22-osios Lietuvos jaunųjų mokslininkų konferencijos „Mokslas – Lietuvos ateitis“ teminė konferencija
Proceedings of the 22nd Conference for Junior Researchers „Science – Future of Lithuania“

EKONOMIKA IR VADYBA / ECONOMICS AND MANAGEMENT

2019 m. vasario 13 d. Vilnius

eISSN 2029-7149
13 February, Vilnius, Lithuania Article Number: vvf.2019.022

Finansų valdymas http://jmk.vvf.vgtu.lt
Financial Management

KREDITINIŲ KORTELIŲ SUKČIAVIMO ATPAŽINIMO TYRIMAS

Greta PRATUZAITĖ[1*], Nijolė MAKNICKIENĖ[2]

[1]*Vilniaus Gedimino technikos universitetas, Verslo vadybos fakultetas, Finansų inžinerijos katedra, Saulėtekio al. 11, LT-10221,Vilnius, Lietuva*

[2]*Vilniaus Gedimino technikos universitetas, Verslo vadybos fakultetas, Finansų inžinerijos katedra, Saulėtekio al. 11, LT-10221,Vilnius, Lietuva*

Gauta 2019 m. sausio 17 d.; priimta 2019 m. vasario 4 d.

Santrauka. Įvairios institucijos susiduria su iššūkiu prisitaikyti prie naujųjų technologijų, įtraukti dirbtinį intelektą į savo vykdomą veiklą ir reaguoti į vykstančius pokyčius. Dirbtinis intelektas gali ženkliai palengvinti įvairių anomalijų, tokių kaip kreditinių kortelių mokėjimų sukčiavimas, stebėjimą bei rasti būdus kaip galima būtų išvengti vyraujančių sukčiavimų. Šio tyrimo tikslas yra aprašyti dirbtinio intelekto sąvoką finansų sektoriuje, naudojamus algoritmus bei ištirti duomenų bazę, kuri rasta Kaggle duomenų bazėje. Tyrimo metu buvo aprašyta, kas yra dirbtiniai neuroniniai tinklai, keli dažnai naudojami algoritmai finansų sektoriuje, pavyzdžiui, vektorių palaikymo mašinos, K–artimiausio kaimyno modeliai. Taip pat buvo pristatyta ir išanalizuota kreditinių kortelių transakcijų duomenų bazės duomenys, naudojami apmokant dirbtinį neuroninį tinklą. Analizuojant duomenų bazę, buvo naudota Python programavimo kalba, kad būtų braižomi grafikai. Tyrimo metu buvo susipažinta su dirbtiniais neuroniniais tinklais, buvo išsiaiškinta, kad tiriama duomenų bazė yra sudaryta iš realių transakcijų ir dėl konfidencialumo buvo transformuoti duomenys, tačiau duomenų bazė tinkama, norint apmokyti neuroninį tinklą.

Reikšminiai žodžiai: kreditinės kortelės, bankai, mokėjimai, sukčiavimai, dirbtinis intelektas.

Įvadas

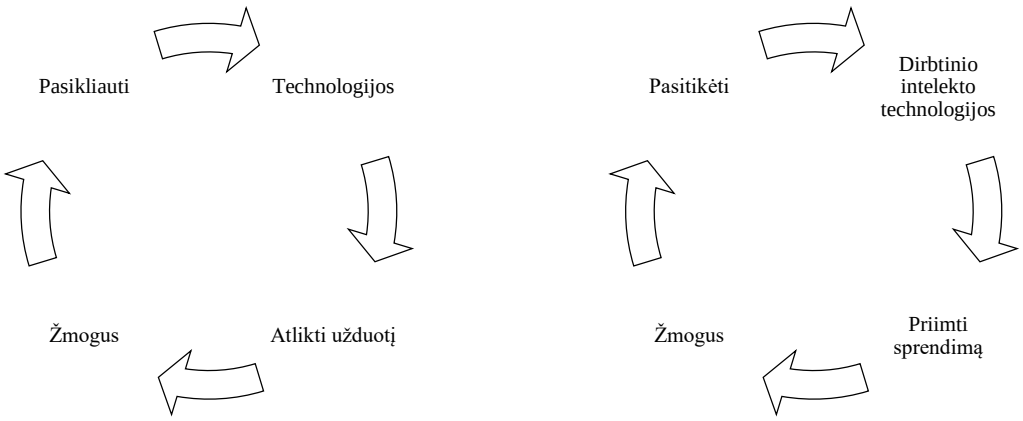
Turbūt kiekvienas pritartų, kad dabar inovacijos ir technologijos yra stipriai augančios ir besikeičiančios. Bet kad jų sėkmė būtų pasiekta, išties svarbus yra pasitikėjimas jomis. Žinomos kompanijos, tokios kaip „Apple”, „Amazon”, „Netflix” stengiasi protingai ir apdairiai rizikuoti su naujomis technologijomis, darbuotojams leidžia ieškoti išmanių ir kitokių idėjų, bet taip pat jos neužmiršta ir pasitikėjimo, kurį turi kelti klientui, svarbos (Botsman, 2017). Įvairios institucijos susiduria su iššūkiu prisitaikyti prie naujųjų technologijų, įtraukti dirbtinį intelektą į savo vykdomą veiklą ir reaguoti į vykstančius pokyčius. Pasaulis tampa vis labiau automatizuotas ir vis daugėja užduočių, priklausančių nuo robotų ir juos valdančių algoritmų. Žmonės ar pačios įmonės siekia sukurti programas, kurios veiktų lyg tikros žmogaus smegenys bei galėtų atlikti žmonėms prilygstančias užduotis.

Kiekviena įmonė stengiasi sukurti kažką geresnio bei efektyvesnio, o nuolatinis tobulėjimas skatina pokyčius pasaulyje. Atkreipus dėmesį į tai kas vyksta pasaulyje technologijose, dirbtinis intelektas jau mažai ką bestebina. Tačiau finansų sektoriuje, konkrečiau bankuose, dirbtinis intelektas plinta, bet ne taip stipriai kaip norėtųsi. Priežasčių šiam reiškiniui gali būti daug, tiek senesnės sistemos, dideli duomenų srautai, atsakomybės, apribojimai, darbuotojų kaita ir baimės, kad bus prarastas klientų pasitikėjimas. Jau kuris laikas ypač dideli nuostoliai yra susiję su kreditinių kortelių sukčiavimu. Bankai siekia atpažinti nesąžiningus mokėjimo kortelių sandorius, kad klientai nebūtų apmokestinami pirkiniais, kurių, pavyzdžiui, jie nepirko, kad nebūtų pasinaudota jų duomenimis ir vykdoma nesąžininga veikla. Ši problema išties svarbi ir yra naudojama daug dirbtinio intelekto algoritmų, kurie mokosi ant mokėjimo kortelių duomenų bazių. Šiame straipsnyje bus nagrinėjamas dirbtinis intelektas ir jo panaudojimas bankuose, pristatyti dažniausiai naudojami algoritmai bei kreditinių kortelių sukčiavimo duomenų bazė iš Kaggle duomenų bazės.

