia straipsnyje analizuojamas tyrimas, kuriame buvo apklausiami aukščiausio lygio vadovai, atsakingi už pinigų plovimo prevenciją. Respondentai išreiškė vienbalsę nuomonę, jog tolesnis dirbtinio intelekto taikymas pinigų plovimo prevencijos srityje yra prioritetas. Tačiau sykiu tyrimo rezultatai identifikavo respondentų įvardytas grėsmes, dėl kurių praktiškai dirbtinio intelekto metodai veikloje taikomi vangiai.

Reikšminiai žodžiai: pinigų plovimas, dirbtinis intelektas, skaitmeninis intelektas.

Įvadas

Pinigų plovimo prevencija, siekiant atpažinti šį finansinį nusikaltimą, vis dar išlieka viena aktualiausių temų globaliu mastu įmonėms, vykdančioms finansines operacijas, įskaitant bankus, finansų technologijų įmones, kriptovaliutų prekybos įmones ir kitas šio sektoriaus įmones. Šį faktą sustiprina tai, jog, 2021 metų duomenimis, Bazelio pinigų plovimo rizikos indeksas padidėjo nuo 5,22 iki 5,3 dešimtiems balų

sistemoje (Basel Institute on Governance, 2021). Taip pat, remiantis naujausiais Jungtinių Tautų duomenimis, apskaičiuota, jog kasmet išplaunama nuo 800 bilijonų iki 2 trilijonų dolerių, o tai sudaro nuo 2 proc. iki 5 proc. viso pasaulio BVP (United Nations, 2022). Siekdami kovoti su pinigų plovimo problema, bankai ir kitos finansų sektoriaus įmonės įpareigotos stebėti, įvertinti ir praneši apie įtariamą pinigų plovimą. Didelis duomenų kiekis, susijęs su pinigų plovimo reikalavimų laikymusi, ir didėjantis nusikalstamų metodų sudėtingumas verčia finansų įstaigas nuolat ieškoti naujų priemonių, kurios padėtų įgyvendinti reguliavimo įsipareigojimus. Siekiant šio tikslo, technologijos, kurių veikimas paremtas dirbtiniu intelektu, tampa neatsiejamos nuo pinigų plovimo prevencijos. Tai pagrindžiama prognozėmis, jog pasaulinė kovos su pinigų plovimu įrankių rinka nuo 2020 m. iki 2025 m. augs 15,6 proc., vadinasi, nuo 2,2 mlrd. USD 2020 m. iki 4,5 mlrd. USD 2025 m. (Anti-money Laundering Market, 2022).

Tyrimo problema – kaip, naudojantis skaitmeniniu intelektu, sukurti įrankiai gali būti pritaikomi pinigų plovimui atpažinti?

Tyrimo objektas – skaitmeninio intelekto metodai.

Tyrimo tikslas – išanalizuoti skaitmeninio intelekto modelių pritaikymo galimybes pinigų plovimui atpažinti.

Tyrimo uždaviniai:

1. Pagrįsti tikrąją pinigų plovimo prevencijos kainą pasaulinėje ekonomikoje.

2. Apibrėžti skaitmeninio intelekto sampratą ir pranašumą prieš kitas technologijas.
3. Palyginus moksliniuose tyrimuose taikomus skaitmeninio intelekto metodus, kurie galėtų būti pritaikyti pinigų plovimui atpažinti, identifikuoti tinkamiausius metodus.

Tyrimo metodai: mokslinės literatūros analizė, lyginamoji analizė, aprašomoji analizė, duomenų aprašymas.

1. Pinigų plovimo samprata

Pinigų plovimo (angl. *Money Laundering*) reiškinys pirmą kartą užfiksuotas JAV 1980 m. pabaigoje, kai buvo pastebėtas didesnis nei įprastai narkotikų ir iš jų gaunamų pajamų srautų augimas. Pinigų plovimas buvo siejamas su procesu, kai nelegaliai gautos pajamos buvo verčiamos legaliomis (Georgieva, 2020). Šio proceso metu slepiama neteisėtų pajamų kilmė, neteisėtos pajamos naudojamos ir maskuojamos, kad atrodytų teisėtos. Ekonomistė Vainienė (2008) paantrina, kad paprastai pinigai plaunami vykdant sudėtingas pinigines operacijas, kuriomis siekiama paslėpti tikrąją pinigų kilmę ir dažniausiai grynuosius pinigus paversti negrynaisiais (Zhang & Trubey, 2018).

Tokia neteisėta veikla, pasak Unger (Under et. al., 2006), turi ne mažiau kaip aštuoniolika skirtingų pinigų plovimo sąvokų apibrėžčių. Šis faktas įrodo šios nusikalstamos veiklos nagrinėjimo aktualumą įvairiose srityse. 1 lentelėje pateiktos oficialios teisinėje veikloje vartojamos pinigų plovimo sąvokų apibrėžtys.

Apibendrinant pinigų plovimo sąvokų analizę, išskirtini keli aspektai. Pirma, tiek analizuoti autoriai, tiek teisės aktai pabrėžia tai, jog pinigų plovimas yra neteisėtu būdu gauti pinigai. Antra, šis reiškinys dažnu atveju įvardijamas kaip procesas, tai reiškia, kad jis yra sudėtingas, daugiasluoksnis veiksmas. Trečia, šia nusikalstama veikla siekiama nuslėpti neteisėto turto kilmę arba padėti bet kokiam nusikalstamoje veikoje dalyvaujančiam asmeniui išvengti atitinkamai teisinių tokios nusikalstamos veikos pasekmių.

2. Tikroji pinigų plovimo kaina pasaulinei ekonomikai

Tikrosios pinigų plovimo kainos ir masto nustatyti neįmanoma dėl kelių priežasčių, kurias įvardija Pol (2020).

