



Kauno technologijos universitetas

Elektros ir elektronikos fakultetas

Savitvarkių neuroninių tinklų taikymas kompanijų vertybinių popierių analizei

Baigiamasis magistro projektas

Karolis Poškus

Projekto autorius

Prof. Rimvydas Simutis

Vadovas

Kaunas, 2019



Kauno technologijos universitetas

Elektros ir elektronikos fakultetas

Savitvarkių neuroninių tinklų taikymas kompanijų vertybinių popierių analizei

Baigiamasis magistro projektas

Valdymo technologijos (6211EX014)

Karolis Poškus

Projekto autorius

Prof. Rimvydas Simutis

Vadovas

Doc. Tomas Tekorius

Recenzentas

Kaunas, 2019



Kauno technologijos universitetas

Elektros ir elektronikos fakultetas

Karolis Poškus

Savitvarkių neuroninių tinklų taikymas kompanijų vertybinių popierių analizei

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Karolio Poškaus, baigiamasis projektas tema „Savitvarkių neuroninių tinklų taikymas kompanijų vertybinių popierių analizei“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Poškus, Karolis. Savitvarkių neuroninių tinklų taikymas kompanijų vertybinių popierių analizei. Magistro baigiamasis projektas / vadovas Prof. Rimvydas Simutis; Kauno technologijos universitetas, Elektros ir elektronikos fakultetas.
Studijų kryptis ir sritis (studijų krypties grupė): elektronikos inžinerija, inžinerijos mokslai.

Reikšminiai žodžiai: savitvarkiai neuroniniai tinklai, klasterizavimas, kompanijų vertybinių popierių analizė.

Kaunas, 2019. 50 p.

Santrauka

Darbe išanalizuotas dirbtinio intelekto metodų taikymas finansuose. Plačiau išnagrinėjama dirbtinių savitvarkių neuroninių tinklų (SOM) metodas ir šio metodo taikymas finansų rinkų ir vertybinių popierių analizei. Analizuojami savitvarkiai neuroniniai tinklai (SOM) ir programinė įranga SOM realizavimui. Pristatoma atlikto tyrimo metodika. Pateikiama sukurta schema kompanijų vertybinių popierių atrankai ir portfelio formavimui, pasitelkiant savitvarkius neuroninius tinklus. Atliktas keturių mėnesių rinkos stebėjimo tyrimas leidžia teigti, jog pasirinktas klasterizavimo algoritmas pasitelkiant SOM leidžia pakankamai gerai prognozuoti finansų rinkų kitimą.

Poškus, Karolis. Application of self-organizing neural networks to corporate securities analysis. Master's Final Degree Project/ supervisor Prof. Rimvydas Simutis; Faculty of Electrical and Electronics Engineering, Kaunas University of Technology.

Study field and area (study field group): electronic engineering, engineering sciences.

Keywords: self-organizing neural network, clustering, corporate securities analysis.

Kaunas, 2019. 50 pg.

Summary

The paper provides an overview of artificial intelligence methods applications in finance. The method of self-organizing neural networks (SOM) and the application of this method for the analysis of financial markets and securities is more widely explored. Software for SOM implementation is presented. Research methodology is provided. The created scheme is presented for the selection of company securities and portfolio formation through SOM. The four-month market monitoring survey suggests that the chosen clustering algorithm with the SOM allows for a good forecast of the evolution of financial markets.

Turinys

Lentelių sąrašas	7
Paveikslų sąrašas	8
Santrumpų sąrašas	9
Įvadas.....	10
1. Dirbtinio intelekto metodai.....	11
1.1.1. Dirbtinio intelekto taikymas finansuose.....	11
1.1.2. Dirbtinio intelekto rūšys	11
1.1.3. Mašininis mokymasis	11
1.1.4. Dirbtinio neurono modelis.....	12
1.1.5. Dirbtinis neuroninis tinklas	13
1.2. Savitvarkiai neuroniniai tinklai	14
1.2.1. Savitvarkių tinklų algoritmas	17
1.3. Savitvarkių neuroninių tinklų taikymas finansų rinkose	18
1.4. Programinė įranga	25
1.4.1. Viscovery SOMine	25
1.4.2. Matlab/Deep Learning Toolbox	26
1.4.3. VisiSOM.....	26
1.4.4. Neural Data Analysis (NDA)	27
1.5. Apibendrinimas	28
2. Savitvarkių tinklų taikymas kompanijų klasterizavimui	29
2.1. Tyrimo organizavimas.....	29
2.1.1. Rodiklių parinkimas.....	29
2.2. Vertybinių popierių atranka.....	31
2.3. Klasterizavimo algoritmo realizavimo žingsniai.....	34
2.4. Pelningiausio klasterio atranka.....	38
3. Pasiūlyto klasterizavimo algoritmo realizavimas ir testavimas	42
3.1. Rezultatų aptarimas	44
3.1.1. Portfelio rizikos įvertinimas	45
Rezultatai ir išvados	47
Literatūros sąrašas	48

Lentelių sąrašas

1 lentelė. Suminio pelno kitimas.	41
2 lentelė. SP&500 indekso ir pelningų portfelių Šarpo rodikliai.	43

Paveikslų sąrašas

1 pav. Biologinis neuronas	12
2 pav. Dirbtinio neuroninio tinklo elementas.....	12
3 pav. Tiesioginio sklidimo neuroninis tinklas.....	13
4 pav. Dirbtinio neuroninio tinklo architektūra.....	13
5 pav. DNT klasifikavimas	14
6 pav. Save tvarkančio neuroninio tinklo scheminė struktūra.....	15
7 pav. Antros dimensijos žemėlapis sudarytas iš trečios dimensijos duomenų rinkinio.....	16
8 pav. Artimi vektoriai iš įėjimų erdvės yra atvaizduojami priešingoje žemėlapio pusėje	17
9 pav. Topografinio žemėlapio, kurį sukelia i-tosios įvesties duomenų komponentas, spalvinimas...17	
10 pav. 50 investicinių fondų, kurie investuoja į pasaulines akcijas.....	21
11 pav. Investicinių fondų analizei, sudaryta SOM žemėlapio principu.....	21
12 pav. Šalies kredito rizikos žemėlapis, grįstas WSJ duomenimis elementas.....	23
13 pav. Įprastas „Viscovery SOMine“ programinės įrangos langas.....	24
14 pav. „Viscovery SOMine“ optinė ženklų atpažinimo funkcija	25
15 pav. „Deep Learning Toolbox“ programos langas.....	26
16 pav. „VisiSom“ programinė įranga.....	27
17 pav. „Neural Data Analysis“ langas.....	27
18 pav. Finviz.com portale nustatyti reikiami rodikliai.....	32
19 pav. Pasirinktas duomenų filtras.....	32
20 pav. Eksportuojamų duomenų iš finviz.com portalo į MS excel failą pavyzdys.....	32
21 pav. „Viscovery SOMine“ duomenų apdorojimo schema.....	34
22 pav. Antrajame duomenų apdorojimo žingsnyje aprašomos duomenų charakteristikos, bei vardinės vertės.....	34
23 pav. „Viscovery SOMine“ klasterio žemėlapio kūrimo schema	35
24 pav. „Viscovery SOMine“ programoje pasirenkamas nustatomi žemėlapio charakteristikos.....	35
25 pav. Pavyzdys: „Viscovery SOMine“ technologijų sektoriaus žemėlapis.....	36
26 pav. Rodiklių išreikštumo žemėlapiai	36
27 pav. Pavyzdys: Įmonių plataus vartojimo prekių C2 klasterio portfelis. Raudonai apibrėžta įmonių suvedimo laukelis (angl. <i>Tickers</i>) ir įmonių savaitinis pelningumas (angl. <i>Week performance</i>).....	38
28 pav. MS Excel faile išsaugotų duomenų apie klasterių vidutinį pelningumą pavyzdys.....	38
29 pav. Atitinkamai pavaizduoti plataus vartojimo prekių, pramoninių prekių, paslaugų, sveikatos ir technologijų sektorių žemėlapiai, ir jų klasteriai.....	39
30 pav. Pasirinktas klasterizavimo algoritmas.....	40
31 pav. Penkių pelningiausių įmonių atrankos procesas	40
32 pav. MS Excel skaičiuoklėje apskaičiuojamas savaitinis viso portfelio pelnas.....	41
33 pav. finance.yahoo.com platformoje gaunami istorinės vertybinių popierių kainos savaitiniam pelnui apskaičiuoti.....	41
34 pav. Akcijų pelningumo tikrinimo grafikas (2018-2019 m.).....	42
35 pav. Pagal pasirinktą klasterizavimo algoritmą suformuotų portfelių vidutinio pelningumo pokyčio grafikas.....	43
36 pav. Pagal pasirinktą klasterizavimo algoritmą sudarytų portfelių suminiai pelnai.....	43
37 pav. Vaizduojamas bendras rinkoje esančių sektorių pelningumas (2018.12 – 2019.05).....	44

Santrumpų sąrašas

Santrumpos:

CR - bendrasis likvidumo rodiklis

DNT – dirbtinis neuroninis tinklas

DI – dirbtinis intelektas

NT – Neuroninis tinklas

SOM – savitvarkis neuroninis tinklas

EPS – Vienai akcijai tenkančio pelno augimas

P/E (PE) – Kainos – pelno santykis

ROE – Grynojo pelno ir nuosavo (akcinio) kapitalo santykis

Įvadas

Darbo aktualumas:

Dirbtinio intelekto metodai vis plačiau taikomi įvairiuose paslaugų ir technologinių procesų srityse. Šiame darbe analizuojamas dirbtinio intelekto metodo, konkrečiai savitvarkių neuroninių tinklų taikymas vertybinių popierių analizei.

Investuotojai nori, kuo geriau priimti sprendimus, į kurias įmones investuoti. Yra du pagrindiniai būdai kaip tai atlikti: fundamentinė analizė ir techninė analizė. Pirmoji remiasi investuotojo sukauptomis žiniomis apie įmones, tačiau joje galima neišvengiama klaidų dėl žmogiškojo faktoriaus. Tuo tarpu techninės analizės specialistai, pasitelkdami duomenis apie finansų rinkų kitimą ir atsižvelgdami į istorinius duomenis bando prognozuoti, į kurias įmones verta investuoti ateityje. Stengiamasi atrasti, kuo efektyvesnius būdus techninei analizei atlikti. Tačiau tai nėra lengva užduotis, nes finansų rinkos pastoviai kinta. Taigi iškyla problema apdoroti ir išgauti informaciją iš finansų rinkose sukauptų didžiųjų duomenų. Paprasti statistiniai metodai to negeba padaryti, todėl tam yra taikomi dirbtinio intelekto metodai. Toki kaip genetiniai algoritmai ar dirbtinių neuroninių tinklų algoritmai. Atlikti tyrimai rodo, jog savitvarkiai neuroniniai tinklai yra potencialūs naudojimui didelėms duomenų bazėms klasterizuoti ir atvaizduoti. Jie geba žmogaus protu sunkiai suvokiamus daugialypius duomenis susisteminti pagal panašumą ir atvaizduoti taip, kad žmogui tai būtų suprantama.

Šiame darbe aprašomos savitvarkių neuroninių tinklų taikymo galimybės: finansuose, versle, finansų rinkose. Pristatomas pasirinktas įmonių iš investavimo platformos atrankos algoritmas.

Darbo tikslas: *Išanalizuoti savitvarkių neuroninių tinklų taikymo galimybes vertybinių popierių analizei. Analizės rezultatus pateikti, kaip rekomendacijas investuotojams.*

Šiam tikslui pasiekti, keliami šie uždaviniai:

1. atlikti literatūros analizę apie savitvarkių neuroninių tinklų taikymą finansų rinkose.
2. atrinkti svarbiausius akcijų vertinimo fundamentinius rodiklius ir kriterijus. Atlikti įmonių atranką pagal juos.
3. pasirinkti klasterizavimo algoritmą ir atlikti įmonių atranką.
4. pasirinkti tyrimo vykdymo laiką, būtiną surinkti duomenų apimtį ir sistemos vertinimo kriterijus.
5. atlikti eksperimentinį pasiūlyto atrankos metodo tyrimą.

1. Dirbtinio intelekto metodai

1.1.1. Dirbtinio intelekto taikymas finansuose

Pasak Schindler'io [1] dirbtinis intelektas (DI) ir mašininis mokymasis yra sparčiai pradedami taikyti įvairioms finansinių paslaugų sektorių reikmėms. Šių taikymų įsisavinimą lėmė du tiekimo veiksniai, toki kaip technologijų pažanga ir finansinių sektorių duomenų pasiekiamumas ir infrastruktūra, ir paklausos veiksniai, toki kaip pelningumo poreikiai, konkurencija su kitomis firmomis, ir finansinio reguliavimo reikalavimai. Toliau pateikiami, kai kurie esami ir potencialūs dirbtinio intelekto ir mašininio mokymosi taikymai finansuose:

- Finansinės institucijos ir pardavėjai naudoja DI ir mašininį mokymąsi įvertinti kredito kokybę, nustatyti draudimo sutarčių kainą ir prekiauti jomis, ir automatizuoti klientų sąveiką.
- Institucijos optimizuoja ribotą kapitalą su DI ir mašinių mokymosi metodais, taip pat grįžtamojo patikrinimo modelius (angl. *back-testing models*) ir analizuoja didelių pozicijų pardavimo, įtaką rinkai.
- Rizikos draudimo fondai, brokeriai ir prekybininkai, ir kitos firmos naudoja DI surasti signalus aukštesnėms grąžoms ir optimizuoti praktybos vykdymą.
- Tiek viešojo, tiek privataus sektoriaus institucijos gali naudotis šiomis technologijomis norminių reikalavimų laikymosi, priežiūros, duomenų kokybės vertinimo ir sukčiavimo nustatymo srityse.[1]

1.1.2. Dirbtinio intelekto rūšys

Dirbtinis intelektas (DI) susideda iš duomenimis varomų metodų, tokių kaip dirbtiniai neuroniniai tinklai, mašininis mokymasis, genetiniai algoritmai, neaiškioji (angl. *fuzzy*) logika [2]. Šiame darbe pagrindė atkreipsime dėmesį į mašininį mokymąsi ir dirbtinius neuroninius tinklus (DNT).

1.1.3. Mašininis mokymasis

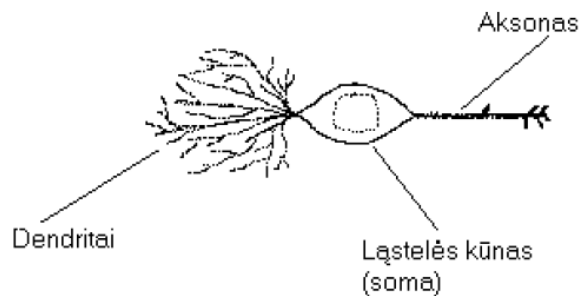
Anot Schindler'io yra kelios mašininio mokymosi (angl. *machine learning*) algoritmų kategorijos. Šios kategorijos skiriasi priklausomai nuo žmogaus įsikišimo lygio, reikalingo ženklinant duomenis:

- Prižiūrimame mokymesi (angl. *supervised learning*) algoritmas yra pamaitinamas mokymo duomenų rinkiniu, kuris turi žymėjimus tam tikrai daliai pastebėjimų (žmogaus atliktų). Pavyzdžiui, duomenų rinkinyje gali būti žymenys tam tikruose duomenų taškuose, kuriuose nurodomi tie, kurie yra apgaulingi, ir tie, kurie nėra apgaulingi. Tada algoritmas apsimokęs „sužinos“ bendrą klasifikavimo taisyklę, kurią jis naudos atspėti žymenis tolimesnėms prognozėms duomenų rinkinyje.
- Neprižiūrimas mokymasis (angl. *unsupervised learning*), tai situacijos, kuriose pateikti duomenys į algoritmą neturi žymėjimų. Algoritmo yra paprašoma duomenyse aptikti šablonus (angl. *patterns*) nustatant stebėjimų klasterius, kurie priklauso nuo panašių pagrindinių savybių. Pavyzdžiui, neprižiūrimi mašininio mokymosi algoritmai gali būti nustatomi ieškoti vertybinių popierių, kurie turi savybes (charakteristikas, parametrus) panašias į nelikvidų vertybinį popierių, kurio kainą sunku nustatyti. Jei algoritmas randa tinkamą klasterį nelikvidžiam vertybiniam popieriui, kitų vertybinių popierių kainos nustatymas tame klasteryje gali būti panaudotas padėti įkainoti nelikvidų vertybinį popierių. Šiame darbe tyrimo dalis buvo atlikta neprižiūrimo mokymosi algoritmais grįsta programine įranga.

- Skatinamasis mokymasis (angl. *reinforcement learning*) patenka tarp prižiūrimo ir neprižiūrimo mokymosi. Šiuo atveju, algoritmas yra pamaitinamas nežymėtais duomenimis, pasirenka veiksmą kiekvienam duomenų taškui, ir gauna grįžtamąjį ryšį (tikriausiai iš žmogaus), kuris padeda algoritmui apsimokyti. Pavyzdžiui, skatinamasis mokymasis gali būti naudojamas robotikoje ir save valdančiose mašinos.
- „Gilusis mokymasis“ (angl. *deep learning*) yra mašininio mokymosi forma, kuri naudoja algoritmus, veikiančius „sluoksniais“, grįstais smegenų struktūra ir jų funkcija. Giliojo mokymosi algoritmai, kurių struktūra vadinama neuroniniais tinklais, gali būti naudojami prižiūrimame, neprižiūrimame ar skatinamame mokymosi.[1]

1.1.4. Dirbtinio neurono modelis

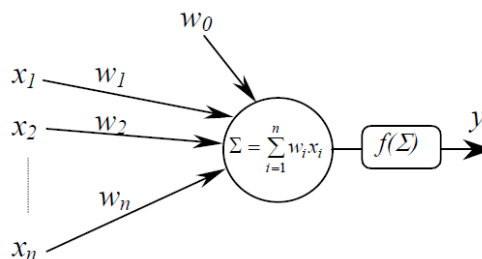
Pasak Myers'o [3] biologinį neuroną 1 pav. sudaro ląstelės kūnas (soma) ir joje esantis branduolys, kuriame yra informacija apie neurono slenkstinę įtampą, iš kitų neuronų informaciją priimančios ataugos - dendritai ir informacijos išėjimo struktūra – aksonas. Vienos ląstelės aksonas su kitos ląstelės dendritu jungiasi per sinapses. Sinapsėmis vadinamos aksono jungtims su kitų neuronų dendritais.