* Autorius susirašinėti. El. paštas *greta.pratuzaite@stud.vgtu.lt*

Dirbtinis intelektas finansiniame sektoriuje

Dirbtinis intelektas (angl. *Artificial Intelligence*), tai tam tikra mokslo šaka, kai yra kuriamos nuovokios programos, analizuojami taip vadinami „protingi agentai“ (angl. *Intelligent Agents*). Protingi agentai, tai tokios mašinos, kurios mąsto ir bando padidinti savo šansus įveikti tam paskirtas užduotis. Tokios mašinos siekia atkartoti žmogaus smegenų veikimą, bando atlikti veiksmus, kuriuos geba atlikti pats žmogus, pavyzdžiui, atskirti objektus ar žmogaus jausmus. Dirbtinį intelektą galima būtų įvardyti ir kaip kompiuterinių sistemų plėtrą, kai yra atliekamos užduotys, susijusios su žmogaus intelektu, pavyzdžiui, supratimas kalbos ar automobilio vairavimas (Botsman, 2018). Šis terminas gali dažnai klaidinti, bet jokiu būdu tai nėra žmogaus protas, jis tik imituojamas ir tik intelektiniai rezultatai yra taikomi dirbtiniam intelektui konkrečiose situacijose. Ateityje dirbtinio intelekto staigus šuolis pasitikėjime turėtų tik augti, tai byloja paveikslas, kuris pateiktas žemiau (1 paveikslas). Dabar žmogus vis dažniau pasikliauna technologijomis, kurios atlieka tam tikras užduotis, o ateityje šis pasitikėjimas tik augs, tačiau vis labiau bus įtraukiamas dirbtinis intelektas, kuris pats priims sprendimą, pateiks reikiamą informaciją žmogui ir taps vis protingesniu.



1 paveikslas. Pasitikėjimas dirbtiniu intelektu dabar ir ateityje (Botsman, 2017)

1.1. Dirbtinio intelekto panaudojimas ir svarba finansų sektoriuje

Finansų sektoriuje dirbtinis intelektas gali būti panaudojamas šiose srityse: buhalterija, auditas, reklama, valiutų keitimas, investicijos, klientų aptarnavimas, korupcija, kredito analizė, pinigų plovimas, sukčiavimo aptikimas ir pan. Bet gana dažnai tai yra susiję su pinigų apsauga, siekiant apsisaugoti nuo sukčiavimo. Būtent tokios kompiuterinės sistemos veikia taip, kad yra stebimos kliento išlaidos ir gaunami signalai, jei tik įvyksta kokie nors netikėti ir įtarimus keliantis mokėjimai. Pasak Oksfordo universiteto profesoriaus D. Kroening, bankai patiria didelį spaudimą iš aplinkos, įmonių, kurios neturi senos programinės įrangos ir tiek duomenų kaip bankai (Allan, 2018). Bankuose naudojami ir automatizuoti asistentai klientams, kurie teikia greitą pagalbą išspręsti savo klausimus per internetinę pranešimų sistemą, naudodami savo asmeninius kompiuterius ar išmaniuosius telefonus. Naudodamiesi mašininio mokymosi technologijomis, pokalbių svetainės nuolat tobulėja, atsižvelgiant į jų sugebėjimą tiksliai nustatyti kliento problemą ir atsakyti į klausimus. Todėl šiandien jie geba atpažinti daugybę bendrinių klausimų, kurių gali paklausti klientas.

Įvairūs bankai pasaulyje jau naudoja arba ruošiasi ateityje plėsti dirbtinio intelekto technologijas. Štai keletas pavyzdžių:

- „JPMorgan Chase” bankas naudoja programą pavadintą „LOXM”, kuri efektyviau ir greičiau vykdo klientų pavedimus bei „CoiN” platformą, kuri analizuoja teisinius dokumentus ir sugeneruoja kredito sutartis ženkliai greičiau – per kelias sekundes;
- Vokietijos skaitmeninis bankas „Number26“ naudoja programą pavadintą „Pulse26”, kuri analizuoja klientų istoriją ir prognozuoja jų galimus veiksmus;
- Britanijos bankas „Atom Bank“ naudoją „Virtual Agent“, bendraujančią su klientais ir atsakančią į jų užduodamus klausimus iki tam tikro lygio;
- Amerikos bankas taip pat naudoja asistentą pavadinimu „Erica”, kuris atlieka kasdienius sandorius, susijusius su finansiniais poreikiais. Jis naudoja dvi formas dirbtinio intelekto – natūralios kalbos apdorojimas, kad būtų suprantama kalba bei mašininį mokymąsi, kad būtų įmanoma surinkti įžvalgas pagal gautus kliento duomenis, kuriuos galima paversti naudingais patarimais (Crosman, 2018);
- Skandinavijos bankai irgi naudoja dirbtinį intelektą, kad būtų sumažintas krūvis darbuotojams. Dirbtiniai intelektai, kurių pavidalas yra kaip asistentų: „Aida” – SEB banke, „Nina” – Swedbank ’e, „Nova” – Nordea banke. Visi jie yra pavadinti moterų vardais iš dalies gal todėl, nes klientai jaučiasi patogiau, kai bendrauja su moterim.

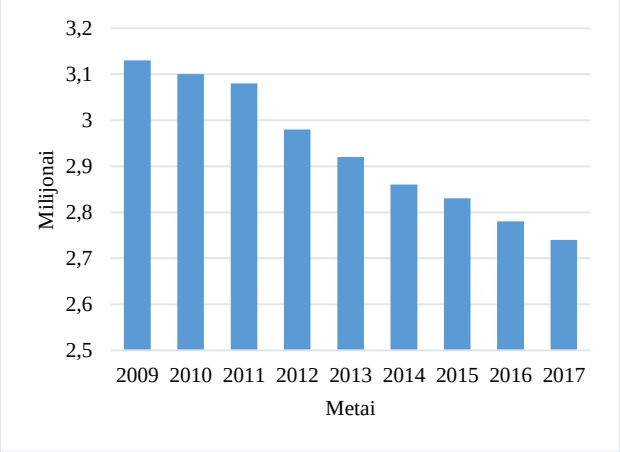
Pagal 2017 metais atliktą „PwC“ kompanijos „Global Digital IQ Survey“ apklausą, net 54 % organizacijų investuoja į dirbtinį intelektą ir toks skaičius taip išaugo per tris metus net iki 63 %. Tačiau bankai susiduria išties su dideliais iššūkiais, tai ir senesnės sistemos, kurias reikia pritaikyti, didžiulės duomenų bazės, klientų jautrumas, procesų patvirtinimas ir pasitikėjimas bei darbuotojų pasikeitimai, kurie gali sutrikdyti darbų eigą.

Taip pat „Accenture" kompanija pateikė ataskaitą apie bankines technologijų vizijas 2018, kur apskaičiavo, kad tokios naujosios technologijos kaip dirbtinis intelektas, turi būti naudojamos tikslingai, išauga būtinybė įgyti žmogiškųjų žinių, kad tikslai būtų įgyvendinami, todėl potencialiai gali augti organizacijų užimtumo lygis apie 10 procentų nuo 2018 iki 2022 metų. Be to net 77 procentai bankų planuoja naudoti dirbtinį intelektą savo veikloje, kad per ateinančius metus automatiškai atliktų reikiamas užduotis įvairiausiais mastų dydžiais.

Daug perspektyvų, susijusių su dirbtiniu intelektu yra pastebima kovoje su pinigų plovimu, kredito sprendimais ir įvairia analize, kur dirbtinio intelekto naudojimas būtų paprasčiausias, tačiau nėra plačiai naudojamas. Dažniausiai yra naudojami „Chat Box”, kur įvairūs asistentai komunikuoja su klientais iki tam tikro lygio.