Pinigų plovimą ypač sunku nustatyti dėl šių toliau išvardytų veiksnių:

1) dėl pinigų plovimo proceso daugiasluoksniškumo;
2) dėl nuolat kintančių nusikaltėlių vykdomų schemų, kurias nusikaltėliai naudoja siekdami apeiti esamus reguliavimo mechanizmus ir atitikties procesus;
3) dėl išaugusio duomenų srauto finansų institucijų paslaugoms persikėlus į skaitmeninę erdvę, dėl šios situacijos finansų įstaigos priverstos ieškoti naujų technologijų duomenims apdoroti;
4) dėl finansų institucijų reguliavimo ir nuolatinio lenktyniavimo tarpusavyje, siekiant sukurti reguliavimo ir atitikties metodus, kurių taikymo sritis būtų plati, naujoviška ir veiksminga, kad būtų galima stebėti ir prireikus laiku sustabdyti neteisėtą veiklą.

Šiuolaikinė kova su pinigų plovimu yra iš esmės neveiksminga (Pol, 2020; Economist, 2021). Finansinės atskaitomybės, skaidrumo ir sąžiningumo komisija (FACTI) – su Jungtinėmis Tautomis (toliau – JT) susijusi institucija, kurios tikslas – pasiekti tvarų vystymąsi kovojant su finansiniais nusikaltimais – apskaičiavo, kad

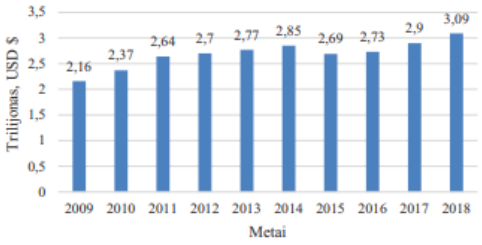
1 lentelė. Pinigų plovimo sampratos teisinės sąvokos

Autorius ir metai	Sąvokos paaiškinimas
Lietuvos Respublikos pinigų plovimo prevencijos įstatymas, 2021	1) turto teisinės padėties pakeitimas arba turto perdavimas žinant, kad šis turtas yra gautas iš nusikalstamos veikos arba dalyvaujant tokioje veikoje, siekiant nuslėpti arba užmaskuoti neteisėtą turto kilmę arba siekiant padėti bet kokiam nusikalstamoje veikoje dalyvaujančiam asmeniui išvengti teisinių šios veikos pasekmių; 2) turto tikrojo pobūdžio, tikrosios kilmės, šaltinio, vietos, disponavimo, judėjimo, nuosavybės ar kitų su nuosavybe susijusių teisių nuslėpimas arba užmaskavimas žinant, kad šis turtas yra gautas iš nusikalstamos veikos arba dalyvaujant tokioje veikoje; 3) turto įgijimas, valdymas ar naudojimas, įgijimo (perdavimo) metu žinant, kad šis turtas gautas iš nusikalstamos veikos arba dalyvaujant tokioje veikoje.
Lietuvos Respublikos baudžiamojo kodekso (toliau – BK) 216 straipsnis[6]	…tas, kas siekdamas nuslėpti ar įteisinti savo paties ar kito asmens pinigus ar turtą, žinodamas, kad jie įgyti nusikalstamu būdu, atliko su tuo turtu ar pinigais ar jų dalimi susijusias finansines operacijas, sudarė sandorius ar naudojo juos ūkinėje, komercinėje veikloje ar melagingai nurodė, kad tai gauta iš teisėtos veiklos, bus baudžiamas bauda arba laisvės atėmimu iki septynerių metų.

Šaltinis: sudaryta autorės, remiantis Lietuvos Respublikos pinigų plovimo prevencijos įstatymu bei Lietuvos Respublikos baudžiamuoju kodeksu.

dėl nusikaltėlių vykdomo pinigų plovimo kasmet prarandama 3,6 % pasaulio BVP (Financial Accountability Transparency & Integrity, 2020).

1 pav. šis faktas tik patvirtinamas. Tiriamuoju periodu nuo 2009 m. iki 2018 m. pajamos iš nusikalstamos veiklos nuosekliai auga, o tai reiškia, kad problema yra įsisenėjusi ir blogėjanti.



1 paveikslas. Jungtinių Tautų įvertintos pajamos globaliu mastu iš nusikalstamos veiklos sudarė 3,6 % BVP (Pol, 2020)

Analizuojant 2 lentelę svarbu akcentuoti, jog net 50 proc. konfiskuotų neteisėtų lėšų yra iš pinigų plovimo tiek Europos, tiek pasauliniu mastu. Taip pat net 50 proc. atitikčiai reguliuoti skiriamų lėšų yra paskiriama pinigų plovimo temai, nors kovos su pinigų plovimo sėkmės rodiklis Europos mastu yra vos 0,55 proc., o pasauliniu mastu – tik 0,05 proc.

2 lentelė. Pinigų plovimo prevencijos politikos patiriami kaštai (Pol, 2020)

	Europa (Eur)	Pasaulinis (USD)
Kasmet sugeneruojamos lėšos iš nusikalstamos veiklos	110 B	3 T
Kasmet konfiskuojamos neteisėtos lėšos / turtas	1,2 B	3 B
Konfiskuotų neteisėtų lėšų arba turto dalis iš pinigų plovimo	50 proc.	50 proc.
Kovos su pinigų plovimu sėkmės rodiklis	0,55 proc.	0,05 proc.
Atitikties išlaidos	144 B	304 B
Atitikties išlaidų tenkančių pinigų plovimui	50 proc.	50 proc.

Apibendrinant pateiktą statistiką apie pinigų plovimo prevenciją, galima teigti, kad šiandien vykdama politika, nukreipta prieš pinigų plovimą, reikalavimai ir kiti būtini žingsniai neduoda teigiamų rezultatų. Šis faktas tik sustiprina požiūrį, jog būtent naujos technologijos, tokios kaip dirbtinio intelekto metodų taikymas pinigų plovimo prevencijai gali ne tik sumažinti kaštus, patiriamus dėl gaunamų baudų už finansų institucijų pažeidimus, bet ir palengvinti specialistų kasdienę veiklą, proaktyviai užkertant kelią šiam finansiniam nusikaltimui.