1 pav. Biologinis neuronas [4]

Verikas ir Geležinis [5] siekdami išlaikyti panašumą į biologines neuronines sistemas, dirbtinio neuroninio tinklo neuroną apibūdino taip:

- Į neuroną įprastai ateina keletas įėjimo reikšmių. Šios įėjimo reikšmės gali būti bendros visam neuroniniam tinklui arba ateiti iš kitų neuronų. Neuronai sujungti ryšiais, kurie turi savo koeficientus (svorius). Neuronai turi savo slenkstines reikšmes. Neuronas sužadina, tada, kai pasiekia sužadavimo reikšmę lygią svoriniai signalų sumai atėmus slenkščio reikšmę.
- Neurono išėjimo reikšmė skaičiuojama pagal jo sužadavimo signalą, taikant neurono perdavimo funkciją.



2 pav. Dirbtinio neuroninio tinklo elementas [5]

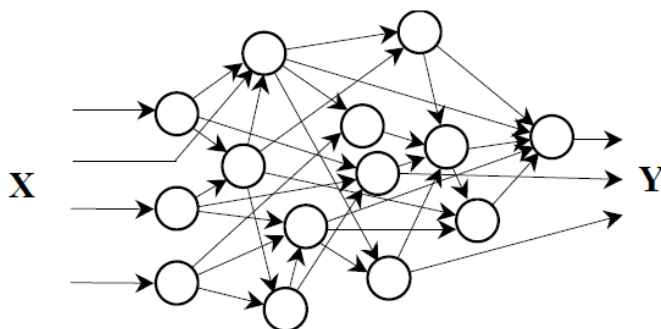
- 2 paveikslėlyje neurono įėjimai pažymėti $x_1 \dots x_n$; atitinkami svoriai pažymėti $w_1 \dots w_n$; w_0 – slenkščio reikšmė; $f()$ – perdavimo funkcija; y – neurono išėjimas [5].

1.1.5. Dirbtinis neuroninis tinklas

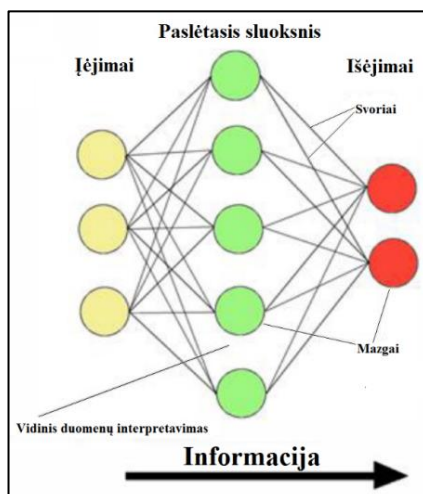
Dirbtinis neuroninis tinklas (DNT) yra rinkinys tarpusavyje paprastai lygiagrečiai sujungtų informaciją apdorojančių dirbtinių neuronų. Šie imituojantys neuronai yra dirbtinio neuroninio tinklo informacijos apdorojimo elementai. Tinklo apdorojimo geba yra laikoma ryšius sudarančiuose svoriuose, kurie yra gaunami mokymosi proceso metu iš mokymosi šablonų (angl. *patterns*) rinkinio [2].

Dažniausiai neuroniniame tinkle signalai sklinda tiesiogiai: informacijai atėjus įėjimus sklinda toliau per elementus esančius paslėptuose sluoksniuose ir galiausiai pasiekia išėjimo neuronus. Tokia dirbtinio neuroninio tinklo struktūra laikoma stabilia. Kita vertus, jei dirbtinis neuroninis tinklas yra su grįžtamais ryšiais ir vadinamas rekurentiniu, toks tinklas gali būti nestabilus, nes dažnai turi sudėtingą dinamiką. Praktikoje dažniausiai yra taikomi tiesioginio sklidimo neuroniniai tinklai, nors pasitaiko atvejų, kada taikomi ir rekurentiniai [5].

Taip pat galimi neuroniniai tinklai, kurie yra tiesioginio sklidimo ir neturi sluoksninės struktūros. Tokio tinklo pavyzdys parodytas 3 paveiksle [5]. Tačiau neuroniniai tinklai paprastai turi sluoksninę struktūrą. Įėjimo, paslėptąjį ir išėjimo sluoksnius 4 pav. [6]



3 pav. Tiesioginio sklidimo neuroninis tinklas [5]

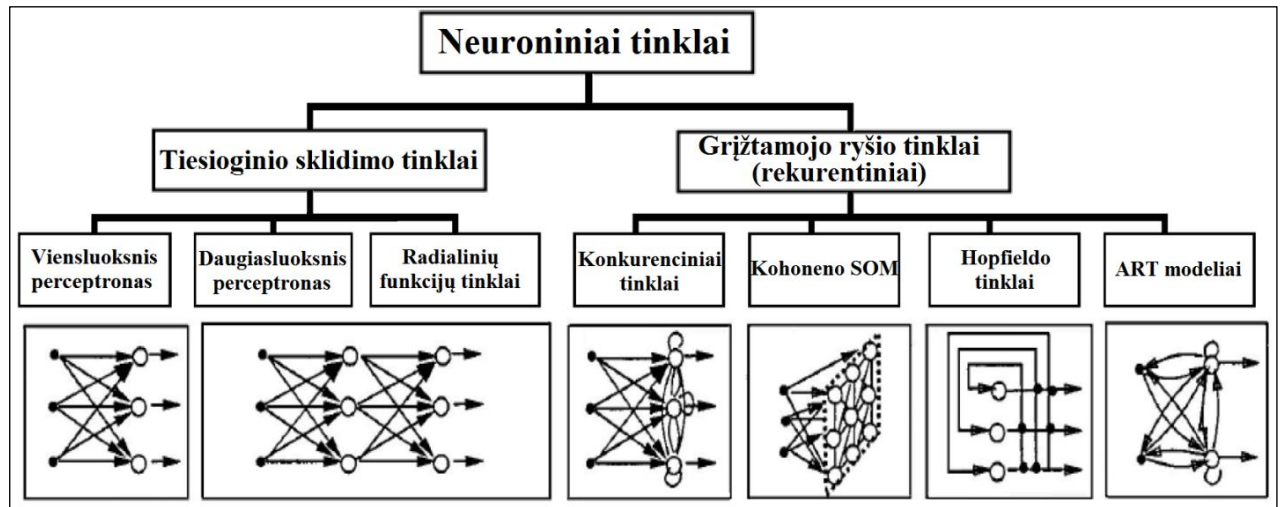


4 pav. Dirbtinio neuroninio tinklo architektūra [6]

Pasak Andriuškos [6] neuroniniai tinklai klasifikuojami pagal keturis pagrindinius aspektus 5pav.:

- „pagal topologiją – viensluoksniai, daugiasluoksniai, atsinaujinantys ir savitvarkiai neuroniniai tinklai;

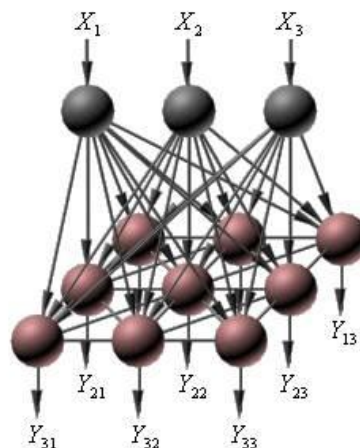
- pagal ryšių pobūdį – dinaminiai ir statiniai;
- pagal signalų sklaidimo kryptį – tiesiniai ir rekurentiniai;
- pagal taikymą – klasifikavimui, prognozavimui, grupavimui ir funkcijų apibendrinimui. [6]



5 pav. DNT klasifikavimas [6]

1.2. Savitvarkiai neuroniniai tinklai

SOM yra tokio tipo neuroninis tinklas, kuriam apsimokyti nereikia priežiūros (Kohene'as, 1995). Pristaikydamas prie treniravimosi duomenų SOM suformuoja savo išėjimus pats. Paprasčiausias SOM susideda iš M neuronų, esančių ant įprastos žemos dimensijos gardelės, įprastai 1-os ar 2-jų dimensijų. Aukštesnių dimensijų gardelės taip pat yra įmanomos, bet jos dažniausiai nenaudojamos, nes jų vizualizavimas yra problemiškas. Gardelės dažniausiai yra heksagonalinės arba stačiakampės [7].



6 pav. Save tvarkančio neuroninio tinklo scheminė struktūra [7].

6 pav. pavaizduoti pilki kamuoliukai formuoja įėjimo sluoksnį, rausvi išėjimo sluoksnį. Rodyklės jungiančios įėjimo ir išėjimo sluoksnius yra ryšiai su savo svoriais.

Yra žinoma, kad SOM vykdo dviejų tipų duomenų kompresijas:

- duomenų dimensijos redukavimą su minimaliu informacijos praradimu

- duomenų įvairovės redukciją, dėl terminalo sudėties prototipų atskyrimo (Duomenų grupių klasterizavimas ir kvantavimas)

Kol prižiūrimas neuroninis tinklas, mokosi modeliuoti ryšius tarp įėjimų ir žinomų išėjimų, neprižiūrimas neuroninis tinklas (tarp jų ir SOM) mokosi klasterizuoti ar grupuoti duomenų šablonus (angl. *patterns*) tik patikrindamas panašumą tarp įėjimų. Klasterizavimą gali sudaryti duomenų klasifikavimas arba segmentavimas į grupes, priklausomai nuo natūralios duomenų struktūros, ir visai nežinant (netaikant) iš anksto nustatytų klasifikacijų. Atitinkamo algoritmo rezultatas yra šablonų panašumo laipsnio maksimizavimas tame pačiame klasteryje ir panašumo minimizavimas tarp skirtingų klasterių [8].

Klasterizavimo duomenys tarnauja keliems tikslams susijusiems su duomenų analizavimu ir modeliavimu. Pirmasis tikslas - klasterizavimas leidžia mums patikrinti didelius duomenų rinkinius ir iškart rasti duomenų šablonus (angl. *patterns*), kurie atrodo labai skiriasi nuo likusio duomenų rinkinio. Šablonai paprastai yra toki skirtingi nuo likusių duomenų, todėl jie yra savo atskirame klasteryje. Todėl klasterizavimas yra naudingas būdas iš anksto apdorojant duomenis, tam, kad išmesti nuokrypius ir pataisyti duomenų suvedimo klaidas, kurios gali turėti žalingą poveikį paskesniai modeliavimui. Antra, klasterizavimas leidžia sukurti natūralias grupavimo struktūras, kurios suteikia mums alternatyvų vaizdą į duomenis. Stebint ir modeliuojant kiekvienos grupės, kurios skiriasi duomenų savybėmis, elgesį gali būti labiau įžvalgus būdas, nei stebint iš anksto nustatytų grupių elgesį. Trečia, kai natūrali grupių struktūra susidarė, mes galime tai naudoti kaip būsimą duomenų prognozavimo ir klasifikavimo priemonę [8]

Smith'o [8] teigimu save organizuojantys žemėlapiai (angl. *self-organizing maps*) yra geriausiai žinomas neprižiūrimas neuroninių tinklų klasterizavimo metodas. Jo privalumas palyginus su tradiciniais metodais, tokiais kaip k-vidurkių algoritmas (angl. k-means algorithm) (Harigan, 1975) slypi geresnėse vizualizavimo galimybėse, dėl dviejų dimensijų klasterių žemėlapiu. Dažnai šablonai aukštos dimensijos įėjimų erdvėje turi labai komplikotą struktūrą, bet ši struktūra yra padaroma aiškesne ir paprastesne, kai jie yra suklasterizuojami žemesnės dimensijos erdvėje. Kohen 'as (1982, 1988) sukūrė savitvarkių neuroninių tinklų būdą automatiškai atpažinti stiprias savybes didelių duomenų rinkiniuose. Savitvarkiai neuroniniai tinklai sudaro žemėlapi iš aukštos dimensijos įėjimų erdvės į žemos dimensijos savybių erdvę, taip, kad klasteriai, kurie formuoja tampa matomais šioje dimensijoje (sumažinus dimensijų skaičių) [8]

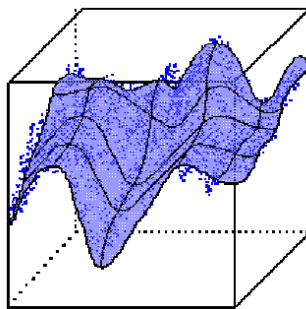
Paprastas SOM algoritmas yra pasikartojantis. Kiekvienas neuronas i turi dimensinį prototipinį vektorių $w_i = [w_{i1} \dots w_{id}]$ arba i -ojo neurono svorį. Kiekviename treniravimosi žingsnyje, imties duomenų vektorius x yra atsitiktinai pasirenkamas iš treniravimo grupės. Atstumai tarp x ir visų prototipinių vektorių yra apskaičiuojami. Geriausiai atitinkantis elementas (BMU) arba laimintis elementas, čia paženkinamas x_{i^*} , yra tinklo elementas su prototipu arčiausiai esančiu x (Kask'is, 1998):

$$|w_{i^*} - x| \leq |w_i - x|, \quad \forall i \neq i^*;$$

Toliau, proto tipiniai vektoriai yra atnaujinami. Laimintis elementas ir jo topo loginiai kaimynai yra nukeliami arčiau įėjimo vektoriaus, į įėjimų erdvę, pagal taisyklę $\Delta w_{i^*}^r = \eta(x^r - w_{i^*}^r)$, kur η yra apsimokymo sparta, ir $\Delta w_{i^*}^r$ yra i -ojo neurono svorio modifikacija. Galiausiai, atnaujinimo taisyklė visiems vektorių elementams i yra:

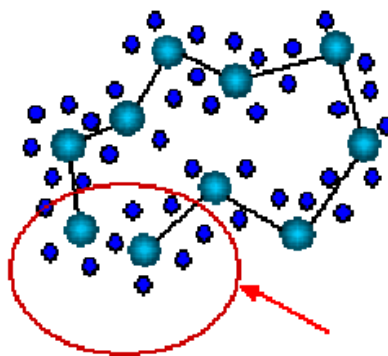
$$\Delta w_i^r = \eta \Lambda(|i - i^*|) (x^r - w_i);$$

Čia $\Lambda(|i - i^*|)$ yra branduolys sutelktas į laimėjusįjį elementą. Branduolys gali būti, pavyzdžiui, Gausinis: $\Lambda(\mathbf{a}) = \exp(-\mathbf{a}^2/\sigma^2)$, kur σ yra kaimyninis spindulys. Abu dydžiai tiek mokymosi sparta η tiek kaimynystėje esantis spindulys σ laikui bėgant monotoniškai mažėja. Per apsimokymą, SOM elgiasi tarsi lankstus tinklas, kuris susilanksto į “debesį” 7 pav., suformuotą apsimokymo duomenų. Dėl kaimynystėje esančių ryšių, kaimyniniai prototipai yra patraukiami į tą pačią pusę, ir tų prototipų vektoriai primena vienas kitą. Neuronų skaičius esantis išėjimo sluoksnyje apibrėžia modelio vektorių maksimalų skirtumą. Apmokytas SOM gali klasifikuoti savo įėjimus: Laimintis elementas (angl. *BMU*) apibrėžia įėjimo vektorių klasę.

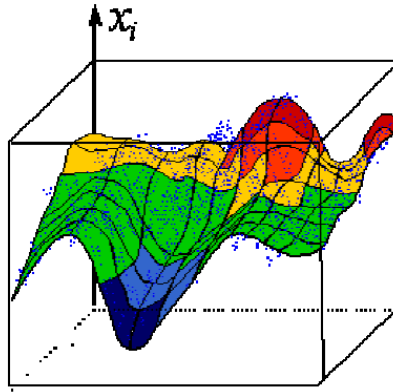


7 pav. . Antros dimensijos žemėlapis sudarytas iš trečios dimensijos duomenų rinkinio [7]

SOM suformuoja žemos dimensijos treniravimo rinkinio tinklą. Sutvarkyta SOM gardelė gali būti kaip patogi vizualizavimo platforma parodanti SOM ypatybes (ir tuo pačiu duomenų). Vizualizavimo tikslas yra pristatyti didelį kiekį detalios informacijos, tam, kad tyrėjui būtų suteikta kokybiška informacija apie duomenų ypatybes. Įprastai ypatybių, kurios turi būti vizualizuojamos yra daugiau nei naudojamų vizualinių dimensijų. Tiesiog jų nėra visų įmanoma parodyti viename paveikslėlyje. Kiekvienas vektorius iš daugialypės įėjimų erdvės turi savo koordinatę gardelėje. Kuo arčiau yra dviejų vektorių koordinatės, tuo šie vektoriai yra arčiau vienas kito įėjimų erdvėje. Bet atvirkščias teiginys nėra teisingas.



8 pav. Artimi vektoriai iš įėjimų erdvės yra atvaizduojami priešingoje žemėlapyje [7].



9 pav. Topografinio žemėlapiu, kurį sukelia i -tosios įvesties duomenų komponentas, spalvinimas. [7].

Kai įėjimo duomenys yra pateikiami SOM tinklui, elementai esantys išėjimo sluoksnyje konkuruoja vienas su kitu, tam, kad vienas jų būtų priimtas laimėtoju. Laimintis išėjimo elementas bus elementas, kurio įeinantieji ryšių svoriai yra artimiausi įėjimo duomenų šablono. Eukalidinio atstumo atžvilgiu. Taigi įėjimas yra pateikiamas ir kiekvienas išėjimo vienetas varžosi, tam kad atitiktų įėjimo struktūrą (angl. *pattern*). Išėjimas, kuris yra arčiausiai įėjimo struktūros (šablono) yra paskelbiamas laimėtoju. Laimėjusiojo elemento ryšių svoriai tada yra pakoreguojami. Tai yra paslenkami įėjimo struktūros link faktoriumi apibrėžtu mokymosi sparta. Tai yra pagrindinė besivaržančių neuroninių tinklų esmė. SOM sukuria topo loginį žemėlapi 9 pav. ne tik pakoreguodamas laiminčio elemento svorius, bet taip pat pakoreguodamas šalia esančių išėjimo elementų, kurie yra kaimynystėje laimėjusio elemento svorius. Taigi ne tik laimintis elementas yra pakoreguojamas, bet ir visi elementai esantys kaimynystėje yra paslenkami arčiau įėjimo šablono. Dažnai pradedant nuo atsitiktinų svorio reikšmių, išėjimo elementai lėtai susilygiuoja (susiderina) taip, kad, kai įėjimo struktūra yra pateikiama, kaimynystėje esantys elementai reaguoja į įėjimo struktūrą.

Kai treniravimas progresuoja, kaimynystės (angl. *neighborhood*) dydis aplinkui laimintįjį elementą sumažėja. Visų pirma dideli išėjimo elementų skaičiai bus atnaujinti, bet kai treniravimas tęsiasi tai vis mažesnis ir mažesnis kiekis bus atnaujintas iki kol pabaigoje treniravimosi tik laimintis elementas bus pakoreguotas. Panašiai ir apsimokymo sparta sumažės, kai treniravimas progresuos.