Dirbtinio intelekto pagalba gali būti apdorojami labai dideli kiekiai duomenų, mažinamas žmogiškųjų klaidų faktorius. Taip pat vyrauja efektyvesnis išlaidų panaudojimas, suprantami realūs klientų poreikiai pagal įvairiausius kriterijus, šablonus, kuriuos arba sunkiau pastebėti žmonėms, arba iš vis yra nepastebima. Tačiau neišvengiama ir rizikos bei pavojų, kuriuos kelia ši technologija. Tikrai stinga pilno pasitikėjimo, kuris jau buvo minėtas kiek anksčiau, pasitaiko neteisingo vystymo ir tikrai reikia daug duomenų, kad būtų galima iš ko mokytis. Susiduriama ir su teisinių aktų trūkumu bei klaidų atsakomybė kelia daug dilemų, kas kaltas, kai įvyksta didelė žala.

Įsitraukimas į banko sektorių gali atnešti ne tik džiugesį, bet ir neišvengiamą rūpestį, kad bus mažinamas darbuotojų skaičius (Marria, 2018). Pavyzdžiui, įvykusi skaitmenizavimą iš tiesų sumažino personalo poreikį padaliniuose, vis daugiau jų buvo uždaroma. Remiantis Europos banko federacijos duomenimis 2017 metų pabaigoje banke dirbo apie 2,7 milijonai žmonių, maždaug 40 tūkstančių mažiau negu 2016 metais (2 paveikslas).



2 paveikslas. Darbuotojų skaičių pokyčiai Europoje (Sudaryta darbo autorių, remiantis European Banking Federation, 2017)

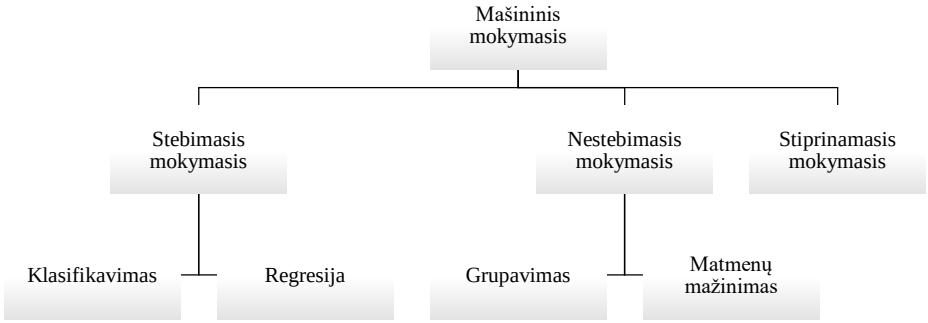
Dirbtinis intelektas ypač yra naudojamas kaip pokalbių dėžutės bei virtualūs asistentai. Jie keičia kai kuriuos darbus ir kuo daugiau jais bus naudojamasi, tuo daugiau jie patys apsimokys ir gerins savo gebėjimus pagelbėti klientams (Marria, 2018). Taip pat dirbtinis intelektas naudojamas atpažinti sukčiavimams bei pinigų plovimui. Būtent tokios įdiegtos sistemos bankuose atnešė daug sėkmės ir sutrumpina darbuotojų laiką bei padidina efektyvumą. Kibernetinės grėsmės yra viena iš svarbesnių rizikų, priklausančių finansinių paslaugų grupei, o pagal įvairius tyrimus kibernetinių duomenų ir duomenų saugumo didinimas – tai pirmasis pasaulinių bankų prioritetas. Dirbtinis intelektas ir pažangios duomenų analizės bus pagrindinis užkertant kelią kibernetinėms atakoms.

Remiantis „ThreatMetrix" duomenimis, skaitmeniniai sandoriai Europoje 2018 metų pradžioje įvyko 30 % daugiau elektroninių išpuolių nei praėjusių metų pirmąjį ketvirtį. Tad be viso saugumo valdymo, siekiant išvengti finansų sistemos pažeidžiamumo, svarbu mokytis ir itin gerai suprasti technologinius sprendimus. Taip pat kompanija „BBVA" nurodė, kad 69 procentai Europos kompanijų stinga pagrindinių žinių apie įtaką kibernetinių atakų, nepaisant to, kad 80 procentai jų turėjo tam tikrą kibernetinį incidentą per pastaruosius metus.

1.2. Neuroniniai tinklai

Gana dažnai dirbtinio intelekto nagrinėjimas prasideda nuo sąvokos mašininis mokymasis (angl. *Machine Learning*). Tai yra mokslo šaka, kuri dirbtinio intelekto pagalba kuria programas, kurios pačios gerina savo rezultatus, pačios apsimoko, kai yra naudojama vis daugiau duomenų. Gana greitai tai tapo kaip dažniausiai naudojamas dirbtinio intelekto mokymosi būdas (Chollet, 2017). Mašininis mokymasis gali būti skirstomas įvairiais aspektais, bet pagal

įvairius šaltinius galima rasti, kad jis yra skirstomas į stebimąjį, nestebimąjį bei stiprinamąjį mokymąsi (Jha, 2017; Hexagon, 2018).



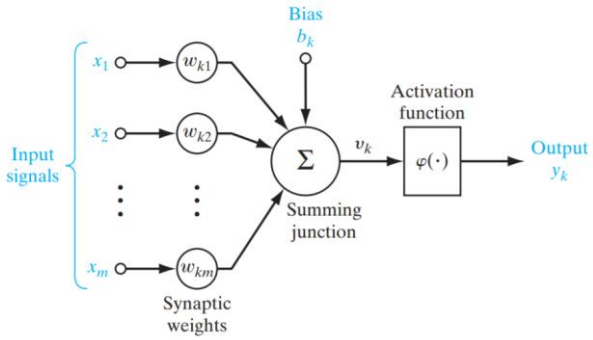
3 paveikslas. Mašininio mokymosi struktūra (Sudaryta darbo autorių, remiantis Jha, 2017)

Stebimasis mokymasis išsiskiria tuo, kad tam tikri duomenys yra pažymėti ir iš to yra ieškoma taisyklių – būdų kaip turėtų būti prognozuojami nauji duomenys. Tad pirmas žingsnis iš turimų duomenų suformuoti modelį jau su naujais duomenimis ir iš kurių po to yra duotas naujas rezultatas. Stebimasis mokymasis yra išskaidomas į šias kategorijas – klasifikaciją arba regresiją. Klasifikacijos tikslas yra naujus stebėjimus priskirti tam tikrai kategorijai, o regresijos – prognozuoti duomenis.

Nestebimasis mokymasis neturi duomenų pažymėjimo ir būtent dėl tos priežasties yra galimos tik tokios funkcijos: grupavimas (angl. *Clustering*) ir matmenų mažinimas. Grupavimas gali būti naudingas, kai yra grupuojami tam tikri objektai į sluoksnius (klasterius), pavyzdžiui, klientai skirstomi pagal savo galimybes ar įpročius. Taisyklių ieškojimas gali turėti naudos, kai, pavyzdžiui, klientai įsigyja vieną prekę ar paslaugą, bet atitinkamai tuo pačiu įsigyja ir dar kitą. Taip pat egzistuoja ir kitas mokymosi metodas – stiprinamasis (angl. *Reinforcement Learning*), kuris dažniau yra naudojamas akademiniais tikslais ar kuriant žaidimus.