3. Skaitmeninio intelekto taikymas pinigų plovimo prevencijai

3.1. Sukčiavimo atpažinimo technologijų taikymas pinigų plovimui atpažinti

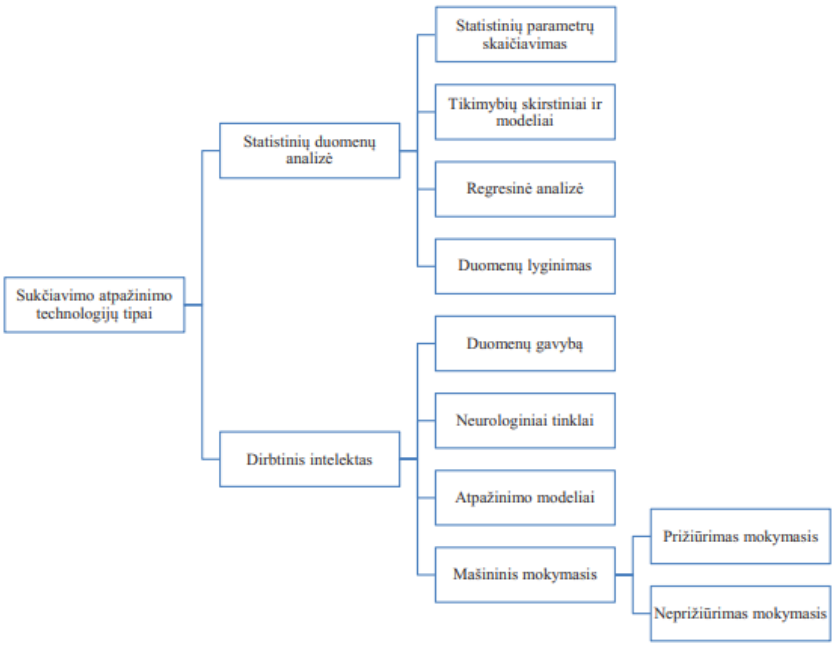
Šiame darbe nagrinėjama, kaip skaitmeninio intelekto (angl. *Computational Intelligence*, CI) technologijos gali prisidėti prie pinigų plovimo prevencijos. Pasak Donepud (2017), tradicinės dirbtinio intelekto (toliau – AI) technologijos susiduria su tokiomis problemomis, kaip didelis, vis sudėtingėjantis taisyklių rinkinys, taisyklių nesuvienodinimas tarp valstybių, todėl AI ne visada atliepia poreikį optimaliai mokytis ir laiku aptikti nuokrypius nuo standartų. Pasikartojančios AI klaidos atvėrė kelią skaitmeninio intelekto įrankiams atrasti (Sadiku, Foreman, Musa, 2018).

Pasak Kanade (2021), labai svarbu suprasti, kad sukčiavimo aptikimas – tai procesas, ar visuma veiksmų, kurių imamasi siekiant aptikti ir užblokuoti sukčių bandymą apgaule įgyti pinigų ar turto. Dažnu atveju tai daugiasluoksnis veiksmas, todėl įmonėms, teikiančioms finansines paslaugas, jį atpažinti tampa vis sunkiau ir klasikinių technologijų nebepakanka. Organizacijos diegia šiuolaikines sukčiavimo aptikimo ir prevencijos technologijas bei rizikos valdymo strategijas, siekdamos kovoti su augančiais nesąžiningais sandoriais įvairiose platformose (Kobadilov, Omarov, Yermekbayeva, 2020).

Mokslinėje literatūroje išskiriamos 2 pav. pateikti technologijų metodai taikomi sukčiavimui nustatyti (Kanade, 2021).

Sukčiavimo aptikimo statistinė duomenų analizė atlieka įvairias statistines operacijas: sukčiavimo duomenų rinkimas, sukčiavimo nustatymas ir sukčiavimo patvirtinimas atliekant išsamius tyrimus. Kalbant apie technologijas, kurių veikimas pagrįstas dirbtiniu intelektu, jos gali būti pritaikomos ne tik sukčiavimui atpažinti, bet ir sukčiavimo prevencijai, nes mokydamasis iš istorinių duomenų gali greičiau pastebėti atsikartojimą, kas praeityje buvo pripažinta pinigų plovimu. Šios sukčiavimo atpažinimo technologijos padeda įmonėms sustiprinti vidinį saugumą ir supaprastinti verslo procesus. Padidėjus efektyvumui, dirbtinis intelektas tapo pagrindine technologija, padedančia užkirsti kelią sukčiavimui (Duhart & Hernández-Gress, 2016).

Apibendrinant audito kompanijos *Deloitte* (2022) išvadas, kuriose teigiama, kad sudėtingėjant finansiniams produktams, sparčiai diegiant naujas technologijas, klasikinės sistemos nebeatliepia finansų institucijoms keliamų reikalavimų, todėl būtent dirbtinis intelektas geba



2 paveikslas. Sukčiavimo atpažinimo technologijų tipai (Kanade, 2021)

pagerinti įvairias galimybes siekiant atpažinti modelius, numatyti ateities įvykius, sukurti tinkamas taisykles, priimti tikslingus sprendimus, komunikuoti su tiksline auditorija.

3.2. Skaitmeninio intelekto metodų palyginimas

Prieš pradedant analizuoti skaitmeninio intelekto metodus, būtina apsibrėžti šią sąvoką. Skaitmeninis intelektas (toliau – CI) yra dirbtinio intelekto pogrupis, adaptacinių mechanizmų gebėjimas išmokti konkrečią užduotį iš besikeičiančių duomenų ar eksperimentinio stebėjimo (Siddique, 2013).

Mokslinėje klasikinėje literatūroje išskiriami šie pagrindiniai skaitmeninio intelekto modeliai: dirbtiniai neuroniniai tinklai (angl. *Artificial Neural Network*), evoliucinis skaičiavimas (angl. *Evolutionary Computation* ar *Genetic Algorithm*), neraiškiosios logikos sistemos (angl. *Fuzzy Logic*). Tačiau Donepud (2017) savo darbe papildomai akcentuoja ir šiuos modelius: Bajeso tinklas (angl. *Bayesian / Belief Networks*), mokymosi teorija (angl. *Learning Theory*), tikimybinė logika (angl. *Probabilistic Reasoning*).