Rezultatas yra ryšių svoriai, kurie pateikia tipišką arba proto tipinę įėjimų struktūrą, pogrupių įėjimų, kurie patenka į konkretų klasterį. Procesas, kurio metu yra paimama aukštesnės dimensijos duomenys ir sumažinami iki keleto klasterių yra vadinamas segmentavimu. Aukštos dimensijos įėjimo erdvė yra sumažinama į dviejų dimensijų žemėlapi. Jei naudojamas laiminčiojo išėjimo indeksas, jis būtinai padalina įėjimo struktūras į keletas kategorijų ar klasterių. SOM taip pat turi gebą generalizuoti. Tai reiškia, kad tinklas gali atpažinti ar charakterizuoti įėjimus, su kuriais niekada anksčiau nėra susidūręs. Naujas įėjimas yra įsisavinamas su žemėlapiu sritimi, kuriame jis pavaizduojamas. Toliau, įėjimo vektoriai net ir su prarastais duomenimis gal būti naudojami ieškant ar prognozuoti prarastų duomenų vertes, remiantis apmokytu žemėlapiu [7].

1.2.1. Savitvarkių tinklų algoritmas

Bendru atveju savitvarkių neuroninių tinklų algoritmas aiškinamas taip: pradinėje stadijoje nustatomi svoriai mažomis atsitiktinėmis reikšmėmis ir nustatomas kaimynystės pradinis dydis $N_m(0)$, toks, kad būtų pakankamai didelis (tačiau mažesnis nei neuronų skaičius mažiausioje masyvo dimensijoje),

ir nustatomos parametrų funkcijų $\alpha(t)$ ir $\sigma^2(t)$ vertės tarp 0 ir 1. Tuomet atliekami algoritmo žingsniai, kaip tai pateikta žemiau:

Etapas 1: Atsitiktine tvarka parinkamas šablonas (angl. *pattern*) \mathbf{x} tam, kad jis būtų pateiktas savitvarkiams neuroniniams tinklams per įėjimo sluoksnį.

Etapas 2: Apskaičiuojamas panašumas (atstumas) tarp šio įėjimo ir svorių tarp kiekvieno neurono j :

$$d_j = \|\mathbf{x} - \mathbf{w}_j\| = \sqrt{\sum_{i=1}^n (x_i - w_{ji})^2},$$

Etapas 3: Parenkamas neuronas su minimaliu atstumu, kaip laimėtojas m .

Etapas 4: Atnaujinami svoriai jungiantys įėjimo sluoksnį su laiminčiuoju neuronu ir jo kaimynystėje esančiais neuronais, pagal apsimokymo taisyklę:

$$w_{ji}(t+1) = w_{ji}(t) + c[x_i - w_{ji}(t)],$$

kur $c = \alpha(t) \exp(-\|r_i - r_m\|/\sigma^2(t))$ visiems j neuronams $N_m(t)$.

Etapas 5: Algoritmas tęsiamas nuo pirmojo etapo Ω epochų; tada mažinamas kaimynystės neuronų skaičius, $\alpha(t)$ ir $\sigma^2(t)$: Kartojama iki tol kol svoriai stabilizuojasi. [8]

1.3. Savitvarkių neuroninių tinklų taikymas finansų rinkose

Šiame poskyryje bus apžvelgti, pagrindiniai tyrimų rezultatai, kaip savitvarkiai neuroniniai tinklai buvo naudojami finansų rinkose. Zherebtsov'as ir Kuperin'is savo darbe [7] taikė savitvarkių neuroninių tinklų (SOM) metodą klasterizuoti DJIA ir NASDAQ100 portfelius tam, kad būtų nustatytos netiesinės koreliacijos tarp akcijų. Autorių straipsnyje buvo pateikiamas SOM metodo taikymas, kaip alternatyva ultrametrinų erdvių metodui. Zherebtsov'as ir Kuperin'is atrado, jog savitvarkių neuroninių tinklų (SOM) metodas yra labiau patikimas ir perspektyvus klasterizuoti blogai sustruktūrizuoti didelei duomenų bazei ir, ypač, NASDAQ 100 akcijų rinkos indeksui, kur netiesinis didelių duomenų apdorojimas yra reikalingas. [7]

Korzeniewsk'is [9] nagrinėjo ir savitvarkiais neuroniniais tinklais sprendė portfelio diversifikavimo problemą. Jis rašė, kad kai investuotojai naudoja statistinius metodus optimizuoti savo vertybinių popierių investavimo sprendimus, viena iš fundamentalių problemų būna sukonstruoti gerai diversifikuotą portfelį, susidedantį iš saikingo (vidutiniško) pozicijų skaičiaus. Tarp daugybės metodų taikomų užduočiai, yra grupė (metodų) kurie remiasi dalinant kompanijas į keletą vienodų (homogenišku) grupių. Vėliau iš kiekvienos šių grupių pasirenkamas vienas atstovas, tam, kad būtų sukurtas galutinis portfelis. Dalinimo žingsnis neturi sutapti su kompanijų sektorių pasirinkimu. Kai dalijimas yra atliekamas kompanijų klasterizavimo priemonėmis, esminė proceso dalis yra parinkti gerą klasterių skaičių. Korzeniewsk'is savo straipsnyje [9] pristatė tuo metu naują portfelio konstravimo metodą grįstą nustatant daugybę portfelio pozicijų ir taip pat pasirenkant klasterių atstovus. Grupavimo metodai naudojami klasterizavimo procese yra klasikinis k-būdas ir PAM (angl. *Partitioning Around Medoids*) algoritmas. Metodas yra testuojamas su duomenimis su 85 dažiausiomis kompanijomis iš Varšuvos fondų biržos 2011-2016 metų laikotarpyje. Rezultatai yra patenkinami atsižvelgiant į bendras galimybes kurti klasteriu grįstą algoritmą, kuris nereikalauja beveik jokio investuotojo įsikišimo.

Enklund'o straipsnyje [10] yra iliustruojamas savitvarkių neuroninių tinklų metodas įmonių finansinei veiklai ir lyginamajai analizei. Šio tyrimo autoriai sukūrė finansinių rodiklių duomenų bazę, kuri parodo 91 tarptautinių medienos ir popieriaus kompanijų periodu nuo 1995 iki 2001

finansinius rezultatus. Tyrime buvo naudojamas savitvarkių neuroninių tinklų metodas analizuoti ir atlikti lyginamąją analizę didžiausių penkių viso pasaulio medienos ir popieriaus kompanijų. Tyrimo rezultatai rodo, jog naudojant savitvarkius neuroninius tinklus (SOM), galime reikšmingai struktūrizuoti, analizuoti ir vizualizuoti didelius kiekius daugialypių finansinių duomenų. [10]

Norint padėti investuotojams valdyti savo portfelius, Joseph'as ir Indratmo'as [11] sukūrė įrankį akcijos biržos duomenų klasterizavimui ir vizualizavimui, naudojantį SOM algoritmą. Jų įrankis skirtas padėti naudotojams atpažinti akcijų grupes, kurios turi panašių kainų kitimo dėsnius per tam tikrą laiką. Tyrėjai atliko vizualinę analizę lygindami gautą rezultato vizualizaciją su Yahoo Finansavimo diagramomis. Apskritai, šio tyrimo autoriai nustatė, jog SOM algoritmas gali analizuoti ir klasterizuoti akcijų biržas pakankamai racionaliai [11].

Eklund'as ir Vanharant'as [12] tyrė, kaip SOM gali padėti įmonėms atlikti konkurentų lyginamąją finansinę analizę. Jie rašė, jog šiomis dienomis didėjant konkurencijai versle, priverčia stebėti panašaus tipo įmones. Ypač konkurentų analizė tampa svarbi toms įmonėms, kurios verčiasi globaliose rinkose. Dėl to didesnės įmonės turi atskirus verslo intelekto skyrius, kurie reguliariai atlieka konkurentų ir rinkos analizę. Lyginamoji analizė lygina kompanijų kiekybinius ar kokybinius rodiklius, tam kad atrastų, kaip įmonės efektyvumas gali būti padidinamas. Finansinė konkurentų analizė, kuri lygina finansinius kompanijų rodiklius, šiomis dienomis yra dažniausiai naudojama. Kompanijos lyginamos iš anksto apibrėžtus finansinius rodiklius, tokius kaip veiklos pelningumas (angl. *operating margin*), nuosavo kapitalo pelningumo rodiklis (angl. *return on equity*) ir kt. Tačiau nėra lengva įvertinti, kompanijas, kai jų daug ir kai, reikia iširti jų visų finansinius rodiklius. Paprastos skaičiuoklės ir grafikai čia nepadeda, nes problema tampa per sudėtinga. Šiai problemai spręsti ir duomenims atvaizduoti geriausiai tinka SOM įrankis. Šiame tyrime, lyginamoji analizė pramoninių prekių industrijoje buvo atlikta savitvarkais neuroniniais tinklais. Metodas vėliau buvo įvertintas šios industrijos ekspertų. Apklausa parodė, jog įmonių vadovai SOM metodą vertino, kaip daug geresnį, nei daugumą savo kitų metodų, ypač dėl jo teikiamų vizualinių galimybių. Pastebėta, jog šis metodas, buvo naudingas, kada įmonių vadovams reikėdavo priimti strateginius sprendimus. [12]

Plačiau panagrinėsime Deboeck 'o straipsnį [13], nes jame naudojama programinė įranga yra ta pati (Viscovery SOMine), kaip šio magistrinio darbo tyrimo dalyje. Pirmas šio tyrimo pavyzdys iliustruoja SOM naudojimą investiciniams fondams. Straipsnyje rašoma, jog pasirinkdami investicinius fondus investuotojai paprastai atsižvelgia į daugybę kriterijų: kiek pelno gauta praityje, į riziką (kintamumą), pelno gavimo riziką, fondo dydį, portfelio apyvartos rodiklį, išlaidas ir kitus panašius finansinius rodiklius.

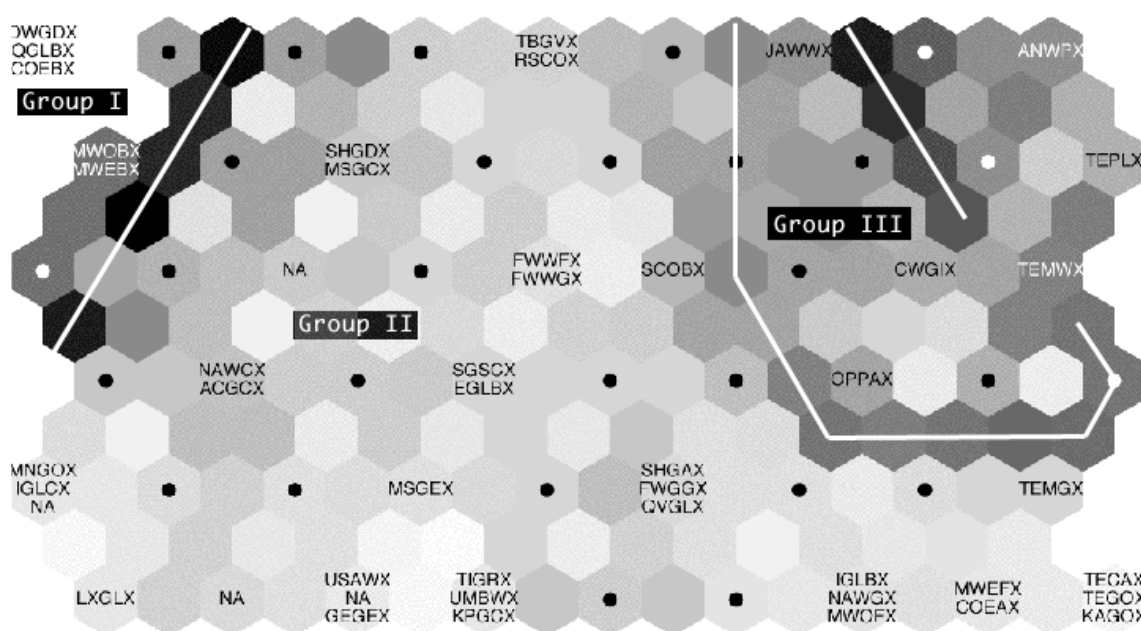
Dauguma skaičiuoklių (tokios, kaip MS Excel) leidžia klasifikuoti duomenų sąrašus tik atsižvelgiant į du ar tris kriterijus; Dauguma programų yra ribotos atvaizduoti tris dimensijas.

Kita vertus, su SOM yra įmanoma įtraukti bet kokį dimensijų skaičių ir gauti sumažintą iki dviejų dimensijų duomenų žemėlapi, be savavališko tyrėjo prielaidų darymo ar kitų metodų, kurie iškraipytų duomenis.

Kad tai iliustruoti tyrėjai naudojo Morningstar investicinių fondų duomenų bazę, informacijos surinkimui apie investicinius fondus. Pirmajame pavyzdyje jie kreipė dėmesį į investicinius fondus, kurie investuojami į pasaulio akcijas. Subkriterijai, kurie buvo naudojami Morningstar sąrašė buvo: (1) įmonės vadovo valdymo įgaliojimai daugiau arba lygu 3 metams, (2) akcininkų susiskirstymas lygus arba didesnis nei B+, (3) Morningstar reitingas lygus ar didesnis nei 4, ir (4) išlaidų rodiklis lygus ar mažesnis nei 1%. Atsižvelgiant į tai įsigyjama 50 investicinių fondų, kurie fokusuojasi į

investicijas pasaulio akcijų rinkose. Pagrindiniai kintamieji parinkti iš Morningstar investicinių fondų duomenų bazės kiekvienam iš šių fondų pasiskirstė į tris klasterių grupes:

- Pirmoji grupė įtraukia tokias akcijas, kaip Colonial Globa Equity B (COEBX), Dean Witter Global Div Growth (DWGDX), MFS World Equity B (MWEBX) ir Oppenheimer Quest Global Value B (QGLBX);
- Trečioji grupė įtraukia (ANWPX), Capital Growth & Inc (CWGIX), Janus Worldwide (JAWWX), Oppenheimer Global A (OPPAX), Templeton World I (TEMWX), ir Templeton Growth I (TEPLX) ir
- Antroji grupė susideda iš visų likusių fondų (fondų simboliai gali būti rasti Morningstar tarptautiniame fondų sąrašė).



10 pav. 50 investicinių fondų, kurie investuoja į pasaulines akcijas.[13]

10 pav. vaizduojama, kaip SOM vizualizuoja skirtumus tarp investicinių fondų, kuriuos Morningstar įvertino 4-iais ar 5-iais. Detalesnė informacija apie skirtumus tarp panašių fondų, gali pagelbėti pasirinkti fondą, taip pat kaip ir turto paskirstymą ar portfelių sudėtį. 11 pav. esanti lentelė rodo kiekvienos grupės originalių įėjimo duomenų vidurkius, kiekvienai iš šių fondų grupių. Šios lentelės, sudarytos SOM metodu pagalba, galima analizuoti įvairius fondų rodiklius.

Group No	Manager	Tot Ret	Annlzd	Annlzd	Decile	Assets	Turnover	Front	Defer	Expense	
Funds	Tenure	12 Mo	3 Yr	5 Yr	Rank	\$MM	Ratio	Load	Load	Ratio	
1	5	2.8	19.0	13.8	14.1	4.0	658.2	80.8	0.0	4.6	2.3
2	39	3.3	20.1	14.1	14.6	6.7	272.4	70.7	2.2	0.1	1.7
3	6	7.2	22.4	16.1	16.0	5.8	6638.3	52.7	4.8	0.0	1.0

11 pav. Investicinių fondų analizei, sudaryta SOM žemėlapiu principu.[13]

Iš lentelės esančios 11 pav. galime pastebėti, kad :

- I grupė turi vadovus su kelių metų valdymo įgaliojimu (vidutiniškai 2.8 metus, priešingai nei III grupės fondų vadovai, kurie vidutiniškai turi 7.2 metų teisę valdyti);
- I grupės fondai sukūrė vidutiniškai 200 bazinių taškų per metus mažiau nei fondai III-ioje grupėje.
- Apyvartos rodiklis I grupėje yra daug didesnis (82proc.) palyginus su III-ios grupės apyvartos rodikliu (52proc.)
- Vidutinis išlaidų rodiklis I-ame fonde yra du kartus didesnis (2.3proc.) nei III fondo vidutinis išlaidų rodiklis 1proc.

Apibendrinant galima pastebėti, jog vadovai su ilgesniu įgaliojimu valdyti ir, kurie valdo daugiau turto, pasiekia mažesnę apyvartą, ir moka mažesnius mokesčius ir taip pat vidutiniškai sukuria 200 bazinius taškų daugiau nei jų kolegos. Reikėtų pabrėžti, jog šie rezultatai tinka tik vadovams, kurie atsakingi už fondus investuojančius į pasaulio akcijas, tačiau ne vadovams iš kitų grupių. [13]

Pasak Marvins'o [14] moderni portfelio teorija teigia, jog diversifikavimas yra didžiausią efektyvumą turinti strategija, kad gautume mažą rizikos/pelno rodiklį (santykį). Jos straipsnio autorės tyrime klasterizavimo technologija yra taikoma tam, kad pastebėti, kurie aktyvai yra skirtingi vienas nuo kito. Vertybiniai popieriai gali būti atskiriami klasterizavimo metodu, kuris maksimizuoja panašumą grupių viduje ir minimizuoja panašumą tarp grupių. Tai atliekant su vertybiniais popieriais galima atrasti aktyvų kombinaciją, kuri sudarytų gerai diversifikuotą portfelį. [14]

Silva'o ir Marques'o straipsnyje [15] dėmesys buvo atkreiptas į laiko eilučių klasterizavimą (angl. *time series clustering*). Tyrime pristatytas metodas taikantis savitvarčius neuroninius tinklus. Jis buvo taikomas istoriniam duomenų rinkiniui aprašančiam 49 vertybinių popierių kainų kitimą. Duomenų rinkinys apėmė informaciją surinktą nuo 1998 iki 2009 metų. Metodas gebėjo suklasterizuoti vertybinių popierių duomenis į klasterius su panašia elgsena, kas rodė, jog SOM gali būti taikomas laiko eilučių klasterizavimui (angl. *time series clustering*) pagal jų bendrą elgseną.

Metodas atsižvelgdamas į istorinis duomenis gebėjo suklasterizuoti draudimo kompanijas į vieną klasterį, ir finansines institucijas į kitus du klasterius. Taip pat metodas parodė, jog tam tikras vertybinis popierius išsiskyrė savo elgsena palyginus su kitais vertybiniais popieriais. Tokio pobūdžio analizė yra aktuali portfelio pasirinkimui, kadangi portfelio sudarymo su per daug panašiais vertybiniais popieriais atsižvelgiant jų istorinius duomenis reikia vengti, kaip priemonės sumažinti rizikos poveikį. Vietoj to, portfelio diversifikacija yra pagrindinis tikslas. [15]

Antrame Deboeck,o [13] straipsnio pavyzdyje dėmesys kreipiamas į rizikas investavime įvairiuose akcijų biržose visame pasaulyje. SOM yra naudojamas analizuoti šias rizikas ir atrasti šalių grupes su panašiais rizikos dėsniais. Iš 52 šalių atrenkamos 5 grupės. Visų šalių rinkos lyginamos su USA šalies rinka, kaip etalonu.