Dirbtiniai neuroniniai tinklai išties primena žmogaus smegenis, nes informacija yra gaunama per mokymosi procesą ir patys neuronai geba prisiminti pateiktą informaciją ir ją kaupti (Haykin, 2009). Galima sakyti, kad paties roboto smegenys yra sudarytos iš programinės įrangos tinklo, kurio tikslas vis atkartoti žmogaus smegenyse esančias neuronų jungtis.

Neuroniniai tinklai yra sudaryti iš neuronų, kurie vienas su kitu yra sujungti ir pagal jungtis geba perduoda signalus. Veikimas susideda iš to, kad tinklas priima įvesties signalą ir kelis kartus apdoroja, kas suteikia gilesnį ir naudingesnį mokymąsi. Šių metodų naudojimas reiškia, kad algoritmas gali pagerinti užduoties atlikimą vis kartojant naudojamus duomenis. 4 paveikslas vaizduoja neurono modelį, iš kurio galima sudaryti dar didesnį neuroninį tinklą.



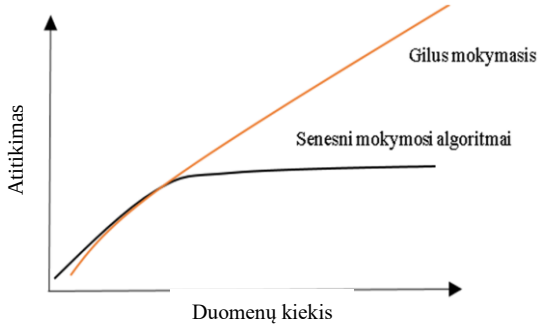
4 paveikslas. Neurono modelis (Haykin, 2009)

Kiekviena sinapsė turi savo svorį, žymimą wkj. Signalas xj prie sinapės j įvesties yra prijungtas prie neurono k, kai yra padauginama iš sinapsės svorio wkj (žr.(1)). Galiausiai po to visi įvesties signalai yra sudedami ir pasiekia aktyvavimo funkciją, skirta apriboti neurono išėjimo amplitudę ir įgyvendinti išėjimą, žymimą yk (Haykin, 2009).

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad (1)$$

Gilus mokymasis (angl. *Deep Learning*), tai tam tikras mašininio mokymosi tipas (Chollet, 2017). Šis mokymasis yra vadinamas gilusis vien dėl to, kad yra sudarytas iš daugiau neuroninių sluoksnių ir kuo daugiau duomenų yra

naudojama, tuo geresni rezultatai yra pasiekiami (Brownlee, 2016). Būtent tai leido sukurti be pilotes mašinas, garsų ir vaizdų atpažinimo programas, geresnę rezultatų paiešką sistemose ir pan.

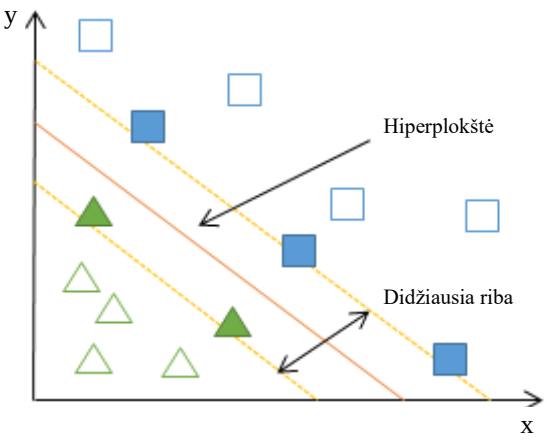


5 paveikslas. Gilaus mokymosi grafikas (Sudaryta darbo autorių, remiantis Brownlee, 2016)

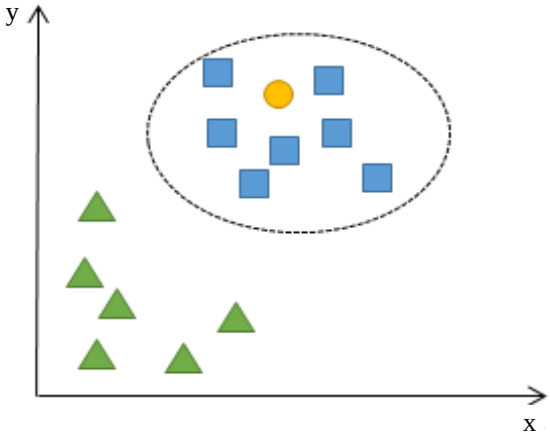
Dirbtinio neuroninio tinklo architektūra gali apibrėžti kaip veikia keli neuronai, kurie yra išdėstomi vienas kito atžvilgiu. Tokios architektūros yra struktūrizuotos kompozicijos. Tam tikros architektūros mokymas apima taikymą suderinti veiksmus siekiant sureguliuoti neuronų svorius bei slenksčius. Neuroninių tinklų architektūros pagal disponavimo funkciją, gali būti skirstomi į vieno sluoksnio, daugiasluoksnius, pasikartojančius bei apraizgytus tinklus (Silva et al., 2017).

1. Mokėjimo kortelių sukčiavimo atpažinimas taikant algoritmus

Nėra vieno konkretaus algoritmo, kuris išspręstų visas problemas, todėl yra daugybę kitų. Tinkamo algoritmo pasirinkimą dažnai lemia tokie veiksmai kaip duomenų apimtis, kintamųjų tipas, tikslai, trukmės apribojimai ir struktūra, tačiau vis tiek gana sunku nusakyti, kuris algoritmas yra pats geriausias, todėl svarbu yra apsibrėžti norimą problemą. Pagal įvairius straipsnius kiek žemiau buvo pasirinkti tokie algoritmai, kurie sprendzia problemas, susijusias su bankų sektoriais, kas šiuo atveju ir yra aktualiausia.



6 paveikslas. Vektorių palaikymo mašinos modelio pavyzdys (Sudaryta darbo autorių, remiantis Dnl Institute, 2015)



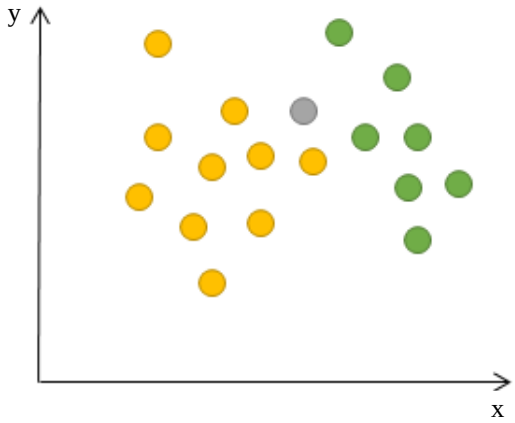
7 paveikslas. K–artimiausio kaimyno modelio pavyzdys (Sudaryta darbo autorių, remiantis Bernd Klein, 2018)

Atraminių vektorių metodas (angl. *Support Vector Machine*) – tai algoritmas, kuris atlieka klasifikavimo užduotis ir priklauso nuo sprendimo planų, kurie atskiria duomenis į skirtingas klases (dažnai į teigiamas ir neigiamas) (Seeja et al., 2014). Tikslas metodo, atrasti N–dimensijų erdvėje hiperplokštes (angl. *Hyperplane*), kurios klasifikuoja duomenis. Siekiant atskirti dvi duomenų klases, yra daugybė galimų hiperplokščių, kurias galima pasirinkti, o tikslas yra surasti tokią plokštę, kuri turėtų didžiausią ribą tarp abiejų duomenų klasių (Gandhi, 2018). Tai svarbu tam, kad būtų nustatyta tiksli duomenų klasė. Šį metodą vaizduoja 6 paveikslas.