Neuroniniai tinklai yra sudaryti iš neuronų, kurie vienas su kitu sujungti ir jungtimis perduoda signalus (Johnston, 2018). Veikimo esmė – tinklas priima įvesties signalą ir kelis kartus apdoroja, o tai suteikia gilesnio ir naudingesnio mokymosi galimybes (Vosyliūtė & Maknickienė, 2022). Šių metodų taikymas reiškia, kad algoritmas gali pagerinti užduoties atlikimą vis kartodamas naudojamus duomenis (Rajawat & Jain, 2020). Šio metodo taikymas praktikoje skirstytinas į penkias grupes: duomenų analizė ir klasifikacija, asociatyvioji

3 lentelė. Skaitmeninio intelekto sampratos analizė (Raj, 2019)

Autorius ir metai	Sąvokos paaiškinimas
Bezdek (1994)	Sistema skaičiavimo požiūriu vadinama intelektualia, jei ji tvarko žemo lygio duomenis, pvz., skaitmeninius duomenis, turi komponentus ir nenaudoja žinių dirbtinio intelekto prasme. Taip pat algoritmas rodo adaptyvius skaičiavimus.
Konar et al. (2006)	Skaitmeninis intelektas gali būti apibrėžtas kaip protingų įrankių ir skaičiavimo priemonių modelis, kuris gali tiesiogiai priimti neapdorotus duomenis ir tiesiogiai juos apdoroti paskirstytu būdu bei toleruoti netikslumą, neapibrėžtumą, dalinę tiesą.
Raj (2019)	Sąvoka „skaitmeninis intelektas“ paprastai reiškia kompiuterio gebėjimą išmokti konkrečią užduotį iš duomenų ar eksperimentinio stebėjimo. Skaitmeninis intelektas paprastai laikomas minkštojo skaičiavimo sinonimu.

atmintis, grupių modelių generavimas ir valdymas. Paprastai šiuo metodu siekiama išanalizuoti ir klasifikuoti medicininius duomenis, pereiti prie veido ir sukčiavimo nustatymo (Berniukevičius & Kurilovas, 2017).

Neraiškioji logika (angl. *Fuzzy Logic*) yra daugiareikšmės logikos forma, kurioje kintamųjų tiesos reikšmė gali būti bet koks realusis skaičius nuo 0 iki 1. Ji naudojama dalinės tiesos sąvokai nustatyti, kai tiesos reikšmė gali svyruoti nuo visiškai teisingos iki visiškai klaidingos (Majhi, 2019). Analizuojant mokslinę literatūrą akcentuojama, jog neraiškioji logika yra vienas labiausiai pritaikomų skaitmeninio intelekto metodų (Majhi, 2019). Ši paradigma susideda iš matavimų ir procesų modeliavimo, sukurtų sudėtingiems realaus gyvenimo procesams (Heidarinia et al., 2014). Priešingai nei AI, kuriam reikalingos tikslios žinios, skaitmeninis intelektas gali susidurti su nepilnumu ir duomenų nežinojimu proceso modelyje.

Ši technika paprastai taikoma įvairioms sritims, kurioms reikalinga kontrolė kaip vaizdo apdorojimas ir sprendimų priėmimas, kai naudojatės vaizdo kamera. Neraiškioji logika čia padeda stabilizuoti vaizdą, netvirtai laikant kamerą (Siddique, 2013).

Evoliucinis algoritmas – tai algoritmas, imituojantis natūralios atrankos procesą. Jis padeda spręsti optimizavimo ir paieškos problemas. Genetiniai algoritmai priklauso didesnei evoliucinių algoritmų klasei. Genetiniai algoritmai imituoja natūralius biologinius procesus, tokius kaip paveldėjimas, mutacija, atranka ir kryžminimas (Rada, 2008). Genetinių algoritmų sąvoka – tai paieškos metodas, dažnai taikomas informatikoje sudėtingiems, neakivaizdiems algoritminio optimizavimo ir paieškos problemų sprendimams rasti. Genetiniai algoritmai yra globalios paieškos euristika (Brabazon et al., 2008). Pasak Rados (2008), pagrindinis šio principo taikymas apima tokias sritis kaip kelių tikslų optimizavimas, kuriems tradicinės matematinės technikos nebepakanka pritaikyti įvairioms problemoms, tokioms kaip DNR analizė.

Bajeso tinklas yra tikimybinis grafinis modelis, vaizduojantis kintamųjų rinkinį ir jų sąlygines priklausomybes per nukreiptą aciklinį grafiką. Bajeso tinklai idealiai tinka įvykiui įvertinti ir numatyti tikimybę, kad bet kuris iš kelių galimų žinomų priežasčių buvo veiksnys (Fong-Rey Liu & Chen, 2011).

Mokymosi teorija aprašo, kaip mokiniai mokydamiesi gauna, apdoroja ir išlaiko žinias. Kognityvinė, emocinė ir aplinkos įtaka, taip pat ankstesnė patirtis turi įtakos supratimui ir pasaulėžiūros įgijimui ir žinių bei įgūdžių išlaikymui (Dunjko & Briegel, 2018).

Tikimybinė logika apima tikimybės ir logikos naudojimą neapibrėžtose situacijose. Tikimybinė logika išplečia tradicines logikos tiesos lenteles tikimybinėmis išraiškomis. Tikimybinės logikos sunkumas yra jų polinkis padauginti tikimybinių ir loginių komponentų skaičiavimo sudėtingumą (Triepels et al., 2018).

4 lentelėje lyginami skaitmeninio intelekto metodai.