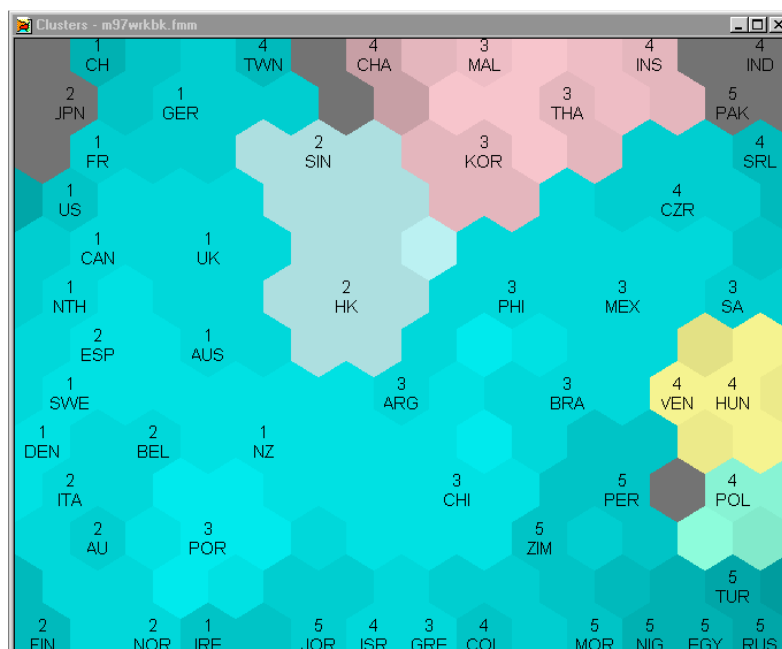
52 šalys yra suklasterizuojamos į penkias grupes: 1) tos, kurios labiausiai panašios į JAV; 2) kitos išsivysčiusios šalys 3) brandžios ir besivystančios rinkos 4) naujai besivystančios rinkos 5) pasienio rinkos. Šalys yra suklasterizuojamos į grupes, kur:

1) Pirmojoje grupėje labiausiai panašios į JAV yra įtrauktos JAV, Australija, Kanada, Danija, Prancūzija, Vokietija, Airija, Nyderlandai, Naujoji Zelandija, Švedija, Šveicarija ir Jungtinė Karalystė.

- 2) Į antrąją grupę įtrauktos kitos išsivysčiusios šalys tokios kaip Austrija, Belgija, Suomija, Hong Kogas, Italija, Japonija, Norvegija, Singapūras ir Ispanija.
- 3) Į trečiąją grupę įtrauktos šalys Argentina, Brazilija, Čilė, Graikija, Korėja, Malaizija, Meksika, Filipinai, Portugalija, Pietų Afrika ir Tailandas.
- 4) Ketvirtoji grupė, naujai atsirandančios rinkos yra šalių, tokių kaip Kinija, Kolumbija, Čekijos Respublika, Vengrija, Indija, Indonezija, Izraelis, Lenkija, Šri Lanka, Taivanas ir Venesuela.
- 5) Penktoji grupė - užsienio rinkos. Šiai grupei priklauso tokios šalys, kaip Egiptas, Jordanas, Marokas, Nigerija, Pakistanas, Peru, Rusija, Turkija ir Zimbabvė.

Šiame pavyzdyje, kaip ir minėta, JAV buvo kaip "šalis etalonas", su kuria lyginamos kitų šalių rinkos. Esamos šalys padalytos į penkias grupes, kurios maždaug tokio pat dydžio. Grupavimo kriterijai nebuvo aiškiai pateikti: tokios šalys, kurios vadinamos užsienio rinkomis (angl. frontier markets) turėjo daug trūkstamos informacijos.

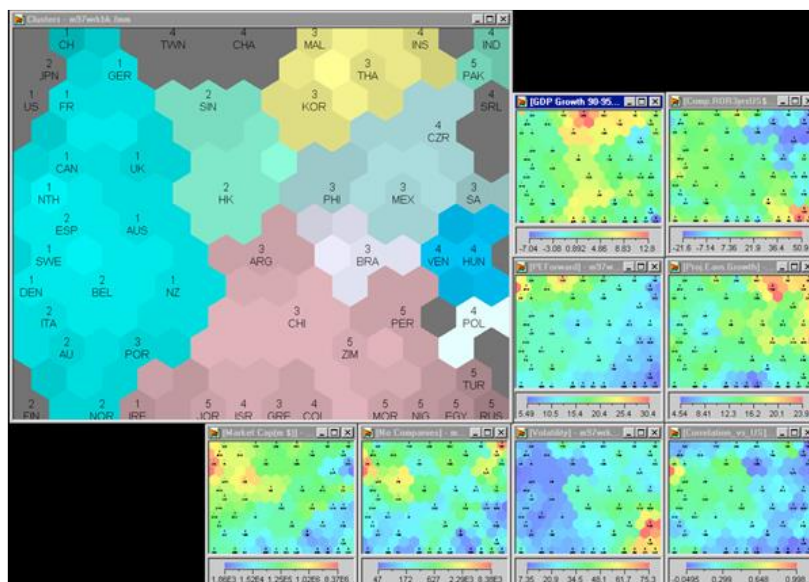
Naudojant tuos pačius duomenis ir atikus analizę interaktyviu ir intuityviu būdu su Viscovery (iš Eudaptics GmbH), kad sukurti SOM buvo gauti gan skirtingi rezultatai. 12 paveikslėlis yra kompanijų skaičiaus ir rodiklių išreikšimo langai parodyti 13-tame paveikslėlyje. Kiekvienam iš šių komponentų(rodiklių) langų, spalvų juosta esanti apačioje lango, rodo verčių diapazoną, kiekvienam komponentui. Žemesnės kiekvieno rodiklio vertės yra vaizduojamos mėlyna spalva, o aukštesnės vertės raudona spalva. Tarpinės vertės vaizduojamos nuo šviesiai mėlynos iki žalios iki geltonos iki oranžinės. Palyginant rodiklių vertes atitinkamose žemėlapių vietose, galima išskaityti netiesines priklausomybes ir taip vizualiai identifikuoti klasterių reikšmę. Žemėlapis grupių sudarytų pagal skirtingas kredito rizikas. Žemėlapis sudarytas SOM pagrindu naudojant WSJ duomenis surinktus iš Gregory Ip. Pirmasis žemėlapis sudarė penki klasteriai. Šis pirmasis žemėlapis buvo su šalių pavadinimų santrumpomis.



12 pav. Šalies kredito rizikos žemėlapis, grįstas WSJ duomenimis.

Kad atrasti netiesines priklausomybes tarp skirtingų rodiklių naudojamų kurti šį žemėlapi, 12 pav. yra vaizduojami skirtingų rodiklių langai. Išdėstant eilę rodiklių langų prie pagrindinio žemėlapi, galima atrasti priklausomybes tarp komponentų (rodiklių). Komponentų, tokių kaip dabartinio akcijos

kainos ir pelno santykio (angl. *current P/E*), būsimo P/E (angl. *forward P/E*), dividendų pelningumas, rinkos kapitalizacija, kompanijų skaičiaus ir kintamumo langai parodyti 13 paveikslėlyje. Kiekviename iš šių rodiklių langų, spalvų juosta esanti apačioje, rodo verčių diapazoną, kiekvienam rodikliui. Žemesnės kiekvieno rodiklio vertės yra vaizduojamos mėlyna spalva, o aukštesnės vertės raudona spalva. Tarpinės vertės vaizduojamos nuo šviesiai mėlynos iki žalios iki geltonos iki oranžinės. Palyginant rodiklių vertes atitinkamose žemėlapių vietose, galima išskaityti netiesines priklausomybes ir taip vizualiai identifikuoti klasterių reikšmę.



13 pav. Įprastas „Viscovery“ programinės įrangos langas [13]

Žemėlapis, kuriame suskirstyti klasteriai, vaizduoja, jog JAV, Indija ir Japonija yra skirtinguose klasteriuose. JAV ir Japonija turi rinkas su didele kapitalizacija; Indija savo biržoje turi didelį skaičių kompanijų sąrašą (vieną didžiausių įmonių skaičių po JAV). Turkija ir Lenkija išsiskiria kaip grupė ir visos kitos įmonės nėra diferencijuojamos. Akivaizdu, jog šiuo atveju, specialus dirbtinis SOM apribojimas į penkis klasterius nesuteikia daug naujos informacijos.

Jei šie specialūs dirbtiniai apribojimai yra panaikinami, yra gaunamas visiškai naujas šalių klasterizavimas, atsižvelgiant į tų šalių rizikas. 13 paveikslėlis vaizduoja žemiau išvardintus klasterius:

- C 1: Australija, Naujoji Zelandija, Kanada ir didžioji dalis Europos
- C 8: didžioji dali Lotynų Amerikos, ir Rytų Europa
- C 6: Meksika, Filipinai, Pietų Afrika ir Čekijos Respublika
- C 3: Korėja, Malaizija, Tailandas, ir Indonezija
- C 2: Singapūras ir Honkongas
- C 4: Vengrija ir Venesuela
- C 7: Brazilija
- C 5: Lenkija
- C 9: Indija ir Pakistanas

Pirmasis klasteris sugrupuoja didžiąją dalį išsivysčiusių rinkų. Antrasis klasteris sugrupuoja Azijos rinkas, kurios yra didesnės ir daug patrauklesnės nei visos kitos; Trečiasis klasteris sugrupuoja

didžiąją dalį ASEAN rinkų; ketvirtasis ir penktasis klasteriai identifikuoja kintančias rinkas arba rinkas, kurios įsiveržė į priekį kitų naujai atsirandančių rinkų. Į šiuos klasterius nepatenka JAV, Japonija, Kinija, Taivanas, Šri Lanka ir Suomija.

Tam, kad palengvinti žemėlapių palyginimus, buvo išlaikytas klasterių skaičius artimas pirmajame žemėlapyje. To pasėkoje, klasteriai nuo pirmo iki penkto yra labai arti pirmiems penkiems klasteriams ankstesniame SOM žemėlapyje; Šalių akcijų rinkos išvardintos nuo šešto iki devinto klasterio sudaro atskirus klasterius. Taigi buvo atrasta, jog šeštam klasteriui priskiriamos šalys kaip Meksika, Filipinai, Pietų Amerika ir Čekijos Respublika; Septintam klasteriui Brazilija; Aštuntajam Lotynų Amerikos šalys (išskyrus Meksiką ir Braziliją); Devintajame klasteryje yra Indija ir Pakistanas.

Šis pavyzdys įrodo keletą dalykų:

Duomenų transformacijos ir modifikacijos vertę: be rinkos kapitalizacijos logaritminių transformacijų ir išvardinto įmonių skaičiaus, JAV ir Japonija užimtų du klasterius, ir Singapūras ir Honkongas nebūtų išskirti iš kitų Azijos rinkų.

Jei nėra įvedama jokių apribojimų klasterių skaičiui, klasterio reikšmingumas (matas naudojamas „Viscovery Somine“ klasterių skaičiui optimizuoti) žymiai krenta, kai yra virš devynių klasterių. Taigi, devyni klasteriai yra geresnė duomenų reprezentacija nei esant mažiau klasterių nei devyni.

SOM žemėlapis, kuris numato duomenų rezultatų savęs organizavimą gana skirtingai grupuoja šalis, kai remiasi kredito rizikomis, lyginant su grupavimu remiantis viena šalimi ir neapibrėžtais kriterijais. SOM turi būti taikomas tik ten, kur naudotojas taip pat supranta ir duomenų prasmę. Žinių srities kompetencija yra būtina ir visada vaidina svarbų vaidmenį, jei norima efektyviai taikyti SOM. [13]

1.4. Programinė įranga

Atlikus programinių įrangų, kurios taiko savitvarkius neuroninius tinklus analizę. Buvo pasirinkta „Viscovery SOMine“ programa, su kuria vėliau buvo atliekamas įmonių klasterizavimas. Toliau aprašoma „Viscovery SOMine“ ir kelios kitos nagrinėtos programinės įrangos pasitelkiančios SOM.

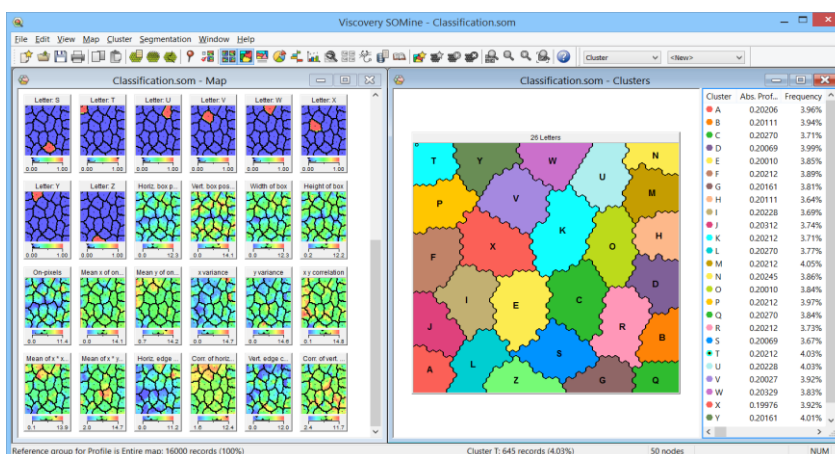
1.4.1. Viscovery SOMine

„Viscovery SOMine“ programa šiomis dienomis yra pirmaujanti tarp kitų duomenų gavybos programų taikančių savitvarkius neuroninius tinklus. „Viscovery“ technologinio pranašumo privalumai nurodomi žemiau:

- greitas ir glaustas modelio kūrimas net ir dideliems duomenų rinkiniams
- geras sudėtingų duomenų ir priklausomybių vizualizavimas
- tradicinės statistikos integravimas pateikiant vartotojui duomenis naujoviškais metodais
- intuityvus abstrakčių modelių ir analizės rezultatų pristatymas
- ekspertų žinių integravimas modeliavimo proceso metu
- išskirtinis prognozavimo tikslumas dėl patentuotos netiesinių ryšių gavybos procedūros [16]

Kitos „Viscovery SOMine“ savybės:

- daugialypių duomenų apdorojimo galimybės
- apibrėžimas(nustatymas) ir automatinis nominalių duomenų valdymas
- įrankiai histogramoms koreguoti
- savitvarkių neuroninių tinklų modelio kūrimas, naudojant iš anksto nustatytus tvarkaraščius
- interaktyvus savitvarkių neuroninių tinklų vizualizavimas ir analizė
- vizualinių klasterių analizė su integruotomis klasterių ribomis ir struktūromis
- statistinės funkcijos, tokios kaip aprašomoji statistika, histogramos, koreliacijos, išsibarstymo grafikas [17]



14 pav. „Viscovery SOMine“ optinė ženklų atpažinimo funkcija (angl. *optical character recognition*) [17]

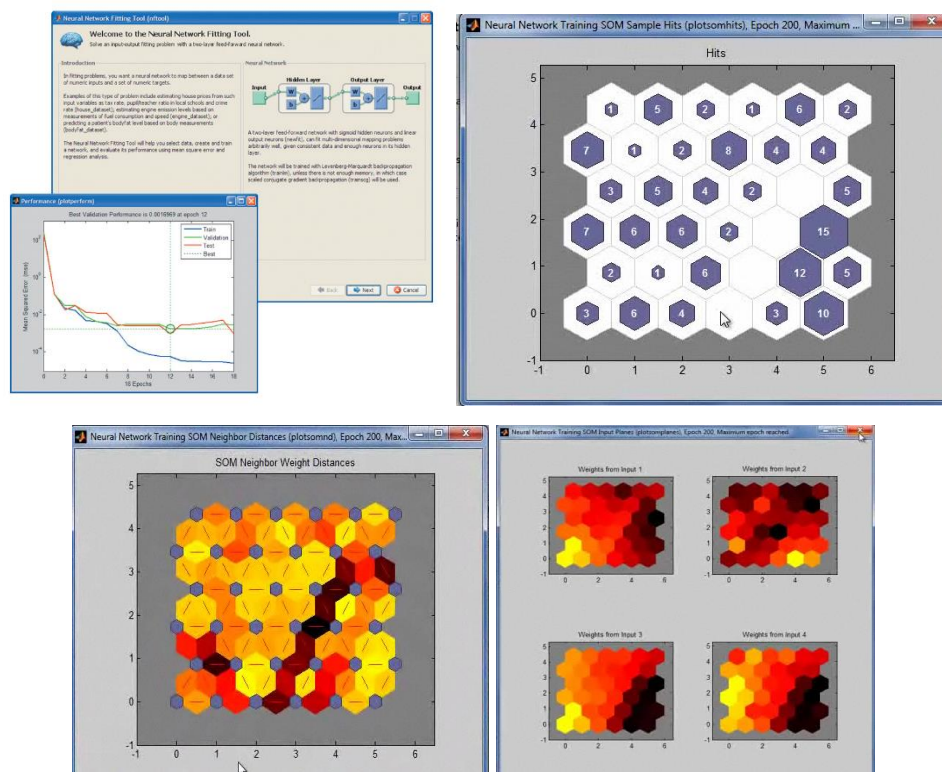
„Viscovery SOMine“ atliekamos funkcijos:

- **klasterizavimas.** Savitvarkiai neuroniniai tinklai esantys „Viscovery SOMine“ supaprastina klasterizavimą ir leidžia programos taikytojams atpažinti homogeniškas duomenų grupes vizualiai. Programoje Viscovery galimi keli klasterizavimo algoritmai (Single Linkage, Ward ir SOM-Ward) automatiškai sugeneruoti klasterius.

- **prognozavimas.** „Discovery“ sujungia netiesinį savitvarkių neuroninių tinklų duomenų vaizdavimą su tiesiniais statistiniais prognozavimo metodais kiekvienai homogeninei subgroupei, tam, kad pagerinti prognozavimo tikslumą.
- **duomenų atvaizdavimas.** Duomenys yra labai suspaudžiami taikant statistinius metodus, kurie leidžia vienam žemėlapiui, kuris naudoja tik kelis megabaitus vietos pavaizduoti duomenų bazes, kurių apimtis daug didesnė.
- **klasifikavimas realiu laiku.** Nauji duomenys žemėlapyje gali būti lokalizuojami labai greitai – iki 100,000 anksčiau nematytų duomenų įrašų gali būti klasifikuojami per sekundę – kas leidžia naujų duomenų įvertinimą realiu laiku.

1.4.2. Matlab/Deep Learning Toolbox

„Deep Learning Toolbox“ (anksčiau „Neural Network Toolbox“) yra Matlab programinės įrangos dalis, kuri yra priemonių rinkinys neuroninių tinklų pagrindu. Šis programinė įranga yra sukurta „MathWorks“ firmos.