K–artimiausio kaimyno metodas (angl. *K–nearest Neighbor*) – tai toks algoritmas, kuris klasifikuoja pagal artimiausius k (nurodytų kaimynų skaičius) kaimynus (7 paveikslas). Tada juos suklasifikuoja į naujus, pagrįstus bei panašius rezultatus. Kiekvienas naujas duomuo lyginamas su esamomis duomenimis naudojant atstumo metriką ir artimiausią esamą duomenį siekiama priskirti tam tikrai klasei (Seeja et al., 2014).

Naivusis Bajeso klasifikavimo metodas (angl. *Naive Bayes*), tai dar vienas algoritmas, kuris naudoja mokymosi

duomenų rinkinį tikslinių klasių prognozavimui (Seeja et al., 2014). Šis klasifikavimo metodas yra tarsi tikimybinis metodas, naudojantis klasės informaciją iš mokymosi duomenų rinkinio ir yra skirtas numatyti būsimas klases pagal tikimybes. Pavyzdžiui, žemiau pateiktame grafike yra pavaizduoti skirtingų klasių objektai ir kai atsiranda naujas objektas be klasės, yra skaičiuojama tikimybė, kuriai klasei jis yra priskiriamas.



8 paveikslas. Naïve Bayes klasifikavimo metodo pavyzdys (Sudaryta darbo autorių, remiantis Holczer, 2018)

Paslėptas Markovo modelis (angl. *Hidden Markov Model*) – tai toks stochastinis modelis, kur modelio objektai yra paslėpti ir modelio objektai gali grąžinti rezultatus, kurie yra stebimi ir analizuojami (Ghahramani, 2001). Šis modelis išsiskiria iš kitų, nes nesprendžia klasifikavimo problemų. Jis naudoja esamus rezultatus, gautus iš paslėptų duomenų ir naudodamas tikimybines funkcijas nusprendžia apie objekto būseną.

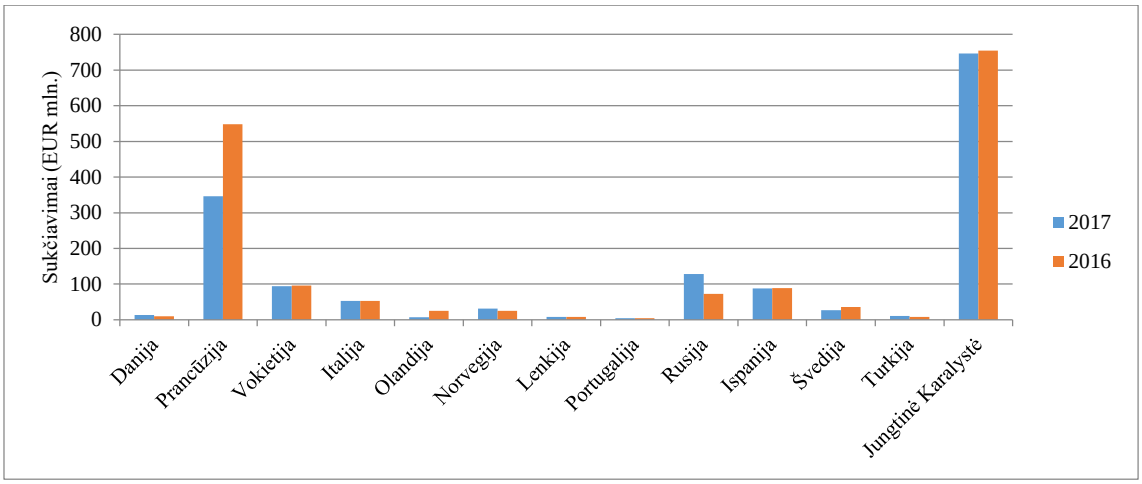
1.3. Mokėjimo kortelių sukčiavimas

Pastaraisiais metais vis didėjanti vartotojų paklausa skatina bankus vis labiau įdiegti skaitmenizavimą. Internetas ir mobilioji bankininkystė tampa įprastu įrankiu (kanalu) klientams. Vis dažniau naudojant šiuos kanalus sudaroma galimybė greitų mokėjimų įvykdymui, patogesniam ir greitesniam įvairių sprendimų ir galimybių naudojimui, tačiau tai didina galimybę nusikaltėliams pasinaudoti klientu ir jo duomenimis.

Kredito kortelių naudojimas yra plačiai paplitęs, o šios priemonės sukčiavimas pastaruoju laikotarpiu nuolat kyla. Finansiniai nuostoliai sukčiaujant paveikia ne tik bankus, bet ir individualius klientus. Jei bankas praranda pinigus, klientai galiausiai moka didesnes palūkanų normą ar didesnius mokesčius. Sukčiavimas taip pat gali turėti įtakos banko reputacijai, gali sukelti nefinansinius nuostolius, kuriuos kiek sunkiau kiekybiškai įvertinti per trumpą laiką, o ilgainiui gali tapti labiau pastebimi. Klientas nebegalės pasitikėti savo banku ir pasirinks kitą – patikimesnį konkurentą.

Europoje yra vykdomas prevencijos mechanizmas nuo sukčiavimo (Fico, 2017). Sukčiavimo išvengimui yra naudojamas bendradarbiavimas su kitais pramonės atstovais ir jų duomenimis, taip pat naudojamas mašininis mokymasis ir dirbtinis intelektas. Įvairūs algoritmai skaičiuoja galimus kliento veiksmus pagal jo vartojimą, stengiamasi atpažinti tokias situacijas, kurios yra nebūdingos tam klientui ir taip atskleisti, kad tai yra sukčiavimas. Tokių technologijų įsitraukimas į šią sritį turi didžiulės naudos, nes klientų duomenys yra apsaugomi, sumažinama rizika pinigų plovimui vykti bei sutaupomi kaštai darbuotojams. Šie įrankiai turi veikti įmonės lygiu, turi integruoti duomenis į tinkamą sprendimą, automatizuoti klientų autentifikavimą bei turi nebenaudoti žmogaus įsikišimo.

Tyrimus atliekanti kompanija „Fico“ pateikia statistiką, vaizduojančia pokyčius Europoje dėl mokėjimo kortelių sukčiavimo. Sumos, kiek pinigų yra prarandama, yra ganėtinai aukštos, o didžiausi praradimai egzistuoja Didžiojoje Britanijoje, Prancūzijoje ir Rusijoje, atitinkamai 747, 346 ir 128 milijonai eurų, kas yra pateikta kiek žemiau ir pagal kitas šalis.