Analizuojant mokslinę literatūrą pinigų plovimo atpažinimo prevencijos praktikoje išbandyti ir davę teigiamų rezultatų yra neraiškiosios logikos ir neuroninių tinklų metodų derinių pritaikymas, nes šių metodų derinys turi pranašumų numatyti, klasifikuoti, grupuoti ir

4 lentelė. Skaitmeninio intelekto technologijų palyginimas (Raj, 2019)

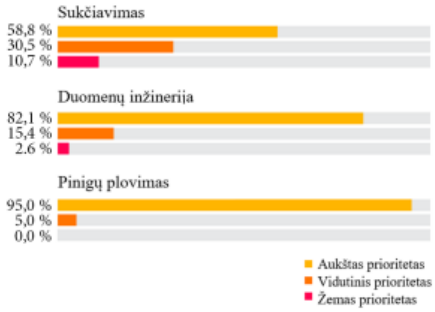
	Technologija	Aprašymas	Taikymas	Privalumai	Trūkumai
Skaitmeninis intelektas	Neraiškioji logika	Naudojamos aibės nariams įvertinti iš daugybės rinkinių	Patentų atpažinimas, kelių tikslų optimizavimas, žinių bazės sistemos	Užtikrina lengvą samprotavimą, įgyvendinimą ir gebėjimą valdyti neapibrėžtumus ir netiesiškumą	Patikimumo stoka
	Neuroniniai tinklai	Kuria paprastą matematinę sistemą, panašią į smegenis, išmoksta ją ir įgyja įžvalgų problemoms spręsti	Netiesinio proceso modeliavimas, struktūros numatymas, anomalijų aptikimas, sprendinių sudarymas	Atsparus gedimams, mokosi iš pavyzdžių, smulkūs pokyčiai neturi didelės įtakos	Reikalingas procesorius su lygiagrečiojo apdorojimo galimybe
	Evoliucinis algoritmas	Remiasi natūraliu atrankos procesu	Intelektuali paieška, kelių tikslų optimizavimas	Lankstus savaime prisitaikantis optimalus sprendimas	Ankstyva konvergencija, lemianti vietinį optimalumą
	Mokymosi teorija	Remiantis ankstesnių rezultatų išmokimu, kad būtų galima numatyti būsimus rezultatus	Diagnozė, aptikimas, nelaimių valdymas	Problemos sprendimas. Taikoma bet kuriai nuspėjamai mokymosi sferai	Trūksta detalumo
	Tikimybinė logika	Nustato galimą problemos rezultatą, pagrįstą ankstesnėmis įžvalgomis	Pramoninės diagnostikos prognozė	Tvarko neapibrėžtumą ir rizikos skaičiavimus	Brangus ir neefektyvus laiko požiūriu

optimizuoti, sutvarkyti duomenis (Jamshidi et al., 2019; Rajawat & Jain, 2020). Vertinant abstrakčiai, skaitmeninio intelekto modeliai parodo gebėjimą mokytis ir prisitaikyti prie naujų situacijų, atrasti neatitikimus, todėl šie metodai gali būti bandomi pritaikyti pinigų plovimui atpažinti ir prevencijai.

3.3. Dirbinio intelekto įrankių pritaikymo pinigų plovimo prevencijai svarba

Jungtinėse Amerikos Valstijose atliktame tyrime (The State of Fraud and Financial Crime in the U.S., 2022) dalyvavo 200 finansų institucijų darbuotojai, kurių turtas sudarė ne mažiau kaip 5 mlrd. dolerių. Apklausti darbuotojai užėmė vadovaujančiąsias pareigas sukčiavimo ir rizikos operacijų, pinigų plovimo, sukčiavimo strategijos, sukčiavimo analizės, technologijų ir duomenų srityse. 177 iš apklausos respondentų buvo atsakingi už sukčiavimo atpažinimo ir prevencijos valdymą, iš kurių 20 – už pinigų plovimo prevencijos valdymą. Kiti respondentai buvo atsakingi už duomenų mokslą ir technologijas.

Tyrimo dalyvių buvo prašoma įvardyti prioritetus pagal skirtingas sritis, kuriose vadovaujančiąsias pareigas užimantys darbuotojai teikia naujų sprendimų kurti arba aptikimui tobulinti ir prevencijos sistemoms, skirtoms kovai su sukčiavimu ir finansiniais nusikaltimais pagal specializacijos sritis (3 pav.).



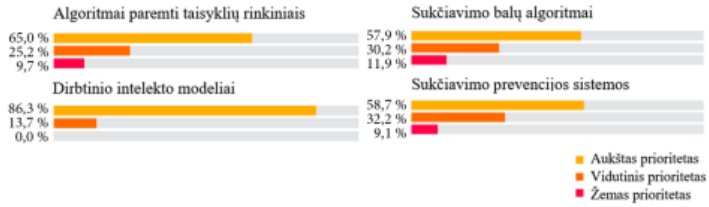
3 paveikslas. Naujovių ir tobulėjimo svarba pagal respondentų veiklos sritis (The State of Fraud and Financial Crime in the U.S., 2022)

Tyrimo duomenys atskleidžia, kad vadovaujančiąsias pareigas einantys darbuotojai pinigų plovimo prevencijos srityje beveik vienbalsiai – 95 proc. – mano, jog naujoviškų kovos sprendimų naudojimas pinigų plovimo prevencijai yra prioritetinis. Taip pat šios nuomonės laikosi vadovai, dirbantys finansinio sukčiavimo srityje: net 82 proc. mano, jog naujausios technologijos ir inovacijos yra viena iš galimybių, padedančių kovoti su finansiniais nusikaltimais. Duomenų mokslo vadovų tik daugiau nei 59 proc. laikosi nuomonės, jog tai prioritetinė sritis.