15 pav. „Deep Learning Toolbox“ programos langas

„Deep Learning Toolbox“ gali apdoroti duomenis neuroniniais tinklais su mokytoju ir neuroniniais tinklais be mokytojo (angl. *supervised and unsupervised*). Įranga turi priemones skirtas neuroninių tinklų treniravimui, duomenų klasterizavimui, vizualizavimui, analizavimui, arba klasifikuoti triukšmingus duomenis ir kita. [18]

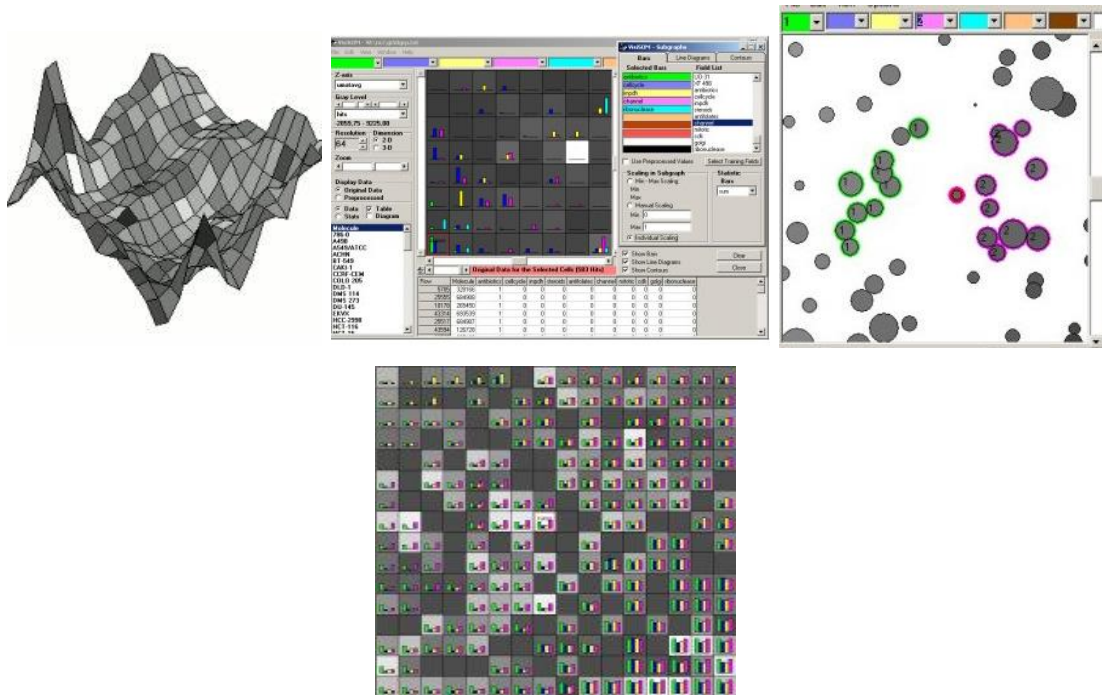
1.4.3. VisiSOM

Pasak Žemaitytės [18] „VisiSOM“ yra bendro duomenų tvarkymo priemonė, kuri suteikia tris pagrindines funkcijas:

1. Aptikti modelių panašumą duomenyse;

2. Sugeneruoti mažas ir įvairius subgrupių rinkinius, kurios reprezentuotų daugumą duomenų;
3. Aptikti elementus su tikrais ar hipotetiniais duomenimis [18]

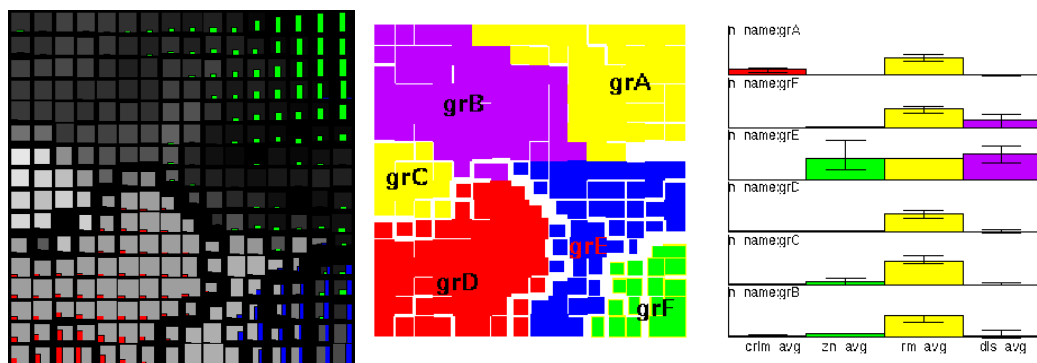
Programinėje įrangoje yra ir savitvarkiai neuroniniai tinklai. Yra galimybės klasterizuoti duomenis, vizualizuoti klasterius, sugeneruoti modelius ar šablonus, galimybė keisti grafikų tipus



16 pav.. „VisiSom“ programinė įranga [18]

1.4.4. Neural Data Analysis (NDA)

„Neural Data Analysis“ programinė įranga turi priemones duomenų manipuliacijai, redukcijai, vizualizacijai ir sprendimų priėmimui.



17 pav. „Neural Data Analysis“ langas [18]

NDA grįsta medžio struktūros savitvarkiais neuroniniais tinklų algoritmu. Programinė įranga taip pat taiko daugiasluoksnio perceptrono tinklo su klaidos skleidimo atgal (angl. *error backpropagation*) algoritmą. Pastarieji algoritmai derinami su statistiniais ir duomenų apdorojimo metodais [18].

1.5. Apibendrinimas

Apibendrinant galima teigti, jog savitvarkių neuroninių tinklų taikymas yra naudingas ypač tuo atveju, kai reikia apdoroti didelius kiekius duomenų, kurių paprastai negali apdoroti įprastos skaičiuoklės. Finansų rinkose savitvarkių neuroninių tinklų taikymas padeda atpažinti akcijų grupes, kurios turi panašių kainų kitimo dėsnius tam tikram laiko tarpe ir taip padeda investuotojams priimti sprendimus. Programinės įrangos grįstos savitvarkiais neuroniniais tinklais verslininkams padeda atlikti tikslesnę konkurentų analizę ir taip priimti sprendimus.

2. Savitvarkių tinklų taikymas kompanijų klasterizavimui

2.1. Tyrimo organizavimas

Kad atrinkti norimas akcijas, techninės analizės specialistai naudoja portalus, kuriuose pateikiama informacija apie vertybinius popierius. Vienas iš patogiausių ir dažniausiai naudojamų tokių portalų yra finviz.com, nes jame galima filtruoti (angl. *to screen*) didelę įvairovę fundamentinių, techninių ir aprašomųjų rodiklių. Šis portalas taip pat pateikia kiekvienos akcijos metinę, pusmetinę, metų ketvirčio, mėnesinę ir savaitės vertybinių popierių kitimo rezultatus, kas taip pat aktualu šiame tyrime. Taip pat portale galima stebėti akcijų kitimo grafikus, jame galima sudaryti investicinius portfelius ir stebėti jų našumą laike.[19]

2.1.1. Rodiklių parinkimas

Išanalizavus literatūrą apie finansinius rodiklius, buvo prieita išvados pasirinkti kelis reikšmingiausius rodiklius, pagal kuriuos buvo atliktas įmonių vertybinių popierių klasterizavimas. Žemiau pateikiami pasirinkti rodikliai:

Vienai akcijai tenkančio pelno augimas (angl. *earnings per share growth rate – EPS*). EPS yra įmonės pelno dalis, paskirta kiekvienai bendrųjų akcijų daliai. Tačiau EPS gali labai skirtis, nuo įmonės akcijų kainos. Kita vertus, EPS yra pagrindinė informacija, kuri daro įtaką įmonės akcijų kainai. Vienai akcijai tenkančio pelno augimas yra įmonės pelningumo ir patikimumo rodiklis.

$$EPS = (\text{Grynasis pelnas})/(\text{Vidutinis skaičius paprastųjų akcijų}) \quad (3.1)$$

Arba:

$$ESP = \frac{I_N - D_P}{S_A}; \quad (3.2)$$

kur I_N ataskaitinio laikotarpio tiesioginis pelnas, D_P – per ataskaitinį periodą apskaičiuoti dividendai pagal privilegijuotas akcijas, S_A – ataskaitiniu periodu esantis vidutinis paprastų akcijų skaičius. Dauguma investicinių fondų renkasi tik tokias kompanijas, kurių vienai akcijai tenkantis pelno augimas viršija 15 – 20 proc. [6][20][21][22]

Grynojo pelno ir nuosavo (akcinio) kapitalo santykis (angl. *return on equity - ROE*). ROE laikomas rodikliu, kuris nurodo kaip efektyviai panaudojamas įmonės turimas kapitalas sukurti pelną. Investuotojui ROE rodiklis parodo, kokio didumo grąžą generuoja įsigyta akcija vienam investuotam investuotojo nuosavybės doleriui:

$$ROE = \text{Grynasis pelnas}/\text{Nuosavas(akcinis)kapitalas}; \quad (3.4)$$

Kuo ROE rodiklis yra didesnis, tuo įmonės veikla laikoma efektyvesne ir tuo akcininkai gauna didesnę pelną sugeneruota iš šios įmonės akcijų. Dažnai prieš įsigydami akciją investuotojai apsibrėžia kokio pelno iš jos tikisi ir būtent ROE parodo, ar realūs įmonės rezultatai gali atitikti išankstinius lūkesčius. [6][21][23]

Akcijos kainos ir pelno santykis (angl. *price/earnings ratio - P/E*). Tai vienas iš pačių populiariausių ir paprasčiausių įmonės vertės nustatymo rodiklių:

$$P/E = \text{Akcijos rinkos kaina (P)} / \text{Grynasis pelnas vienai akcijai (EPS)}; \quad (3.5)$$

PE rodiklis parodo, kiek akcijas perkantis investuotojas moka už vieną bendrovės JAV dolerį. Nustatydamas akcijos rinkos kainos ir jai tenkančio pelno dalies ryšį, šis faktorius leidžia investuotojui palyginti dvi įmones ar daugiau [6][21][24] [25]

Jei turima šio rodiklio vertė per pastaruosius metus lygi 10, tai reiškia, kad investuotojai už vieną dolerį įmonės pelno sutinka sumokėti dešimt dolerių. Yra manoma, kad kuo mažesnis P/E rodiklis, tuo mažesnė ir akcijos kaina. Žemas rodiklis taip pat gali byloti apie pesimistinę investuotojų nuomonę vertinant šios akcijos kainos kilimą. Tuo tarpu aukšta rodiklio vertė gali rodyti, kad tikimasi spartaus akcijos augimo ateityje. P/E laikomas vienu populiariausių akcijos vertinimo rodikliu. Laikoma, kad akcijos pigios kai šis rodiklis mažiau nei 10, o brangios – kai daugiau nei 20. [29][37][38].

Ilgalaikės skolos ir nuosavybės santykis (angl. *long term debt to equity*). Yra metodas nustatyti įmonės finansinį svertą. Kad gauti rodiklį, apskaičiuojamas ilgalaikių įsiskolinimų ir akcininkų nuosavybės santykis:

$$\text{Ilgalaikės skolos, nuosavyb. kof.} = \text{Ilgalaikiai įsipareig.} / (\text{Visas turtas} - \text{Visi įsipareig.}) ; \quad (3.6)$$

Ilgalaikiai įsipareigojimus sudaro daugiau nei vienerius metus trunkančios paskolos ir finansiniai įsipareigojimai. Į ilgalaikius įsipareigojimus įmonei apimtų bet kokius finansavimo ar lizingo įsipareigojimus, kurie turi būti įvykdyti per ilgesnė nei 12 mėnesių laikotarpį (30 metų ir daugiau). Ilgalaikiai įsipareigojimai taip pat taikomi ir valstybėms, nes ir jos gali turėti finansinių įsipareigojimų. [26]

Konkrečiai ilgalaikių įsipareigojimų vertinimui ir skirtas ilgalaikės skolos ir nuosavybės santykio rodiklis. Pasak Andriuškos [6] šis rodiklis parodo „*kokią dalį nuosavo ir skolinto kapitalo bendrovė naudoja finansuoti savo turtą. Kuo šis rodiklis yra didesnis, tuo didesnė išorinių finansavimo šaltinių įtaka ir tuo didesnė tikimybė, kad įmonė nebesugebės sumokėti palūkanų. Kuo šis koeficientas mažesnis, tuo įmonė patrauklesnė investuotojams*“.[6]

Didesnis rodiklis dažniausiai parodo aukštesnį verslo rizikos laipsnį, nes kompanija turi didesnių įsipareigojimų. O potencialūs kreditoriai nenori investuoti į įmones turinčias daug skolų. Kita vertus, įsiskolinimų dydis priklauso nuo verslo tipo. Pavyzdžiui, bankas gali turėti didelį įsiskolinimą, bet jo turtas paprastai yra finansiškai likvidus. Bankas gali turėti didesnę rodiklį nei pavyzdžiui pramonės kompanija, nes banko pajamos yra stabilesnės.[26]

Bendrasis likvidumo rodiklis (angl. *current ratio CR*). Likvidumo rodiklis matuoja kompanijos gebėjimą susimokėti už trumpalaikius įsipareigojimus, kurie yra vienerių metų laikotarpyje. CR investuotojams ir analizės specialistams parodo, kaip bendrovė gali padidinti savo trumpalaikį turtą, siekdama padengti savo dabartines skolas ir kitas mokėtinas sumas. Bendrasis likvidumo rodiklis tiesiogiai priklauso nuo bendrovės apyvartinio kapitalo dydžio – kuo kapitalas didesnis tuo didesnė šio koeficiento vertė. Bendrojo likvidumo koeficientas apskaičiuojamas kaip trumpalaikio turto ir trumpalaikių įsipareigojimų santykis: [6][21][28]

$$\text{Bendrasis likvidumo rodikis} = \text{trumpalaikis turtas} - \text{trumpalaikiai įsipareigojimai} \quad (3.7)$$

Kompanijos akcijų dalis, kuri tenka firmos darbuotojams (angl. *insider ownership*). Koeficientas, kuris procentais parodo, kokia įmonės akcijų dalis priklauso tos įmonės darbuotojams.

$$\text{Darbuotojų akcijų dalis} = (\text{Darbuotojų turimų akcijų sk.} / \text{Visas akcijų sk.}) * 100 \quad (3.8)$$

Firmos darbuotojai yra bendrovės pareigūnai, direktoriai, giminės ar kiti asmenys, turintys prieigą prie pagrindinės įmonės informacijos, dar jai esant viešai nepaskelbtai. Todėl, kai šis rodiklis yra aukštas, jis gali rodyti, jog įmonei galimai sekasi ir investuojantys vidaus įmonės darbuotojai nemato priešasčių, kodėl pelningumas turėtų kristi. Dėl šios priežasties investuotojai ieško akcijų su aukšta šio koeficiento verte. Dažniausiai investiciniai fondai įsigija akcijas tada, kai rodiklis viršija nors 10 – 20 procentų. [6][20][28]

Kompanijos akcijų dalis, kurią valdo žinomi investiciniai fondai (angl. *institutional ownership*). Rodiklis parodantis kiek procentų firmos akcijų priklauso investiciniams fondams:

$$\text{Invest. fond. akcijų dalis} = (\text{Invest. fond. turimų akcijų sk.} / \text{Visas akcijų sk.}) * 100 ; \quad (3.9)$$

Dažnai palankiai vertinamos akcijos, kurių didelę dalį valdo investiciniai fondai. Didelės įmonės dažnai samdo analitikų komandą, kad atliktų išsamų ir brangų finansinį tyrimą, kol grupė investuotojų įsigyja didelį įmonės akcijų paketą. Todėl pastarųjų investuotojų priimami sprendimai, daro įtaką kitų investuotojų priimamiems investavimo sprendimams. Žinoma, jog didesni investiciniai fondai prieš įsigydami įmonės akcijas vykdo itin nuodugnią analizę ir jei fondas nusprendė įsigyti šią akciją, tai labai tikėtina, kad ateityje investicinis fondas tikisi akcijos kylančios tendencijos. Priimta, kad akcijų kaina turi didelę tikimybę kilti aukštyn, kai investiciniai fondai turi bent 15 – 30 procentų kompanijos akcijų. Didesnis koeficientas gali rodyti, jog akcijos kaina pervertinta vien dėl institucinių fondų susidomėjimo. [6][21][30]

2.2. Vertybinių popierių atranka

Anksčiau parinkti rodikliai naudojami vertybiniais popieriams atrinkti. Taip pat, įmonių atrankos metodas bus tikrinamas su skirtingais sektoriais, nes skirtingus sektorius veikia skirtingi ekonominiai veiksniai. Todėl būtų neracionalu pasirinkti tik vieną sektorių, nes yra tikimybė, jog jis gali patirti ekonominį nuosmukį, dėl ko didėtų tikimybė, jog ir pats portfelis sudarytas iš to sektoriaus vertybinių popierių patirs nuostolį. Todėl bus sudaromi 5 portfeliai 5 skirtingiems sektoriams. Finviz.com platformoje pasirenkami 5 sektoriai: sveikatos, technologijų, plataus vartojimo prekių (angl. *consumer goods*), pramoninių prekių (angl. *industrial goods*) ir paslaugų. Iš kurių kiekviename pagal ankstesniame skyrelyje pateiktus rodiklius bus atrinktos akcijos. Rodiklių vertės parenkamos pagal literatūroje pateiktus reikalavimus ir vėliau pakoreguojamos taip, kad kiekviename sektoriuje akcijų būtų tarp 20-30.

Home News Screener Maps Groups Portfolio Insider Futures Forex Crypto Backtests Elite Thu NOV 29 2018 3:56 PM EST Help karolis.pok

s: Paslaugu Order: Sector Asc Signal: None (all stocks) Tickers: Filters

Filters: 8

Descriptive(1)		Fundamental(7)		Technical		All(8)	
P/E	Over 15	Forward P/E	Any	PEG	Any	P/S	Any
Price/Cash	Any	Price/Free Cash Flow	Any	EPS growth this year	Over 20%	EPS growth next year	Any
EPS growth next 5 years	Any	Sales growth past 5 years	Any	EPS growth qtr over qtr	Any	Sales growth qtr over qtr	Any
Return on Equity	Over +10%	Return on Investment	Any	Current Ratio	Over 1	Quick Ratio	Any
Debt/Equity	Under 1	Gross Margin	Any	Operating Margin	Any	Net Profit Margin	Any
Insider Ownership	Low (<5%)	Insider Transactions	Any	Institutional Ownership	Over 10%	Institutional Transactions	Any

Overview Valuation Financial Ownership Performance Technical Custom Charts Tickers Basic TA News Snapshot Stats

18 pav. Finviz.com portale nustatomi reikiami rodikliai

18 pav. rankiniu būdu finviz.com portale pasirenkamas apžvalgos laukelis (angl. *overview*) pasirenkami reikiami rodikliai, kurių vertės nustatomos. Taip pat pasirinkus laukelį (angl. *descriptive*) nustatomas vienas iš penkių sektorių.