9 paveikslas. Sukčiavimo statistika (Sudaryta darbo autorių, remiantis Fico, 2017)

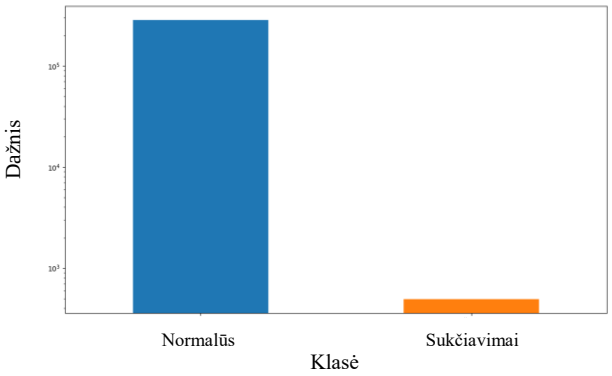
Egzistuoja du sukčiavimo aptikimo būdai – sukčiavimo analizė (angl. *Fraud Analysis*) ir naudotojo elgesio analizė (angl. *User Behaviour Analysis*) (Sorournejad et al., 2016). Sukčiavimo analizė yra atliekama, kai istoriniai duomenys yra analizuojami, stebima, ar įvykę sandoriai yra įprasti ir niekuo neišsiskiriantys, ar vykdoma nesąžininga veikla. Tokie duomenų rinkiniai yra sudaromi pagal klasifikavimo modelį ir yra naudojamas dviejų klasių klasifikavimas – taisyklių indukcija, sprendimų medžiai bei neuroniniai tinklai. Būtent šios priemonės aptinka daugiausiai sukčiavimo gudrybių. Kitas metodas yra susijęs su nekontroliuojama metodika, kuri priklauso nuo naudotojo elgesio. Mokėjimo nustatymas gali būti nustatomas apgaulingai, jei prieštarauja vartotojo elgesiui ir būtent dėl to reiktų turėti teisėta ir tikslų vartotojo elgesio modelį. Sukčiavimo analizė išsiskiria tuo, kad sukčiavimo analizės sistemos mokymasis ženkliai priklauso nuo sukčiavimo įrašų, bet ji negali aptikti naujų įrašų, o analizuojant elgesio analizę yra nukrypstama į naujų sukčiavimų atpažinimą. Tad nors elgesio analizė yra stipri, bet ji dažnai kenčia nuo melagingų signalų.

Technologijos, kurios buvo naudojamos siekiant užkirsti kelią sukčiavimui yra adreso tikrinimo sistemos (angl. *Address Verification System*), kortelės tikrinimo metodas (angl. *Card Verification Method*) ir asmens identifikavimo numeris (angl. *Personal Identification Number*) (Bhatla et al., 2003). Sukčiavimas yra aptinkamas, kai yra atsižvelgiama į kredito kortelių vykdytų operacijų rinkinį, procesų identifikavimą, ar įvykusi operacija priklauso suklastotai ar tikrai sandorių klasei (Maes et al., 2002). Pirmiausias yra svarbu patikrinti esmines sąlygas: ar tinkamas balansas ir įvertinti pagal parinktą modelį kiekvieną sandorį, ar jis yra su didele ar maža sukčiavimo rizikos tikimybe. Jei rizika yra didelė, asmenys yra informuojami perspėjimais. Tyrėjai tikrina šiuos įspėjimus ir pateikia kiekvieno perspėjimo atsakymą, ar tikras sukčiavimas, ar ne. Tada modelį reikia tobulinti ir jis gali būti pagrįstas taisyklėmis, kurioms reikia žmonių priežiūros.

Galime efektyviai naudotis mašinų mokymosi metodika nustatant apgaulės atvejus, kurie yra nustatomi pagal modelį, remiantis duomenų bazėmis. Dažniausiai modelis yra parametrų funkcija, kuri leidžia prognozuoti sandorio sukčiavimo tikimybę. Sukčiavimo aptikime mokymosi metodų naudojimas yra patogus įrankis, nes leidžia aptikti modelius didelės apimties duomenų bazėse, o kiekvieną sandorį apibrėžia daugybė kintamųjų. Taip pat, nesąžiningi sandoriai dažnai koreliuoja tiek laiko, tiek erdvės atžvilgiu. Mokymosi metodai gali būti naudojami nustatant ir modeliuokite egzistuojančias nesąžiningas strategijas bei nustatant naujas strategijas, susijusias su neįprastais kortelių savininkų elgesiais. Kai aptinkamas sukčiavimo metodas, nusikaltėliai pritaiko savo strategijas ar bando visai kitas. Tuo pačiu metu kasdien nauji nusikaltėliai žaidime dalyvauja ir bando naujas ir senas strategijas.

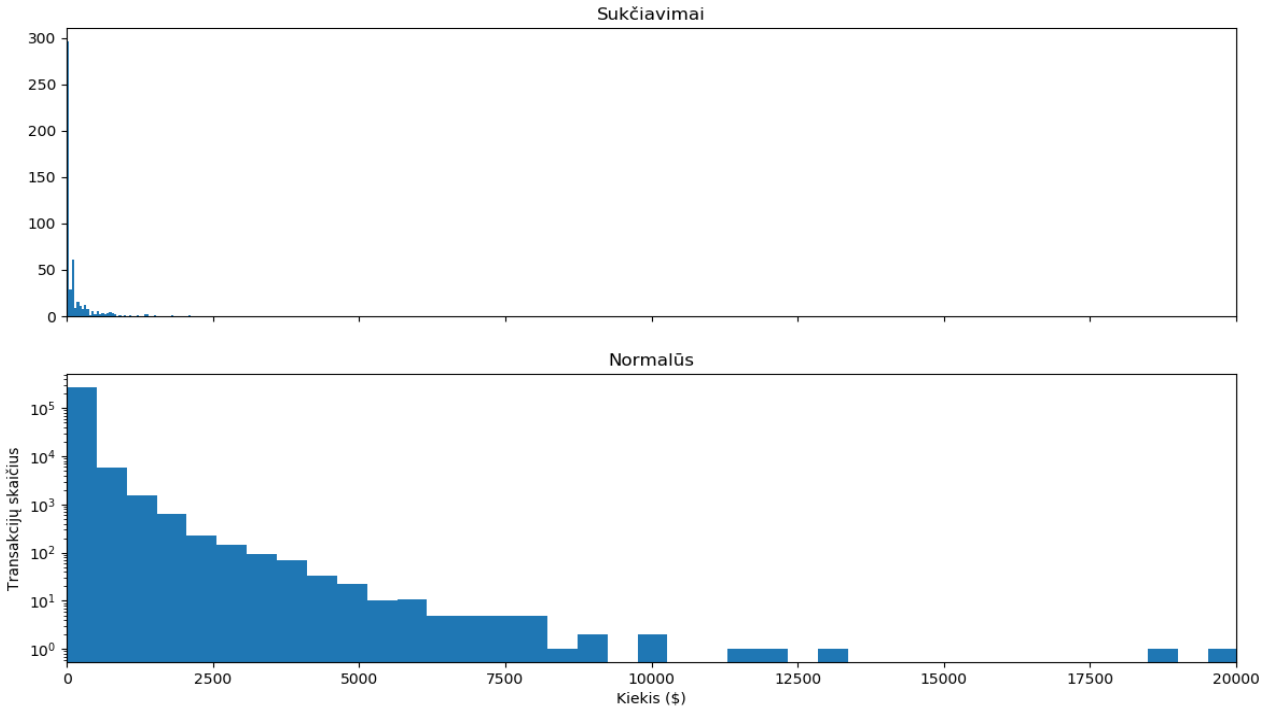
2. Mokėjimo kortelių atpažinimui naudojamų duomenų bazių tyrimas

Yra naudojama daug dirbtinio intelekto algoritmų, kurie gali naudoti mokėjimo kortelių duomenų bazes. Iš Kaggle duomenų bazių sistemos, buvo pasirinkta duomenų bazė, susijusi su kreditinių kortelių sukčiavimu. Šie duomenys apie kreditinių kortelių savininkų operacijas Europoje yra 2013 metų rugsėjo mėnesio. Tačiau šis duomenų rinkinys rodo operacijas, kurios yra įvykusios per dvi dienas bei yra įvykdyti 492 apgaulingi veiksmai iš 284 807 operacijų. Duomenų bazės analizei ir grafikų braižymui buvo naudota Python programavimo kalba.