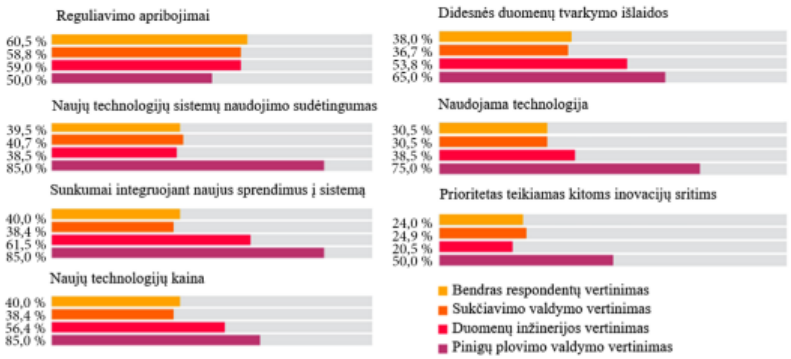
Įmonės, jau taikančios modernias technologijas, buvo linkusios ir toliau teikti pirmenybę šiuolaikinių pinigų plovimo (toliau – AML) ir kovos su sukčiavimo technologijų integracijai. Įmonės, kurios dar nėra integravusios naujų kovos su sukčiavimu ir AML sprendimų, galimas suvokimo sudėtingumas yra didelė modernizavimo kliūtis. Apklausti vadovai, kurie nepriėmė šiuolaikinių AML ir kovos su sukčiavimu strategijų, tai padarė ne todėl, kad nebuvo technologinių sprendimų, kuriuos būtų galima integruoti su esamomis technologijomis ar strategijas, bet todėl, kad buvo įsitikinę, kad tai padaryti bus per sunku. Atlikus apklausą nustatyta, kad 66 proc. vadovų išreiškė susirūpinimą, jog reguliavimo standartai yra per sudėtingi ir jų sprendimas gali neatitikti visų jų atitikties poreikių. Taigi 58 proc. manė, kad šiuolaikinės sukčiavimo schemos buvo pernelyg sudėtingos ir joks sprendimas negalėjo būti veiksmingas, o 49 proc. nerimauja, kad didėjantis sukčiavimo lygis priblokš bet kurią sistemą.

Technologijų, kurioms vadovai skiria dėmesį ir taiko veikloje (4 pav.), tyrimo duomenys atskleidė, jog 86 proc. įmonių, naudojančių dirbtinio intelekto įrankius, mano, jog tai prioritetinė sritis, duodanti teigiamų rezultatų kovojant su sukčiavimu. Taip pat 78 proc. apklaustųjų mano, jog debesijoje veikiančios sukčiavimo platformos lengvina jų kasdienę veiklą.

Svarbu suprasti ir veiksnius, stabdančius skaitmeninio intelekto taikymą pinigų plovimui atpažinti ir prevencijai. Tyrimo respondentai sutinka, kad šie veiksniai (5 pav.) slopina naujoves arba naujų funkcijų įtraukimą į esamus sprendimus. Iš pinigų plovimo prevencijos srityje vadovaujančiąsias pareigas užimančių asmenų



4 paveikslas. Naujų technologijų pritaikymo svarba (The State of Fraud and Financial Crime in the U.S., 2022)



5 paveikslas. Veiksniai slopinantys naujų technologijų diegimą (The State of Fraud and Financial Crime in the U.S., 2022)

net 85 proc. mano, jog naujų technologijų sistemų naudojimas yra sudėtingas, turi sunkumų integruojant naujus sprendimus į esamas sistemas. Net 75 proc. respondentų pasisako, jog naudojama technologija įvardijama kaip veiksnys, stabdantis naujų technologijų taikymą veikloje.

Apibendrinant tyrimą galima teigti, jog aukščiausio lygmens vadovai aiškiai supranta, kad naudojant dirbtinį intelektą sukurti sprendimai yra būdas suvaldyti pinigų plovimo nusikalstamą veiklą. Įmonės, teikiančios finansines paslaugas, ieško būdų, kaip kovoti su vis sudėtingesnėmis finansinių nusikaltimų formomis, taikydamos naujų technologijų sprendimus, tokius kaip dirbtinis intelektas. Tyrimo rezultatai paviešino veiksnius, kurie stabdo naujų technologijų diegimą pinigų plovimo prevencijoje, t. y. nusistatymas, jog dirbtinio intelekto metodai atrodo sudėtingi ir sunkiai pritaikomi.

3.4. Pinigų plovimo duomenų aprašymas

Šiame darbe naudojama pinigų plovimo atpažinimo duomenų bazė, prieinama kaggle.com. Dirbtiniu būdu sukurtos analizuojamos duomenų bazės modeliavimas pagrįstas trimis pinigų plovimo etapais:

1) pinigų įdėjimas;
2) pinigų sluoksniavimas;
3) pinigų integravimas.

Duomenų bazėje šie etapai kategorizuoti tipais, kai 1 taisyklė yra susijusi su pinigų išgryninimu (angl. *Cash-in*), o 2 ir 3 taisyklės – su pervedimu (angl. *Transfer*).

Iš pateiktų duomenų pinigų išsigryninimas sudarė 38 proc., pervedimas – 62 proc. visų duomenų. Sugeneruotų duomenų laikotarpis yra 6 mėn. Per visą laikotarpį buvo užfiksuota 59 proc. pinigų plovimo atvejų, iš kurių išsigryninant pinigus užfiksuota 57 proc. nelegalios veiklos, o pervedant pinigus – 62 proc. tipinių plovimo veiksmų. Dar detaliau analizuojant duomenis tiek antrame, tiek trečiame pinigų plovimo etapuose, nelegalių pervedimų skaičius buvo panašus. Taip pat pastebėta, jog pinigų plovimo veiksmai užfiksuoti pervedant ar išsigryninant mažesnes sumas.

Taigi šioje duomenų bazėje pateikti duomenys leidžia daryti prielaidą, jog daugiau atliktų pinigų plovimo veiksmų pastebėta pervedimo etape arba pinigų sluoksniavimo ir pinigų integravimo etapuose. Taip pat ši nusikalstama veikla atliekama pervedant mažesnes pinigų sumas.