Vėliau finviz.com portale pasirenkamas laukelis (angl. *custom*) 19pav., kuriame nustatome, kokie duomenys mums aktualūs ir, kuriuos vėliau įkeliame į MS excel failą.

No.	<input checked="" type="checkbox"/> Ticker	Company	Sector	Industry	Country	Market Cap.	<input checked="" type="checkbox"/> P/E	Forward P/E	PEG
P/S	P/B	P/Cash	P/Free Cash Flow	Dividend Yield	Payout Ratio	EPS	<input checked="" type="checkbox"/> EPS growth this year	EPS growth next year	EPS growth past 5 years
EPS growth next 5 years	Sales growth past 5 years	EPS growth qtr over qtr	Sales growth qtr over qtr	Shares Outstanding	Shares Float	<input checked="" type="checkbox"/> Insider Ownership	<input checked="" type="checkbox"/> Insider Transactions	<input checked="" type="checkbox"/> Institutional Ownership	<input checked="" type="checkbox"/> Institutional Transactions
Float Short	Short Ratio	Return on Assets	<input checked="" type="checkbox"/> Return on Equity	Return on Investments	<input checked="" type="checkbox"/> Current Ratio	Quick Ratio	<input checked="" type="checkbox"/> Long Term Debt/Equity	Total Debt/Equity	Gross Margin
Operating Margin	Net Profit Margin	Performance (Week)	Performance (Month)	Performance (Quarter)	Performance (Half Year)	Performance (Year)	Performance (YearToDate)	Beta	Average True Range
Volatility (Week)	Volatility (Month)	20-Day Simple Moving Average	50-Day Simple Moving Average	200-Day Simple Moving Average	50-Day High	50-Day Low	52-Week High	52-Week Low	RSI
Change from Open	Gap	Analyst Recom.	Average Volume	Relative Volume	Price	Change	Volume	Earnings Date	Target Price
IPO Date									Defaults

Overview Valuation Financial Ownership Performance Technical Custom Charts Tickers Basic TA News Snapshot Stats

19 pav. Pasirinktas duomenų filtras

Total: 29 #1 save as portfolio | create alert Auto Refresh: 3min | off Page 1/2

Ticker	P/E	EPS this Y	Insider Own	Inst Own	ROE	Curr R	LTDebt/Eq
BGSF	15.80	23.70%	3.20%	43.20%	23.50%	2.10	0.32
MSM	17.21	25.50%	0.60%	95.70%	24.40%	2.30	0.22
ARCB	17.44	87.80%	2.70%	97.10%	13.10%	1.40	0.33
PLCE	18.20	39.30%	1.70%	95.00%	17.70%	1.30	0.00
DLTR	18.29	28.10%	1.20%	94.60%	24.20%	2.20	0.67
ICFI	19.29	36.20%	3.20%	93.20%	11.10%	1.70	0.36
ZTO	19.87	45.50%	0.29%	31.30%	16.40%	5.20	0.00
AMRC	20.22	98.90%	0.50%	46.40%	14.40%	1.40	0.62
ROST	20.23	32.70%	0.50%	91.60%	47.80%	1.70	0.10
ERA	20.32	29.20%	2.40%	85.90%	18.10%	3.60	0.34
ODFL	20.60	22.40%	0.10%	74.30%	26.40%	1.90	0.02
SFM	21.32	22.00%	0.40%	99.90%	29.20%	1.20	0.87
BKNG	21.57	74.80%	0.20%	95.50%	26.40%	2.40	0.87
FWRD	22.85	161.70%	1.10%	99.40%	18.40%	2.70	0.07
NVEE	26.14	83.00%	2.70%	63.10%	14.10%	3.40	0.06
ULTA	27.62	46.90%	1.00%	93.10%	35.80%	2.40	0.00
LOPE	27.87	26.70%	1.30%	99.70%	20.70%	4.60	0.05
CTAS	31.24	33.50%	1.50%	67.60%	28.00%	3.10	0.76
WLDN	31.28	22.30%	2.80%	58.00%	15.50%	1.70	0.00
NSP	32.11	34.60%	2.30%	87.20%	144.40%	1.20	0.69

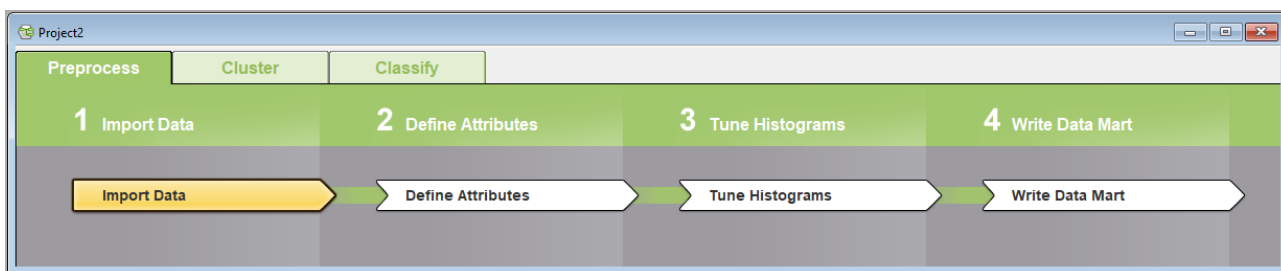
Filters: fa_curratio:o1 fa_debtq:u1 fa_epsyoy:o20 fa_pe:o15 fa_roe:o10 sec:services sh_insi...

20 pav. Eksportuojamų duomenų iš finviz.com portalo į MS Excel failą pavyzdys.

Eksportuoti duomenys 20 pav. išsaugojami MS Excel failuose ir vėliau apdorojami Viscovery SOMine programa, kuri remiasi savitvarkių neuroninių tinklų algoritmu ir geba grupuoti duomenis,

atrasdama juose koreliacijas ir šablonus (angl. *patterns*) , tarp duomenyse esančių rodiklių ir įmonių vertybinių akcijų.

2.3. Klasterizavimo algoritmo realizavimo žingsniai

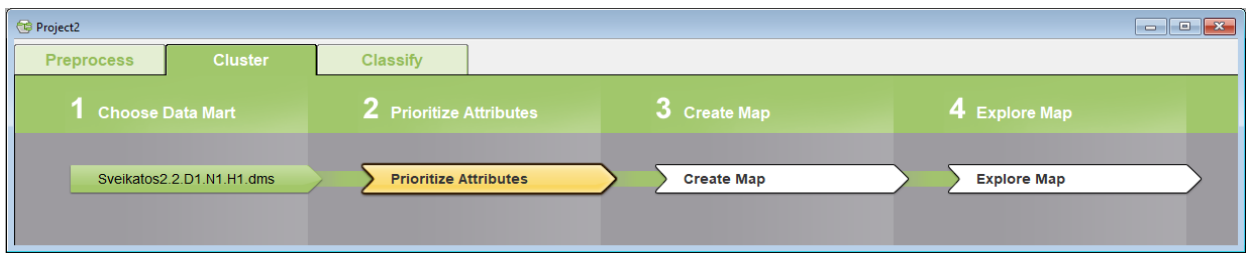


21 pav. Viscovery SOMine duomenų apdorojimo schema

Panagrinėsime pavyzdį, kaip Viscovery SOMine programa atliekamas klasterizavimas. Viscovery SOMine programoje sukuriamas naujas projektas ir duomenys parengiami tolimesniam jų apdorojimui: importuojami duomenys iš MS Excel failų pirmajame Viscovery SOMine duomenų apdorojimo schemos (21pav.) žingsnyje. Pavyzdžiui, pasirenkamas failas su duomenimis iš sveikatos sektoriaus. Programos antrajame žingsnyje 22 pav. apibrėžiamos šių duomenų charakteristikos (skaitinės arba tekstinės), bei vardinės vertės (įmonių tekstiniai simboliai), kurių atžvilgiu programa klasterizuos. Trečiajame žingsnyje galima stebėti kiekvieno rodiklio verčių pasiskirstymą dažnių histogramose ir derinti šias histogramas. Ketvirtajame žingsnyje įrašomi duomenys (angl. *write data mart*).

Attribute	Minimum	Maximum	Mean	Std. Deviation	Missing Values	Sum
Ticker	24 nominal values				0 (0.00%)	
73 P/E	15.34	69.21	42.74	16.16	0 (0.00%)	1025.65
73 EPS this Y	0.202	3.270	0.808	0.784	0 (0.00%)	19.397
73 Insider Own	0.00060	0.04900	0.00948	0.01186	0 (0.00%)	0.22760
73 Inst Own	0.555	1.000	0.848	0.120	0 (0.00%)	20.343
73 ROE	0.1010	0.5570	0.1992	0.1063	0 (0.00%)	4.7800
73 Curr R	1.10	12.40	4.40	3.21	0 (0.00%)	105.60
73 LTDebt/Eq	0.0000	0.7000	0.2304	0.2501	0 (0.00%)	5.5300

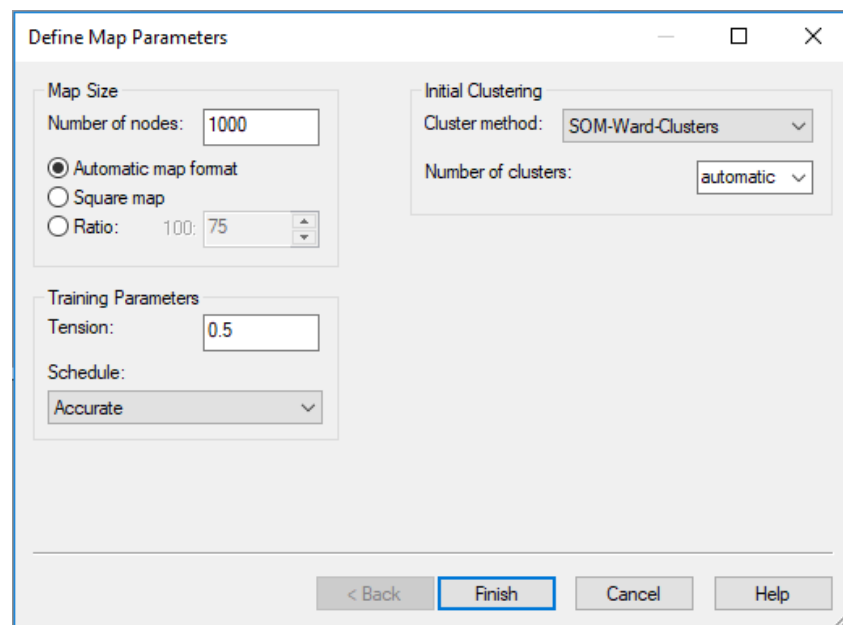
22 pav. Antrajame duomenų apdorojimo žingsnyje aprašomos duomenų charakteristikos, bei vardinės vertės



23 pav. Viscovery SOMine klasterio žemėlapio kūrimo schema

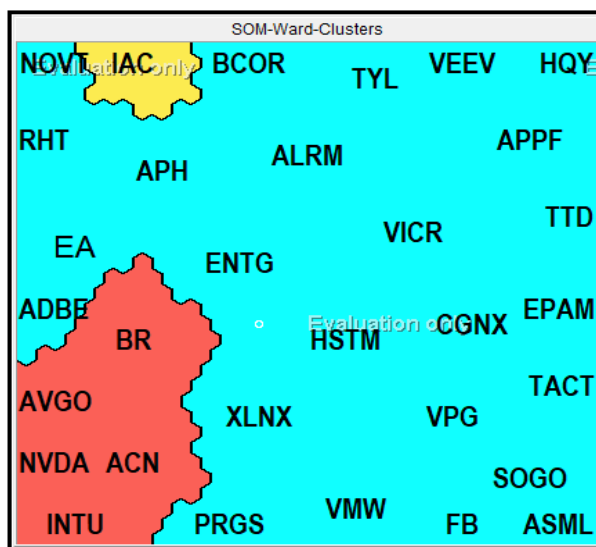
Atsivėrusi Viscovery SOMine klasterio žemėlapio kūrimo schema 23 pav. taip pat turi keturis žingsnius. Pirmajame žingsnyje pasirenkami programos duomenys, kurie anksčiau apdoroti ir išsaugoti Viscovery SOMine duomenų apdorojimo schemas pagalba. Antrajame žingsnyje (angl. *prioritize attributes*) nustatomi rodiklių prioritetai, kiekvienam rodikliui priskyrus skaičių nuo 1 iki 10. Šiame tyrime laikoma, jog visų rodiklių svarba yra tokia pati. Todėl visiems septyniems rodikliams arba atributams priskiriami vienodi prioritetai pasirinkus skaičių 1.

Trečiajame žingsnyje - žemėlapio kūrimas (angl. *Create Map*) 24 pav. sukuriame žemėlapią priskirdami kiek mazgų jame naudosisime (angl. *number of nodes*). Parenkamas klasterizavimo metodas *SOM-Ward-Cluster* ir tikslusis režimas (angl. *accurate*). Sukuriamas žemėlapis, kuris susideda iš trijų klasterių.



24 pav. Visionary SOMine programoje pasirenkamas nustatomi žemėlapio charakteristikos

Atliktus parametrų nustatymo procesą duomenys apie įmonių vertybinius popierius automatiškai suklasterizuojami į tris klasterius pagal 7 rodiklius. Žemiau parodytas, suklasterizuotų įmonių žemėlapias pavyzdys 25 pav.



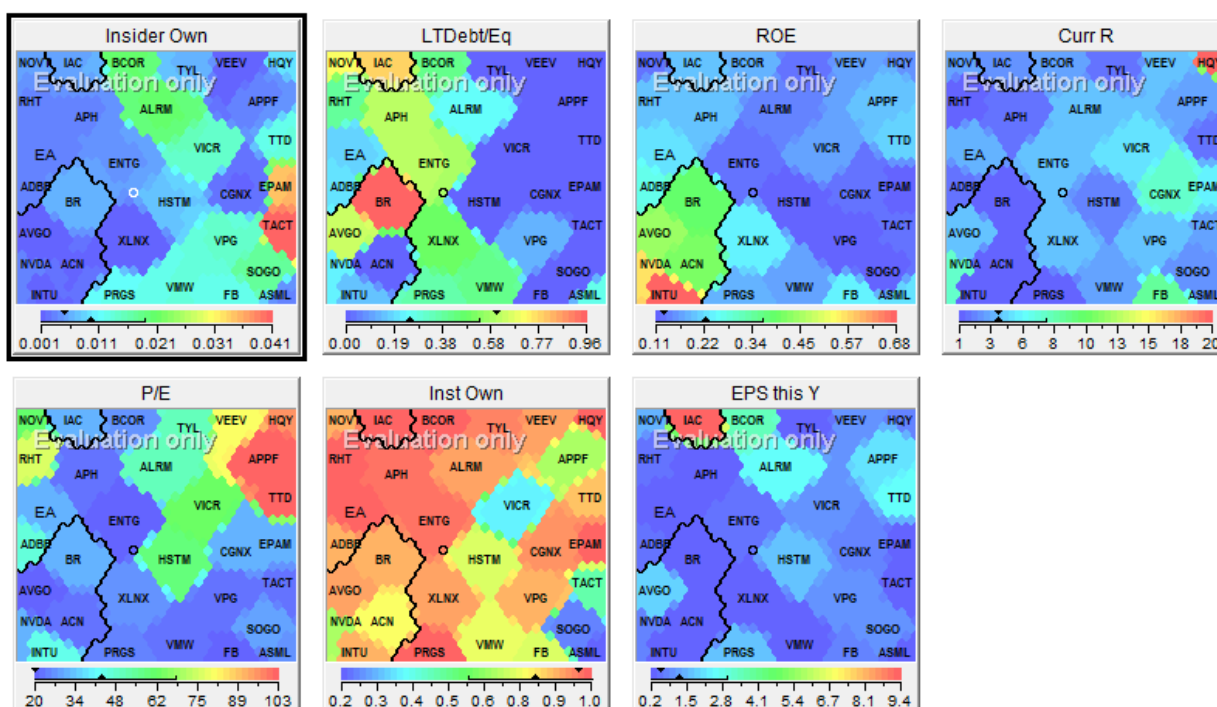
25 Pav. Pavyzdys: „Viscovery SOMine“ technologijų sektoriaus žemėlapis

Iš žemėlapio galima pastebėti, jog pirmajam mėlynos spalvos klasteriui priskiriamos toki įmonių akcijos kaip: NOV, BCOR, TYL, VEEV, HQY, RHT, APH, ALRM, APPF, EA, ENTG, VICR, TTD, ADBE, HSTM, CGNX, EPAM, XLNX, VPG, TACT, PRGS, VMW, FB, ASML.

Antrajame klasteryje yra tokios įmonės: BR, AVGO, NVDA, ACN, ITU.

Trečiajam klasteriui priskiriama įmonė: IAC

Klasterizavimas automatiškai atliekamas pagal rodiklių vertes. Įmonės su panašiais rodikliais įtraukiamos į vieną klasterį. Žemiau esančioje lentelėje pateikiama, kiekvieno klasterio rodiklių vidutinės vertės.



26 pav. Rodiklių išreikštumo žemėlapiai

Kiekvienas rodiklis lemia, žemėlapio struktūrą, kur tam tikros įmonės žemėlapyje atsiras. 26 pav. galima pastebėti, rodiklių išreikštumą kiekvienai įmonei vadinamuose rodiklių „šalčio-šilumos“

žemėlapiuose. Raudonai nuspalvintos žemėlapių dalys vaizduoja įmones, kurios turi aukštas rodiklių vertes, geltonai žaliai vidutines rodiklių vertes, o mėlynai- žemas rodiklių vertes. Atitinkamai galima pamatyti, kurioms įmonėms turi didžiausias rodiklių vertes. Svarbiausia, jog pagal šiuos žemėlapius skirtingiems sektoriams ir išvelgti dėsningumas, koks rodiklis turi didžiausią įtaką to sektoriaus įmonėms. Pavyzdžiui, 26 pav. Matome, jog didžiausią įtaką turi rodiklis Kompanijos akcijų dalis, kurią valdo žinomi investiciniai fondai (angl. *institutional ownership*).

2.4. Pelningiausio klasterio atranka

Žinant trijuose klasteriuose pasiskirsčiusias įmones patikrinamas kiekvieno šių klasterių savaitinis pelningumas (angl. *Week performance*). Tai atliekama vieno iš klasterio įmonių santrumpas (angl. *Tickers*) suvedus į finviz.com platformoje esantį atrankos filtrą (angl. *Screener*) 27 pav. Šie duomenys nukopijuojami į MS Excel failus, kur vėliau patikrinamas kiekvieno klasterio įmonių savaitinio pelningumo vidurkis 28 pav.