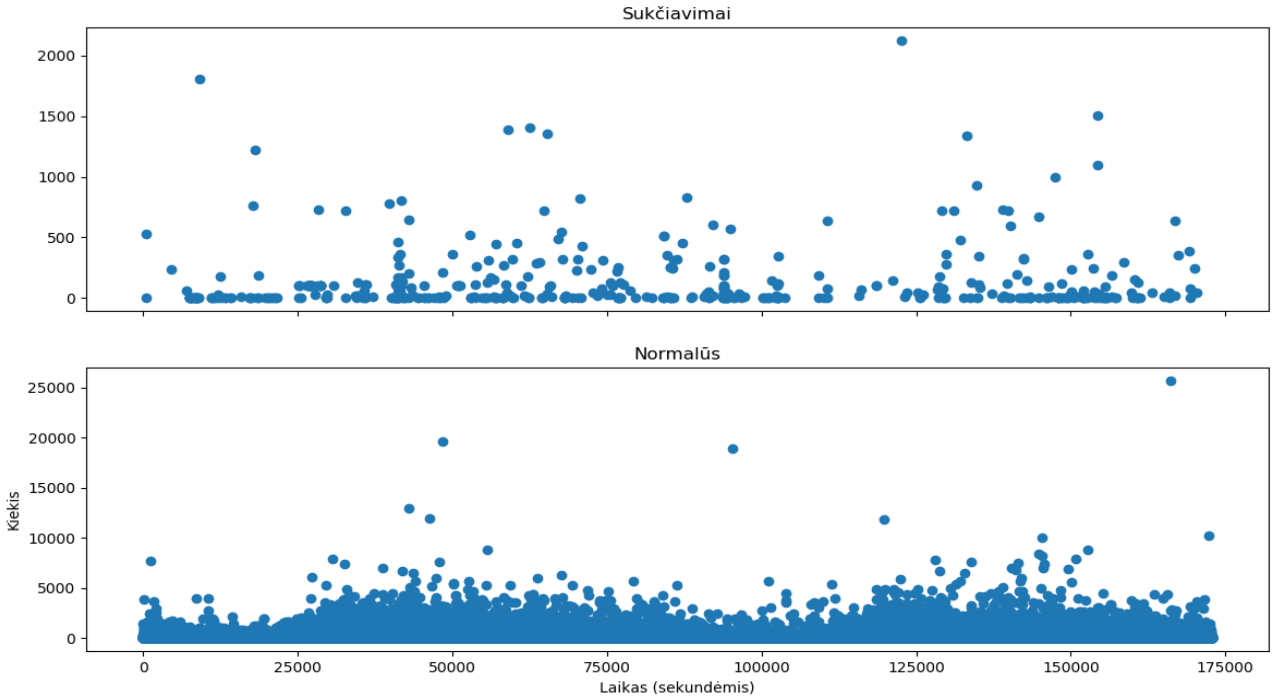
10 paveikslas. Mokėjimo kortelių paskirstymas (Sudaryta darbo autorių, remiantis Kaggle duomenų baze)

Remiantis 10 paveikslu, kuriame yra pavaizduotos dvi mokėjimo klasės (normalios transakcijos ir sukčiavimai), duomenų rinkinys yra nesubalansuotas, nes teigiama klasė (sukčiavimai) sudaro 0,172 % visų sandorių. Dėl konfidencialumo negalima pateikti pagrindinių funkcijų ir kitos papildomos informacijos apie įvykusius sandorius, todėl duomenys buvo transformuoti pagal pagrindinę komponentų analizę (angl. *Principal Component Analysis*) ir būtent dėl šios priežasties duomenys turi tik skaitinę vertę. Ši duomenų bazė susideda iš V1–V28 kintamųjų, kurie buvo gauti naudojant pagrindinę komponentų analizės funkciją. 11 paveikslas vaizduoja transakcijų pasiskirstymą pagal sukčiavimus ir ne sukčiavimus, kas akivaizdžiai byloja, sukčiavimų skaičius yra ženkliai mažesnis.



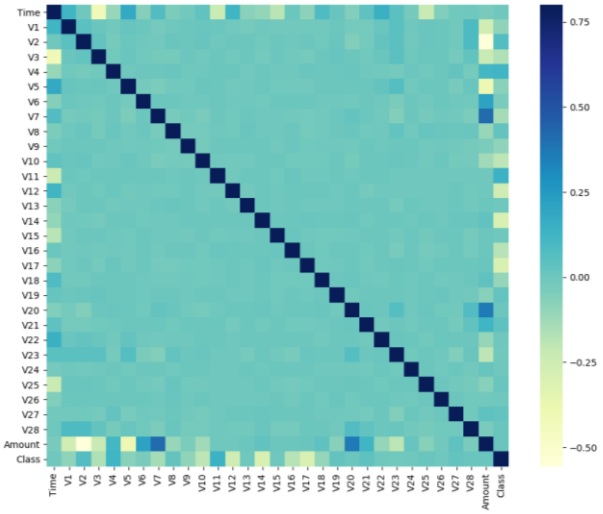
11 paveikslas. Transakcijų skaičius pagal klases (Sudaryta darbo autorių, remiantis Kaggle duomenų baze)

Taip pat šiame duomenų rinkinyje yra nurodyta sandorio suma bei klasė, kuri vaizduoja, ar sandoris buvo sukčiavimas ar ne. Jei sukčiavimas, tada transakcija yra žymima – 1, jei ne – 0. Taip pat svarbus vienas komponentas tai laikas, kuris šioje duomenų bazėje yra atvaizduojamas sekundžių tikslumu po kiekvieno įvykusio mokėjimo (12 paveikslas).



12 paveikslas. Mokėjimo judėjimai laiko atžvilgiu pagal klases (Sudaryta darbo autorių, remiantis Kaggle duomenų baze)

Žemiau pavaizduota koreliacijos matrica rodo, kad nė vienas iš V1–V28 komponentų neturi tarpusavio koreliacijos. Tačiau, jei atkreipsime dėmesį į klasę, ji turi tam tikrą formą teigiamą ir neigiamą koreliaciją su visais V komponentais, bet neturi ryšio su laiku ir suma.



13 paveikslas. Komponentų koreliacija (Sudaryta darbo autorių, remiantis Kaggle duomenų baze)

Duomenų nesikoreliavimas tarpusavyje reiškia, kad duomenys vienas nuo kito nepriklauso, pačios transakcijos yra unikalios ir nepriklausančios viena nuo kitos. Norint apmokyti neuroninius tinklus šis duomenų rinkinys yra kiek per mažas, nes sukčiavimo pavyzdžių nėra tiek daug, tačiau algoritmų analizei ši duomenų bazė yra tinkama.

Išvados

Naujausios technologijomis daro didžiulę įtaką finansų sektoriui. Ir jei pasitikėjimas senosiomis institucijomis imtų ir sumažėtų, žmogui rasti kuom pasitikėti išties pasidaro ganėtinai sunku. Bet šios institucijos nebūtinai turi išnykti, jos gali ir turi prisitaikyti prie pasaulyje vykstančių technologinių pokyčių. Pasaulis tampa vis labiau automatizuotas, bet klausimas kada gali įvykti tai, kad viskas priklausys nuo robotų ir juos valdančių algoritmų. Toks dalykas sukeltų

sąlygas vis mažesnėms abejonėms, bet taip pat gali tapti ir itin pavojingas. Tad iš tiesų svarbu suprasti ir nuspręsti, kur, kiek ir kada kompiuterinis kodas gali kontroliuoti situaciją.