Išvados

Vis didėjantys kaštai patiriami kovojant su pinigų plovimu ir tik 0,5 proc. sėkmingų pinigų plovimo atskleidimo atvejų konstatuoja faktą, jog vykdoma politika prieš pinigų plovimą, reikalavimai finansų įmonėms ir kiti būtini žingsniai neduoda teigiamų rezultatų. Šis faktas tik sustiprina požiūrį, jog būtent naujos technologijos, tokios

5 lentelė. Pinigų plovimo duomenų bazės struktūra (kaggle.com)

Duomenys reikšmė	Pinigų operacijos tipai	Šaltinio identifikacijos numeris	Vietos identifikacijos numeris	Pinigų suma	Data, laikas	Pinigų plovimas	Pinigų plovimo etapas
Kategorija	Išsigyninimas, pervedimas	–	–	–	–	1, 0	None; Type1; Type2; Type3

kaip dirbtinio intelekto metodų taikymas pinigų plovimui atpažinti ir prevencijai, gali ne tik sumažinti kaštus, kurie patiriami dėl gaunamų baudų už įmonių pažeidimus, bet ir palengvinti kasdienę specialistų veiklą proaktyviai užkertant kelią šiam finansiniam nusikaltimui.

Šiame darbe pasirinkta analizuoti dirbtinio intelekto pošakio skaitmeninio intelekto metodų taikymą pinigų plovimui atpažinti ir prevencijai, nes skaitmeninio intelekto įrankiai yra lankstesnis ir skaičiavimas remiasi „minkštaisiais skaičiavimo metodais“, kurie leidžia prisitaikyti prie daugelio situacijų. Taip pat taikomi metodai yra artimi žmogaus samprotavimo būdui, t. y. naudoja netikslią ir neišsamią informaciją bei sugeba prisitaikant atlikti kontrolės veiksmus, ko priešingai dirbtinio intelekto metodai nesugeba.

Išanalizavus mokslinę literatūrą apie skaitmeninio intelekto metodus, galima teigti, kad jie yra tinkami taikyti pinigų plovimo atpažinimo ir prevencijos metodus, nes technologijos rodo gebėjimą mokytis ir prisitaikyti prie naujų situacijų, atrasti neatitikimus bei susieti informaciją. Sykiu svarbu pabrėžti, kad skaitmeninio intelekto metodų vis daugėja, o tai rodo, kad jų veikimas yra pripažintas kaip naudingas ir praktiškai pritaikomas. Taip pat naujausiuose moksliniuose tyrimuose pabrėžiama, jog neuroninių tinklų ir neraiškiosios logikos modelių derinys leidžia tiksliau grupuoti, klasifikuoti informaciją ir sklandžiau optimizuoti užduotis.

Straipsnyje analizuoti tyrimo rezultatai paviešino veiksnius, kurie visgi stabdo naujų technologijų diegimą pinigų plovimo prevencijos srityje, t. y. nusistatymas, jog dirbtinio intelekto metodai atrodo sudėtingi ir sunkiai pritaikomi. Šiuos nuogąstavimus paneigti gali tik tolesnis šios temos tyrimas ir praktinis jos taikymas pinigų plovimo veikloje. Šio darbo tęstinė dalis – praktiškai lyginant skaitmeninio intelekto metodus, įvertinti efektyviausią metodą ar metodų derinį.

Literatūra

Anti-money Laudering Market. (2022). https://www.marketsandmarkets.com/Market-Reports/anti-money-laundering-solutions-market-95490454.html?gclid=CjwKCAjwm8WZBhBUEiwA178UnEPu4IQINJ45Df5rS12jGtpVAsafr7CCMOsyZ-ZYOWInX60-Bbj9mBoCUwkQAvD_BwE

Basel Institute on Governance. (2021). Basel AML Index 2021 (10th ed.). https://baselgovernance.org/publications/basel-aml-index-2021

Berniukevičius, A., & Kurilovas, E. (2017). Dirbtiniai neuroniniai tinklai personalizuotam mokymuisi. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiqya_sm7D6AhVnkIsKHaI2AXoQFnoECC4QAQ&url=https%3A%2F%2Fwww.zurnalai.vu.lt%2FLMR%2Farticle%2Fdownload%2F17755%2F16919%2F&usg=AOvVaw3mk3w2ve7B0CEmK6f2urge

Bezdek, J. S. (1994). What is Computational Intelligence? https://www.researchgate.net/publication/220045330_What_is_Computational_Intelligence

Brabazon, A., O'Neill, M., & Dempsey, I. (2008, November). An Introduction to Evolutionary Computation in Finance. In *IEEE Computational Intelligence Magazine*, *3*(4), 42–55. https://doi.org/10.1109/MCI.2008.929841

Deloitte. (2022). The new physics of financial services: How artificial intelligence is transforming the financial ecosystem. https://www2.deloitte.com/bd/en/pages/financial-services/articles/artificial-intelligence-transforming-financial-ecosystem-deloitte-fsi.html

Donepud, P. K. (2017). Machine learning and artificial intelligence in banking. *Engineering International*, *5*(2), 83–86. https://doi.org/10.18034/ei.v5i2.490

Duhart, B., & Hernández-Gress, N. (2016). Review of the Principal Indicators and Data Science Techniques used for the Detection of Financial Fraud and Money Laundering. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, pp. 1397–1398 https://doi.org/10.1109/CSCI.2016.0267

Dunjko, V., & Briegel, H. (2018). Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, *81*(7), 074001. https://doi.org/10.1088/1361-6633/aab406

Economist. (2021, April 12). The war against money-laundering is being lost: The global system for financial crime is hugely expensive and largely ineffective. https://www.economist.com/finance-and-economics/2021/04/12/the-war-against-money-laundering-is-being-lost

Financial Accountability Transparency & Integrity. (2020, September). *FACTI PANEL Interim Report, 2020*. https://uploads-ssl.webflow.com/5e0bd9edab846816e263d633/5f6b68c7bff4ad6cf6cb53a7_FACTI_Interim_Report_final.pdf

Fong-Rey Liu, K., & Chen, Jia-Shen. (2011). Prediction and assessment of student learning outcomes in calculus a decision support of integrating data mining and Bayesian belief networks. In *2011 3rd International Conference on Computer Research and Development*, Shanghai, China, pp. 299–303. https://doi.org/10.1109/ICCRD.2011.5764024

Georgieva, N. (2020). Concept, definition and characteristics of the money laundering phenomenon. *Journal of Process Management. New Technologies*, *8*(2), 23–37. https://doi.org/10.5937/jouproman8-26220