The screenshot shows the Finviz.com screener interface. At the top, the 'Tickers' field is populated with 'CALM, COLM, DORM, EBF, IPAR, KEQU, MNST, UNF'. Below this, various filters are set to 'Any'. The main table displays performance metrics for 8 tickers. The 'Perf Week' column is highlighted in red, and the 'Perf Week' values are also highlighted in red in the original image.

No.	Ticker	Perf Week	Perf Month	Perf Quart	Perf Half	Perf Year	Perf YTD	Volatility W	Volatility M	Recom	Avg Volume	Rel Volume	Price	Change	Volume
1	CALM	1.80%	2.98%	0.00%	-12.97%	-9.05%	3.10%	1.84%	1.76%	3.70	289.42K	1.05	43.61	1.30%	301,855
2	COLM	-6.49%	-9.63%	-13.32%	-0.19%	12.55%	10.26%	2.43%	2.29%	1.90	350.22K	1.05	92.72	-3.82%	381,719
3	DORM	-3.35%	-9.43%	-4.07%	2.69%	27.88%	-7.11%	2.42%	3.19%	3.00	196.84K	0.67	83.62	-2.39%	130,601
4	EBF	-1.46%	-7.93%	-4.89%	-1.26%	6.98%	1.97%	1.57%	2.10%	-	110.64K	0.97	19.63	-0.05%	106,967
5	IPAR	-6.17%	-8.57%	-0.81%	10.85%	42.36%	4.10%	3.61%	2.98%	2.80	135.95K	0.48	68.26	-1.54%	64,534
6	KEQU	-1.03%	3.40%	-27.96%	-22.47%	-33.87%	-32.36%	1.14%	2.45%	-	7.50K	0.46	22.49	-0.08%	3,371
7	MNST	-0.45%	13.86%	7.48%	11.50%	24.53%	26.66%	1.92%	2.31%	2.30	4.76M	0.70	62.34	-1.05%	3,274,554
8	UNF	-5.57%	1.56%	10.87%	5.16%	-8.27%	9.83%	1.95%	1.68%	-	130.94K	0.83	157.13	-3.27%	107,530

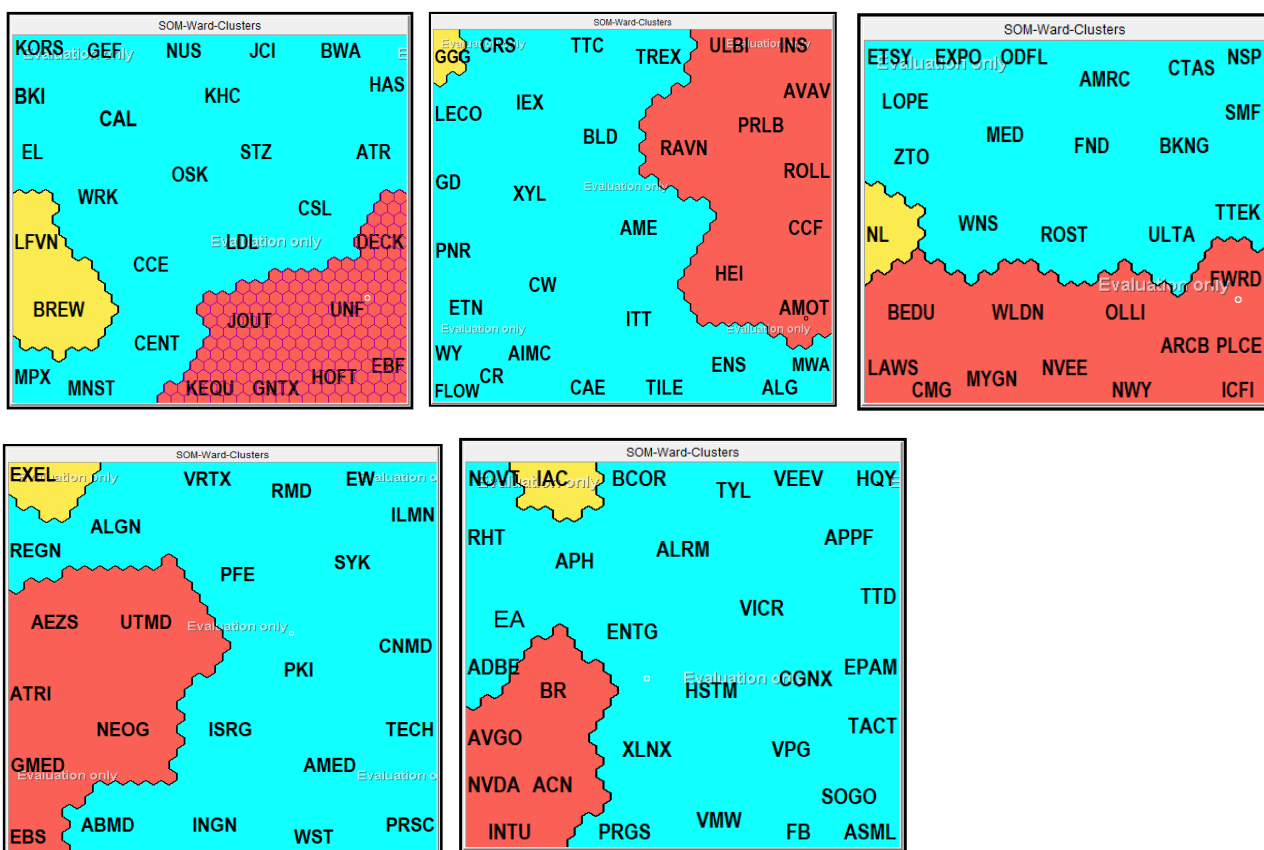
27 pav. Pavyzdys: Įmonių plataus vartojimo prekių C2 klasterio portfelis. Raudonai apibrėžta įmonių suvedimo laukelis (angl. *Tickers*) ir įmonių savaitinis pelningumas (angl. *Week performance*)

1	BC	-0,0538	-0,0455	0,1027	-0,2816	-0,1879	0,0706	0,0267	0,0219	1,8	956.53K	0,89	49,73	-0,0424	855998	C1
2	CAL	-0,1162	-0,2055	-0,1701	-0,3468	-0,2865	-0,1552	0,0449	0,0329	-	379.79K	5,41	23,51	-0,087	2053025	
3	CSL	-0,0282	-0,0144	0,2613	-0,0575	0,1232	0,1978	0,021	0,0144	1,9	395.53K	0,8	120,4	-0,0238	317622	
4	DECK	-0,0089	-0,0192	0,1446	0,247	0,5323	0,0982	0,0285	0,0232	2,8	634.44K	0,6	140,51	-0,039	382578	
5	ELY	-0,0574	-0,068	0,0294	-0,3103	-0,0402	0,0301	0,0251	0,0217	2,2	1.04M	0,99	15,76	-0,0535	1022087	
6	FLO	0,0197	-0,0119	0,1216	0,0751	-0,0024	0,1234	0,0169	0,0172	3,3	1.13M	0,97	20,75	0,0112	1096856	
7	GEF	-0,0131	-0,0055	-0,0487	-0,308	-0,2276	0,0784	0,0249	0,0252	2,9	275.89K	0,8	40,02	-0,0345	221972	
8	HNI	-0,0694	-0,1467	0,0353	-0,1496	-0,0843	0,0028	0,0281	0,0251	2,5	195.08K	1,19	35,53	-0,0492	232039	
9	JCI	0,0084	0,0152	0,1719	-0,0235	-0,0055	0,2185	0,0198	0,0156	2,5	4.56M	0,73	36,13	-0,0225	3322465	
10	MLR	-0,0472	-0,0731	0,1563	0,1747	0,2403	0,1507	0,0333	0,0303	-	32.87K	1,38	31,07	-0,0416	45258	
11	MNRO	0,0487	0,0329	0,0975	0,1416	0,4569	0,1491	0,0392	0,0283	2,4	341.05K	1,12	79	-0,0485	380922	
12	OSK	-0,0381	-0,0876	0,1659	-0,0413	-0,0907	0,1806	0,0261	0,0234	2,2	750.47K	0,61	72,38	-0,0398	457766	
13	PKG	-0,0109	-0,0318	0,1099	-0,1923	-0,1914	0,1398	0,0236	0,0204	2,3	940.25K	0,64	95,13	-0,0319	597787	
14	SHOO	-0,0123	-0,012	0,092	-0,1534	0,0513	0,0631	0,024	0,0232	2,5	562.20K	1,09	32,17	-0,0336	613206	
15	SON	0,0034	-0,0049	0,1	0,0409	0,2372	0,1161	0,0156	0,0147	-	531.88K	0,93	59,3	-0,0161	493200	
16	SRI	-0,026	-0,0066	0,1235	-0,0455	-0,0514	0,1075	0,0278	0,0284	-	251.62K	1,25	27,3	-0,0531	314125	
17	THO	-0,0913	-0,1644	0,1	-0,4523	-0,5165	0,1127	0,0353	0,0355	1,6	922.61K	1,25	57,86	-0,059	1149683	
18	UFS	-0,0172	-0,0926	0,327	-0,0955	0,0728	0,3803	0,0225	0,0184	2,8	828.06K	0,52	48,49	-0,0296	430954	
	Vidurkis	-0,02832														
	Top 5	0,01426														
	St. Nuokrypis	0,02182														
1	CALM	-0,0113	-0,0249	0,0005	-0,1121	-0,0262	0,0359	0,0194	0,0179	3,7	280.72K	0,55	43,82	-0,0126	154983	C2
2	COLM	-0,0257	-0,0356	0,2045	0,1016	0,3046	0,202	0,0213	0,0171	1,9	361.12K	0,68	101,08	-0,0372	246469	
3	DORM	0,021	-0,0695	0,0071	0,0603	0,2141	-0,0745	0,0205	0,0316	3	246.96K	0,76	83,31	-0,0178	187957	
4	EBF	-0,0684	-0,0425	0,1633	-0,0287	0	0,0545	0,0249	0,0182	-	141.36K	0,89	20,3	-0,0324	125836	
5	IPAR	-0,0212	0,0232	0,2149	0,1785	0,5767	0,1458	0,0237	0,026	2,8	131.08K	0,72	75,13	-0,0225	94751	
6	KEQU	-0,0306	-0,3283	-0,0998	-0,3605	-0,2814	-0,3622	0,0691	0,0549	-	10.55K	1,79	21,2	-0,0093	18838	
7	MNST	-0,1122	-0,0912	0,0823	-0,0979	-0,0787	0,0872	0,028	0,0233	2,2	3.80M	2,2	53,51	-0,0127	8340833	
8	UNF	-0,0134	-0,073	-0,0026	-0,2225	-0,1283	-0,0444	0,0181	0,0161	-	114.44K	1,38	136,72	-0,024	158211	
	Vidurkis	-0,03273														
	top 5	-0,01012														
	stdev	0,018345														
1	AAPL	0,0265	0,1169	0,1875	-0,1251	0,1155	0,2112	0,0254	0,0168	2,1	32.79M	1,29	191,05	-0,0207	42246234	C3
2	EL	-0,0199	0,019	0,2409	0,1421	0,0915	0,226	0,0181	0,0179	2,1	2.15M	0,48	159,5	-0,0142	1022541	
3	FIZZ	-0,0326	-0,2253	-0,243	-0,4869	-0,3382	-0,2101	0,025	0,0411	3,8	439.75K	0,48	56,69	-0,0077	212474	

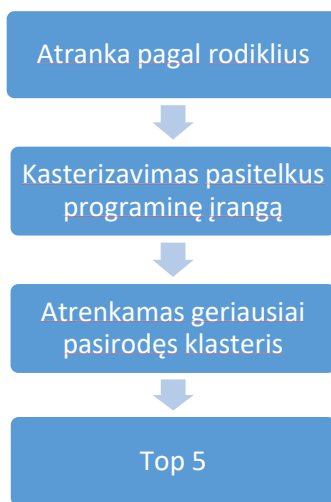
28 pav. MS Excel faile išsaugotų duomenų apie klasterių vidutinį pelningumą pavyzdys

Vėliau patikrinamas Kiekvieno sektoriaus klasterių (kuriuose nemažiau nei 5 įmonės) vidutinis pelningumas

28 pav. pateiktas MS Excel failo pavyzdys, kuriame patikrinama, kurio iš trijų klasterių savaitinis vidutinis įmonių pelningumas didžiausias pasitelkus MS Excel vidurkio formulę (angl. *average*).. Pastebėta, jog suklasterizavus sektorius didžiausias savaitinis vidutinis pelningumas niekada nebūna viename iš klasterių. Tai yra, jei vieną savaitę, pirmajame klasteryje C1 esančios įmonės pasirodė gerai, kitą savaitę geriau gali pasirodyti C2 klasteryje esančios įmonės. Todėl kiekvieną kartą atlikus įmonių klasterizavimą patikrinamas pelningumas ir atrenkamas geriausias klasteris su pelningiausiomis įmonėmis. Beje, klasteriai, kuriuose yra mažiau nei 5-ios įmonės netikrinami, nes portfelis bus sudaromas iš pastarosios savaitės pelningiausio klasterio 5-ių pelningiausių įmonių. Laikoma, jog portfelis turi mažesnę riziką, jei jo įmonių vertybinių popierių vidutinio įmonių pelningumo standartinis nuokrypis yra mažas. Būtent tai padeda pasiekti klasterizavimas, nes į vieną klasterį sutraukiamos įmonės su panašių verčių rodikliais, tuo tarpu ir mažesnis skirtumas yra tarp skirtinguose klasteriuose esančių įmonių savaitinio vidutinio įmonių pelningumo. Tai yra mažesnis standartinis nuokrypis. Taip pat mažesnis standartinis nuokrypis turi būti tam tikras optimalus įmonių skaičius portfelyje.



29 pav. Atitinkamai pavaizduoti: plataus vartojimo prekių, pramoninių prekių, paslaugų, sveikatos ir technologijų sektorių žemėlapiai, ir jų klasteriai.



30 pav. Pasirinktas klasterizavimo algoritmas

30 pav. pavaizduotas pasirinktas klasterizavimo algoritmas parodo įmonių, į kurias ruošiamasi investuoti atrankos procesą. Visų pirma pasitelkiama finviz.com platforma, kurioje nustatomos finansinių rodiklių vertės ir kiekvienam sektoriuj. Pagal nustatytus rodiklius atrinkama apie 20-30 įmonių kiekvienam sektoriui. Antrame etape pasitelkiama „Viscovery SOMine“ programinė įranga. Įmonės suklasterizuojamos į tris klasterius. Trečiame etape pamatuojamas kiekvieno klasterio vidutinis pelningumas. Ketvirtame etape iš geriausiai pasirodžiusio klasterio atrinkamos penkios didžiausią pelningumą turinčios įmonės.

Atrinkus perspektyviausią praeitos savaitės klasterį pagal jo praeitos savaitės vidutinį pelningumą. Iš šio klasterio yra atrinkami penki vertybiniai popieriai, kurie pastarąją savaitę turėjo didžiausią pelningumą. 21 pav. matome MS Excel failo pavyzdį, kuriame atrinkamos 5 pastarosios savaitės įmonės iš perspektyviausio klasterio. Yra tariama, jog būtent į šias penkias įmones ir yra investuojama paskesnę savaitę. Iš jų sudaromas portfelis. Po savaitės vėl patikrinamas šių įmonių gautas pelnas. Tai pelnas kurį gauname.

Žinoma, po savaitės portfelio sudėtis gali keistis. Tai priklausys, kokios įmonės pateks tarp penkių tos savaitės perspektyviausių, pasitelkus atrankos algoritimą.

BC	-0,0538								Ateitis	Istorija	
CAL	-0,1162								kovo 25	kovo 18	
CSL	-0,0282					Top 5	Vertė	Žymuo	Close 18 adj.	Close 17 adj.	
DECK	-0,0089					1	0,0487	MNRO	86,52	79	8,691632
ELY	-0,0574					2	0,0197	FLO	21,32	20,75	2,673546
FLO	0,0197					3	0,0084	JCI	36,94	36,13	2,192745
GEF	-0,0131					4	0,0034	SON	61,53	59,3	3,624248
HNI	-0,0694					5	-0,0089	DECK	146,99	140,51	4,408463
JCI	0,0084								Pelnas	vidurkis %	4,318127
MLR	-0,0472									St. nuokny	2,590322
MNRO	0,0487										
OSK	-0,0381										
PKG	-0,0109										
SHOO	-0,0123										
SON	0,0034										
SRI	-0,026										
THO	-0,0913										
UFS	-0,0172										

31 pav. Penkių pelningiausių įmonių atrankos procesas

Atrinkus 5 perspektyviausias įmonių vertybinius popierius po savaitės yra patikrinamas šių vertybinių popierių gautas savaitinis pelnas procentais 32 pav. Duomenys vidurkiui apskaičiuoti gaunami pasitelkiant finance.yahoo.com platformos istorinių duomenų filtrą 33 pav. Savaitinis pelnas procentais apskaičiuojamas taip:

$$\frac{(\text{Praėjusios sav. vertybinio popieriaus kaina} - \text{Einamos sav. vertybinio popieriaus kaina}) * 100}{\text{Einamos sav. vertybinio popieriaus kaina}}$$

Vėliau apskaičiuojamas visas tos savaitės portfelio, penkių vertybinių popierių pelno savaitinis vidurkis. Taip gaunamas vieno portfelio pelno pokyčio kreivės vienas taškas (vienos savaitės). Toks duomenų apdorojimas atliekamas visiems penkiems sektoriams.