Darbo metu buvo susipažinta su dirbtiniais neuroniniais tinklais, bei apžvelgta kokie yra dažniausiai naudojami algoritmai, norint sukurti dirbtinį neuroninį tinklą, kuris geba atpažinti tinkamas ir klaidingas transakcijas. Buvo ištirta duomenų bazė ir nubraižyti tam tikri grafikai. Matyti, kad duomenys tarpusavyje nesikoreliuoja ir jei būtų siekiama ateityje apmokyti neuroninius tinklus yra kiek per mažai pateiktą sukčiavimo pavyzdžių, tačiau algoritmų skaičiavimams ši duomenų bazė yra tinkama, nes naudojama iš tikrų duomenų ir duomenys tarpusavyje vienas nuo kito nepriklauso, kas reiškia yra daugybe unikalių mokėjimų, kur jų tinkamumas nepriklauso nuo kitų transakcijų. Kreditinių kortelių sukčiavimas yra opi problema ir būtent todėl yra įtraukiamas dirbtinis intelektas su įvairiausiais algoritmais, kad ši problema būtų išspręsta. Tačiau siekiant nustatyti nesąžiningas strategijas, vis sukuriami kiti modeliai ir kitos nusikaltėlių strategijos. Tad analizuojant įvairiausias duomenų bazes ir algoritmus yra susiduriama su iššūkiais, nes tokios duomenų bazės yra labai konfidencialios bei duomenys būna transformuojami, todėl tai analizuoti pasidaro kiek sudėtingiau.

Literatūra

Accenture. (2018). *Building the future – ready bank. Banking technology vision 2018.* Retrieved from https://www.accenture.com/gb-en/_acnmedia/PDF-78/Accenture-Banking-Technology-Vision-2018.pdf

Allan, K. (2018). *How banking is adopting and using AI technology.* Retrieved from https://www.idgconnect.com/idgconnect/analysis-review/1018981/banking-adopting-ai-technology

Bhatla. T. P., Prabhu, V. & Dua, A. (2003). *Understand credit cards fraud.* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.431.7770&rep=rep1&type=pdf

Botsman, R. (2017). *Kuo galite pasitikėti? Kaip technologijos mus suartino ir kodėl gali išskirti.* Vilnius: Eugrimas.

Brownlee, J. (2016). *What is Deep Learning?* Retrieved from https://machinelearningmastery.com/what-is-deep-learning/

Chollet, F. (2017). *The impossibility of intelligence explosion.* Retrieved from https://medium.com/@francois.chollet/the-impossibility-of-intelligence-explosion-5be4a9eda6ec

Crosman, P. (2018). *Where Bank of America uses AI, and where its worries lie.* Retrieved from https://www.americanbanker.com/news/where-bank-of-america-uses-artificial-intelligence-and-where-its-worries-lie

Dnl Institute. (2015). *Support Vector Machine: Simplified.* Retrieved from http://dni-institute.in/blogs/support-vector-machine-simplified/

European Banking Federation. (2018). *Banking in Europe: EBF Facts & Figures 2018.*

European Banking Federation. (2018). *Financial technology.* Retrieved from https://www.ebf.eu/facts-andfigures/financial-technology/

Fico. (2017). *Evolution in card fraud in Europe 2017*. Retrieved from https://www.fico.com/europeanfraud/indexFraud_Detection_Applying_Bayesian_and_Neural_networks/links/0deec52519708c5f7a000000/Credit-Card-Fraud-Detection-Applying-Bayesian-and-Neural-networks.pdf

Gandhi, R. (2018). Support Vector Machine – Introduction to Machine Learning Algorithms. Retrieved from https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms934a444fca47?fbclid=IwAR3JEigdujaka NpbHrffwl9peyjEYMV9PfiRsiAPx0Yf-5_wp4GqdglnRRA

Ghahramani, Z. (2001). *An introduction to Hidden Markov Models and Bayesian Networks.* Retrieved from http://mlg.eng.cam.ac.uk/ zoubin/ papers/ijprai.pdf

Haykin, S. (2009). *Neural networks and learning machines.* New Jersey: Pearson Education,Inc.

Hexagon, C. (2018). *How did you benefit from machine learning today?* Retrieved from https://www.crimsonhexagon.com/blog/how-did-you-benefit-from-machine-learning-today/

Holczer, B. (2018). *Naive Bayes Classifier Explained Step by Step.* Retrieved from https://www.globalsoftwaresupport.com/naive-bayes-classifier-explained-step-step/

Jha, V. (2017). *Machine Learning Algorithm – Backbone of emerging technologies*. Retrieved from https://www.techleer.com/articles/203-machine-learning-algorithm-backbone-of-emerging-technologies/

Kaggle. (2018). Credit Card Fraud Detection. Retrieved from https://www.kaggle.com/mlg-ulb/creditcardfraud

Klein, B. (2018). k–Nearest–Neighbor Classifier. Retrieved from https://www.python-course.eu/k_nearest_neighbor classifier.php

Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002). *Credit card fraud detection using bayesian and neural networks*. Retrieved from https://www.researchgate.net/profile/Karl_Tuyls/publication/248809471_Credit_Card_

Marria, V. (2018). *Is Artificial Intelligence Replacing Jobs In Banking?* Retrieved from https://www.forbes.com/sites/vishalmarria/2018/09/26/is-artificial-intelligence-replacing-jobs-in-banking/#72eebc73c55

Pwc. (2017). Consumer Intelligence Series Bot.Me: A revolutionary partnership. Retrieved from http://pwcartificialintelligence.com/

Seeja, K. R. & Zareapoor, M. (2014). *FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining*. Retrieved from https://www.hindawi.com/journals/tswj/2014/252797/

Silva, I. N., Hernane Spatti, D., & Flauzino, R. A. (2017). *Artificial Neural Network Architectures and Training Processes.*

Chapter 2.

Sorournejad, S., Zojaji, Z., Atani, R. E., & Monadjemi, A. H. (2016). *A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective.* Retrieved from https://www.researchgate.net/publication/310610856_A_Survey_of_Credit_Card_Fraud_Detection_Techniques_Data_and_Technique_Oriented_Perspective

CREDIT CARDS FRAUD IDENTIFICATION INVESTIGATION

Greta PRATUZAITĖ, Nijolė MAKNICKIENĖ

Abstract. Different institutions face the challenge of adapting to new technologies, incorporating artificial intelligence into their own activities and monitoring the changes. Artificial intelligence can make easier monitoring of various anomalies, as well as counterfeiting credit card payments. The aim of this paper is to describe the artificial intelligence concept in the financial sector using the algorithms and to investigate the database that was found in the Kaggle database. During the work, artificial neural networks were described, and algorithms are often used in the financial sector (eg: Vector Support Machines, K–Neighbor Models). A credit card database was also introduced and analyzed, which can be used to teach an artificial neural network. When analyzing the database, Python programming language was used to draw graphs. During my work I managed to learn about artificial neural networks, it became clear that the database under study was made up of real transactions and the data was transformed due to confidentiality, but still the database is suitable for training the neural network.

Keywords: credit cards, fraud, investigation, banks, transactions, artificial inteligence.