Heidarinia, N., Harounabadi, A., & Sadeghzadeh, M. (2014, July). An intelligent anti-money laundering method for detecting risky users in the banking systems. *International Journal of Computer Applications*, *97*(22), 35–39. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.662.6190&rep=rep1&type=pdf

Jamshidi, M. B., Gorjiankhanzad, M., Lalbakhsh, A., & Roshani, S. (2019). A novel multiobjective approach for detecting money laundering with a neuro-fuzzy technique. In *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, Banff, AB, Canada, 2019, pp. 454–458. https://doi.org/10.1109/ICNSC.2019.8743234

Johnston, M. (2018, September 28). Seizing the moment for artificial intelligence. https://blog.evergreengavekal.com/seizing-the-moment-for-artificial-intelligence/

kaggle. (n.d.). *Level up with the largest AI & ML community*. https://www.kaggle.com/

Kanade, V. (2021). What Is Fraud Detection? Definition, Types, Applications, and Best Practices. Fraud detection prevents fraudsters from obtaining money or property through false means. https://www.toolbox.com/it-security/vulnerability-management/articles/what-is-fraud-detection

Kobadilov, B., Omarov, G. B., & Yermekbayeva, D. D. (2020). Financial technologies trends and how they will shape financial markets [Article]. *The Economy: Strategy and Practice*, *15*(2), 151–157. https://esp.ieconom.kz/jour/article/view/208/204

Konar, A. (2006). *Computational intelligence: principles, techniques and applications*. Springer Science & Business Media,

Lietuvos Respublikos Seimas. (2021). Lietuvos Respublikos pinigų plovimo ir teroristų finansavimo prevencijos įstatymas. (2021, gruodžio 31, Nr. XIV-831). https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/addbfac066ed11ecb2fe9975f8a9e52e

Lietuvos Respublikos baudžiamasis kodeksas. (2021, rugsėjo 1d. Nr. 89-2741). https://www.e-tar.lt/portal/lt/legalAct/TAR.2B866DFF7D43/asr

Majhi, S. K. (2019). Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection. *Evolutionary Intelligence*, *14*(1), 35–46. https://doi.org/10.1007/s12065-019-00260-3

Money laundering data. (n.d.). https://www.kaggle.com/datasets/maryam1212/money-laundering-data?select=ML.csv

Pol, R. F. (2020). Anti-money laundering: The world's least effective policy experiment? Together, we can fix it. *Policy Design and Practice*, *3*(1), 73–94. https://doi.org/10.1080/25741292.2020.1725366

Rada, R. (2008). Expert systems and evolutionary computing for financial investing: A review. *Expert Systems with Applications*, *34*(4), 2232–2240. https://doi.org/10.1016/j.eswa.2007.05.012

Rajawat, A. S., & Jain, S. (2020). Fusion Deep Learning Based on Back Propagation Neural Network for Personalization. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9170693

Raj, J. S. (2019). A comprehensive survey on the computational intelligence techniques and its applications. *IRO Journals: Journal of IoT in Social, Mobile, Analytics, and Cloud*, *1*(3), Article-2. https://irojournals.com/iroismac/V1/I3/02.pdf

Sadiku, M. N. O., Foreman, J., & Musa, S. M. (2018). Computational intelligence. *European Scientific Journal*, July 2018 Edition, *14*(21), 56–60. https://doi.org/10.19044/esj.2018.v14n21p56

Siddique, A. (2013, April 22). *Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing*. John Wiley & Sons. https://doi.org/10.1002/9781118534823

Triepels, R., Daniels, H., & Feelders, A. (2018). Data-driven fraud detection in international shipping. *Expert Systems with Applications*, *99*(1), 193–202. https://doi.org/10.1016/j.eswa.2018.01.007

The State of Fraud and Financial Crime in the U.S. (2022). https://www.featurespace.com/the-state-of-fincrime-in-the-us-2022-report/

Unger, B., Siegel, M., Ferwerda J., Kruijf, W., Busuioic, M., & Wokke K. (2006). The amount and the effects of money laundering: Report for the Ministry of Finance February 16, 2006. Utrecht School of Economics and Australian National University. https://www.maurizioturco.it/bddb/2006_02_16_the_amounts_and_.pdf

United Nations. (2022). *Money Laundering*. https://www.unodc.org/unodc/en/money-laundering/overview.html

Vainienė, R. (2008). Ekonomikos terminų žodynas: apie 1400 terminų. Tyto Alba.

Vosyliūtė, I., & Maknickienė, N. (2022). Investigation of financial fraud detection by using Computational intelligence. In *12th International Scientific Conference "Business and Management" 2022, May 12–13*, Vilnius, Lithuania, pp. 390–397. https://doi.org/10.3846/bm.2022.787

Zhang, Y., & Trubey, P. (2018). Machine learning and sampling scheme: An empirical study of money laundering detection. Computational Economics, *54*,1043–1063. https://link.springer.com/article/10.1007/s10614-018-9864-z

INVESTIGATION OF MONEY LAUDERING DETECTION BY USING COMPUTATIONAL INTELLIGENCE

Akvilė EREMINĖ, Nijolė MAKNICKIENĖ

Abstract. The large amount of data related to compliance with money laundering requirements and the increasing sophistication of criminal methods are forcing financial companies to look for new tools to help them meet their regulatory obligations. To achieve this goal, technologies whose operation is based on artificial intelligence become inseparable from the prevention of money laundering. This article makes the case for the true cost of money laundering to the global economy. To achieve this, different technologies used in identifying financial fraud were compared. Special attention in this article is given to a subset of artificial intelligence – digital intelligence methods. According to the latest scientific research, these methods are based on soft computing, which allow to adapt to many situations, predict future events, create appropriate rules, and make targeted decisions. Also, the importance of artificial intelligence in the prevention of money laundering is emphasized by the US study analyzed in the article, in which top-level managers responsible for the field of money laundering prevention were interviewed. The respondents expressed a unanimous opinion that the further application of artificial intelligence in the field of money laundering prevention is a priority.

Keywords: money laudering, computional intelligence, artificial intelligence.