			Ateitis	Istorija	
			kovo 25	kovo 18	
Top 5	Vertė	Žymuo	Close 18 adj.	Close 17 adj.	
1	0,0487	MNRO	86,52	79	$=((J48-K48)/J48)*100$
2	0,0197	FLO	21,32	20,75	2,67355
3	0,0084	JCI	36,94	36,13	2,19274
4	0,0034	SON	61,53	59,3	3,62425
5	-0,0089	DECK	146,99	140,51	4,40846
			Pelno vidurkis %		4,31813
				St. nuokry	2,59032

32 pav. MS excel skaičiuoklėje apskaičiuojamas savaitinis viso portfelio pelnas

Time Period: May 16, 2018 - Apr 01, 2019 Show: Historical Prices Frequency: Weekly Apply

Currency in USD [Download Data](#)

Date	Open	High	Low	Close*	Adj Close**	Volume
Mar 25, 2019	79.16	86.84	78.43	86.52	86.52	1,632,200
Mar 18, 2019	75.34	83.40	75.34	79.00	79.00	1,589,000
Mar 14, 2019	0.2 Dividend					
Mar 11, 2019	78.10	79.93	75.01	75.33	75.14	1,428,100

33 pav. finance.yahoo.com platformoje gaunamos istorinės vertybinių popierių kainos savaitiniam pelnui apskaičiuoti

3. Pasiūlyto klasterizavimo algoritmo realizavimas ir testavimas

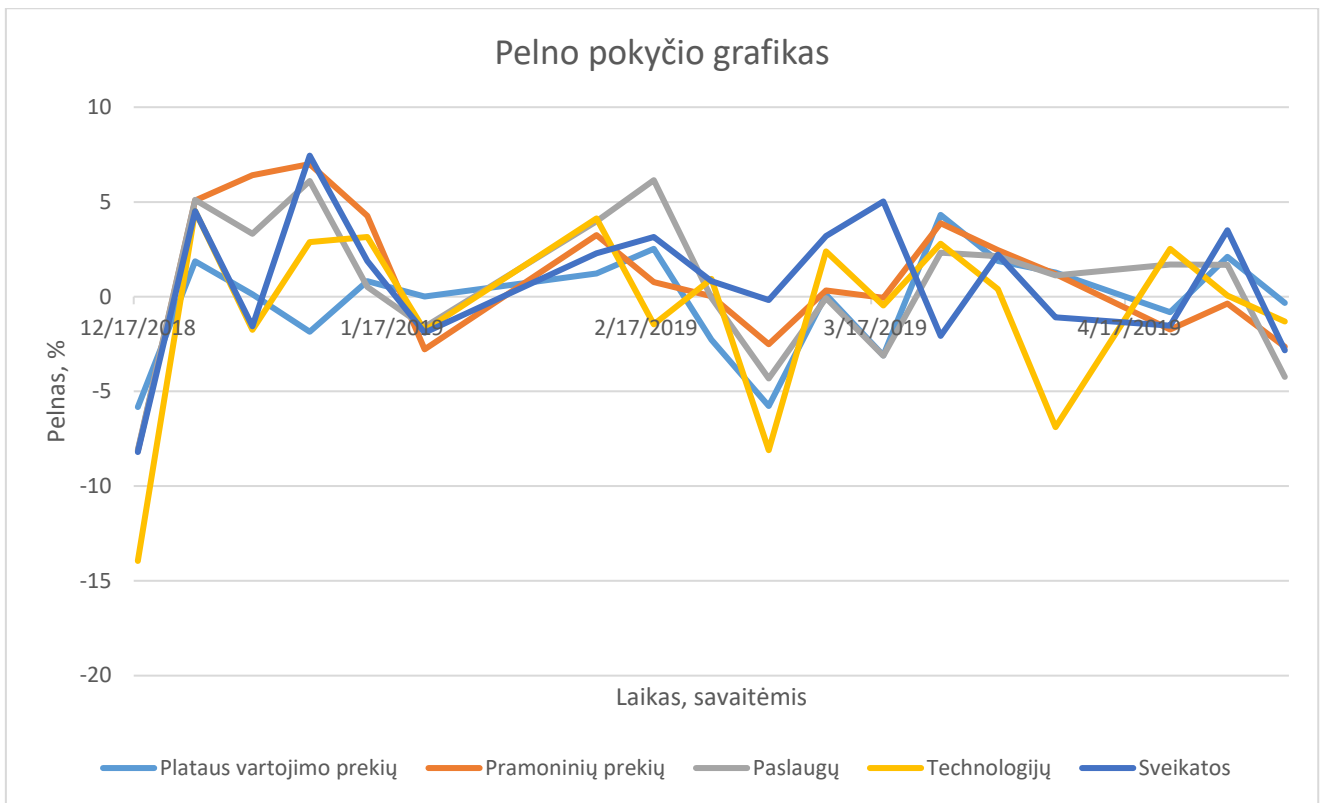
Atsižvelgiant į sudarytą klasterizavimo grafiką vaizduojamą 34 pav., ir anksčiau minėtą klasterizavimo algoritmą 30 pav. pažymėtomis savaitės dienomis pagal 7 rodiklius buvo atrenkama nuo 20 iki 30 kompanijų vertybinių popierių iš kiekvieno sektoriaus. Tuomet atlikus klasterizavimą kiekvieno sektoriaus įmonės buvo atrenkamos iš geriausiai pasirodžiusio klasterio. Geriausiai pasirodęs klasteris yra tas, kurio yra didžiausias vidutinis pelningumas. Taip pat buvo įvesta taisyklė, jog klasteryje negali būti mažiau nei penkios įmonės, nes portfelis iš tokio klasterio būtų pernelyg mažas. Vėliau, iš geriausiai pasirodžiusio klasterio buvo atrenkamos penkios didžiausių pelningumą turinčios kompanijos iš sektoriaus geriausiai pasirodžiusio klasterio.

Gruodis	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
																S								S						S	
Sausis	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
					S							S																			
Vasaris	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28			
									S								S							S							
Kovas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
			S							S							S							S						S	
Balandis	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
						S							S							S								S			
Gegužė	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
				S							S							S							S						

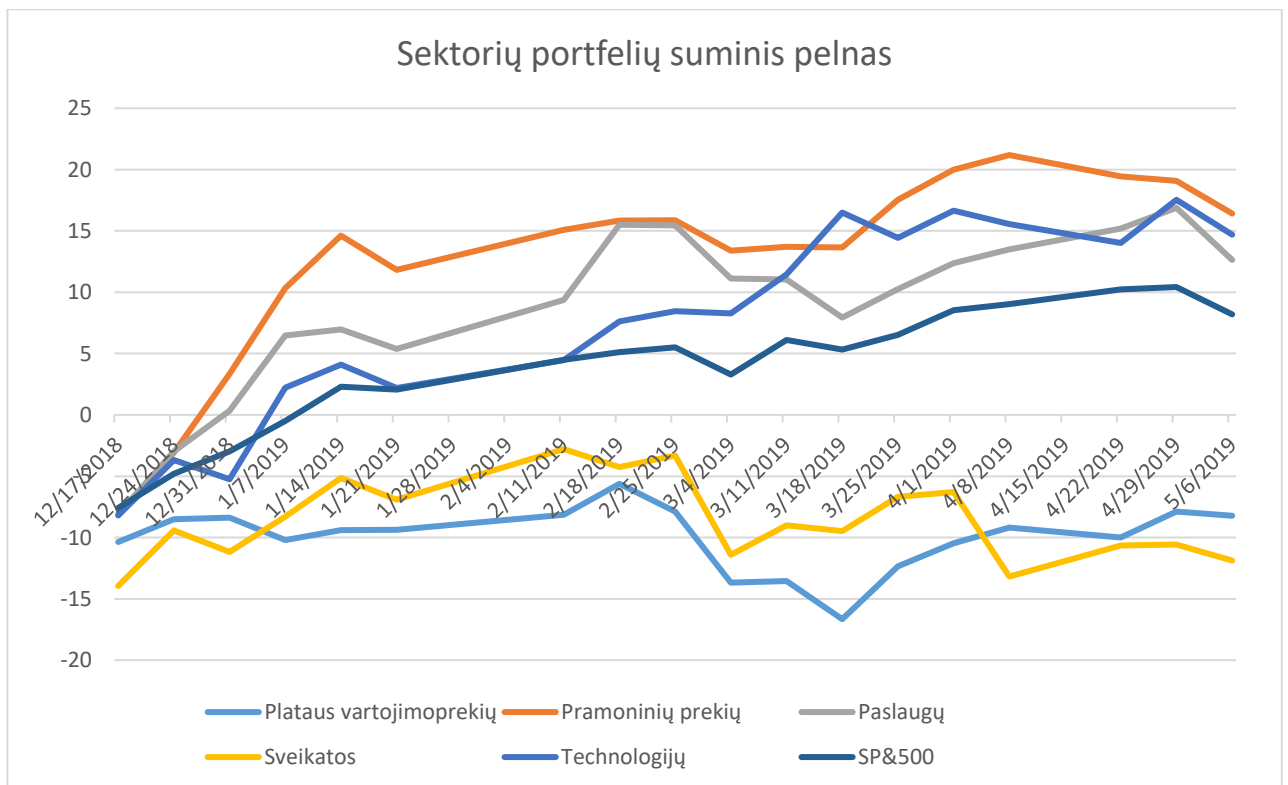
34 pav. Akcijų pelningumo tikrinimo grafikas (2018-2019 m.)

Efektyviausios, didžiausių vidutinį pelningumą turinčios įmonės buvo atrenkamos akcijų portfelių keitimo grafike (34 pav.) pažymėtomis savaitėmis kiekvienam sektoriui ir taip sudaroma vertybinių popierių pelningumo pokyčio grafikas 35 pav. Pastarajame grafike, vaizduojama atskirai kiekvieno sektoriaus didžiausių pelningumą turinčių įmonių portfelių, laike kintantis vidutinį pelningumas procentais.

36 pav. Vaizduojamas suminio pelningumo grafikas. Iš jo galima pastebėti, jog trijų sektorių portfeliams (pramoninių prekių – morkinė linija, paslaugų – pilka linija ir technologijų) suminiu pelningumu pavyko aplenkti SP&500 indeksą. Portfeliams sudarytiems iš kitų dviejų sektorių (sveikatos ir plataus vartojimo prekių) įmonių vertybinių popierių aplenkti SP&500 indekso nepavyko.



35 Pav. Pagal pasirinktą klasterizavimo algoritmą suformuotų portfelių vidutinio pelningumo pokyčio grafikas



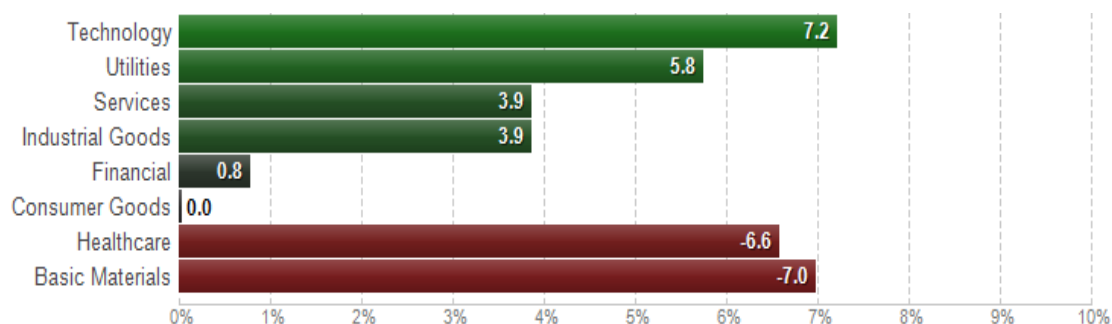
36 pav. Pagal pasirinktą klasterizavimo algoritmą sudarytų portfelių suminiai pelnai.

3.1. Rezultatų aptarimas

Lentelėje 1 rezultatai rodo, jog didžiausią pelningumą turėjo pramoninių prekių sektoriaus portfelis su galutiniu 16.428 % pelningumu. Antroje vietoje pagal pelningumą atsidūrė technologijų sektoriaus portfelis 14.705%, trečioje – paslaugų sektoriaus portfelis, kurio galutinė grąža 12.644%. Visi pastarieji trys investiciniai portfeliai aplenkė SP&500 indeksą, kurio portfolio galutinė grąža 8.193%. Sveikatos ir plataus vartojimo prekių sektorių portfeliams nepavyko aplenkėti SP&500 indekso. Šie du portfeliai patyrė nuostolį. Sveikatos sektoriaus portfelis -11.881% ir plataus vartojimo prekių portfelis -8.221%. Nenuostabu, atsižvelgus į tai, jog finviz.com portale statistiniai rezultatai rodo, jog per šį klasterizavimo laikotarpį 12/17/2018 - 5/6/2019 sveikatos ir plataus vartojimo prekių sektoriai patyrė bendrą ekonominę nuosmukį 37 pav.

1 lentelė. Portfelių suminio pelno kitimas.

Data	Plataus vartojimo prekių	Pramoninių	Paslaugų	Sveikatos	Technologijų	SP&500
12/17/2018	-10.364	-8.150	-8.089	-13.947	-8.200	-7.586
12/24/2018	-8.498	-3.067	-2.976	-9.426	-3.679	-4.806
12/31/2018	-8.368	3.341	0.349	-11.177	-5.238	-2.981
1/7/2019	-10.215	10.339	6.461	-8.290	2.207	-0.503
1/14/2019	-9.388	14.608	6.979	-5.136	4.094	2.284
1/21/2019	-9.377	11.830	5.380	-6.921	2.187	2.061
2/11/2019	-8.150	15.079	9.363	-2.789	4.474	4.501
2/18/2019	-5.620	15.852	15.512	-4.245	7.630	5.112
2/25/2019	-7.894	15.879	15.450	-3.305	8.454	5.505
3/4/2019	-13.664	13.376	11.129	-11.403	8.273	3.295
3/11/2019	-13.553	13.704	11.048	-9.013	11.467	6.109
3/18/2019	-16.654	13.648	7.928	-9.480	16.497	5.331
3/25/2019	-12.336	17.530	10.245	-6.676	14.431	6.520
4/1/2019	-10.458	19.991	12.374	-6.284	16.648	8.537
4/8/2019	-9.184	21.180	13.497	-13.171	15.555	9.041
4/22/2019	-9.996	19.446	15.198	-10.640	14.023	10.227
4/29/2019	-7.885	19.075	16.882	-10.578	17.536	10.422
5/6/2019	-8.221	16.428	12.644	-11.881	14.705	8.193



37 pav. Vaizduojamas bendras rinkoje esančių sektorių pelningumas (2018.12 – 2019.05)

3.1.1. Portfelio rizikos įvertinimas

Atrodytų, jog pelningas portfelis yra geras portfelis. Tačiau ne visada, nes pelningumas procentais priklauso nuo rizikos ir vien tik jis neparodo investavimo sistemos efektyvumo. Taigi reikia įvertinti šio ir portfelio riziką. Būtent tai padaryti padeda Šarpo rodiklis (angl. *Sharpe ratio*), kurį išvedė Nobelio premijos laureatas Vilijam 'as Sharp' as. Šarpo rodiklis apskaičiuojamas iš tikėtino vidutinio pelningumo atėmus pelningumą be rizikos ir gautą rezultatą padalinus iš standartinio nuokrypio [31]

$$\text{Šarpo rodiklis} = (\text{Laukiama portfelio grąža} * 52 - \text{Nerizikinga palūkanų norma}) / \text{Portfelio standartinis nuokrypis} * \sqrt{52}$$

Arba:

$$\text{Šarpo rodiklis} = \frac{R_p * 52 - R_f}{\sigma_p * \sqrt{52}};$$

Čia R_p – laukiama portfelio grąža; R_f – nerizikinga palūkanų norma; σ_p – standartinis nuokrypis.

Pastarojoje formulėje, nerizikinga palūkanų norma (angl. *risk free ratio*) tai koeficientas (skaičius) parodantis palūkanų normą, kurią užtikrintų nerizikingi skolos vertybiniai popieriai. Nors šio koeficiento terminas tik teorinis (visiškos rizikos nebuvimo negalima pasiekti), tačiau koeficientas plačiai naudojamas praktikoje. Jis apskaičiuojamas atsižvelgus į saugiomis laikomų valstybių (pavyzdžiui, JAV ar Vokietijos, priklausomai nuo to, kokia valiuta norma nustatoma) ilgalaikių obligacijų mokamos palūkanas. Šiomis dienomis nerizikinga palūkanų normos koeficientas yra lygus 1.75. [31] [32]

Iš rizikingesnių įmonių į kurias investuojama tikimasi didesnio pelno. Sharp 'o rodiklis padeda išmatuoti pelno ir rizikos santykį. Rizikingumas arba kintamumas (angl. *volatility*) apskaičiuojamas standartinio nuokrypio pagalba. Standartinis nuokrypis rodo investavimo rezultato, reikšmių pasiskirstymą apie vidurkį. Kuo standartinis nuokrypis mažesnis tuo geriau, nes efektyviau prognozuojami investicijų rezultatai. Esant dideliame kintamumui rizika didėja, nes investicijų rezultatai prognozuojami sunkiai.

Šarpo rodiklis padeda įvertinti ar į portfelį verta investuoti. Pavyzdžiui, jei į portfelį yra įtraukiama nauja kompanija. Perskaičiuotas Šarpo rodiklis parodys ar į sudarytas naujas portfelis duoda pakankamą grąžą ir ar rizika, kurios imamas nėra per didelė.

Kuo Šarpo koeficientas didesnis, tuo investicija neša didesnę pelną tenkantį rizikos vienetui. Obligacijų fondų Šarpo rodiklis dažnai kinta tarp 0.6 iki 0.8, akcijų fondų svyruoja tarp 0.1-0.8. Kita vertus, įvairūs ribotos rizikos fondai taikydami rizikos mažinimo priemones pasiekia, kad Šarpo rodiklis būtų didesnis nei 1. Įprastai laikoma, jog Šarpo rodiklis:

- Didesnis nei 1 yra patenkinamas arba geras.
- Didesnis nei 2 laikomas labai geru.
- Didesnis nei 3 laikomas puikiu,
- Mažesnis už 1 laikomas neoptimaliu. [32] [33] [34]

2 lentelėje pateikiamas SP&500 indekso ir kiekvieno sektoriaus portfelio Šarpo rodiklio vertės. Galima pastebėti, jog visi trys portfeliai (pramoninių prekių, paslaugų, technologijų), kurie davė didesnę pelną nei SP&500 indeksas buvo mažiau rizikingi nei SP&500 portfelis. Žinoma, du portfeliai, kurie patyrė nuostolį (plataus vartojimo prekių ir sveikatos) buvo rizikingi, todėl jų Šarpo rodiklio nagrinėti net neverta.

2 lentelė. SP&500 indekso ir pelningų portfelių Šarpo rodikliai.

Portfelis	SP&500	Pramoninių prekių	Paslaugų	Technologijų
	0,455	0,913	0,702	0,817
Standartinis nuokrypis	2,554	3,784	3,842	3,640
ŠARPO RODIKLIS	1,193	1,677	1,257	1,554

Rezultatai ir išvados

1. Atlikta skaitinio intelekto taikymo vertybinių popierių atrankai literatūros analizė. Nustatyta, kad savitvarkiai neuroniniai tinklai (SOM) yra potencialus įrankis kompanijų vertybinių popierių atrankai.
2. Pasiūlyta schema kompanijų vertybinių popierių portfelio formavimui, pasitelkiant savitvarkius neuroninius tinklus. Tyrimui pasirinkta tinkama programinė įranga SOM realizavimui.
3. Atliktas pelningumo analizės 20 savaitių tyrimas 5-iems sektoriams. Rezultatai rodo, trijų sektorių portfelių suminis pelnas procentais aplenkia SP&500 indekso suminį pelną 8.193%. Pramoninių prekių sektoriaus portfelio suminis pelnas 16.428%. Paslaugų sektoriaus suminis pelnas 12.644%. Technologijų sektoriaus portfelio 14.705% . Plataus vartojimo prekių ir sveikatos sektorių portfeliai patyrė nuostolį.
4. Darbe atliktas investicijų rizikingumo vertinimas pagal Šarpo rodiklį parodė, kad trijų pelningų portfelių rizika yra mažesnė nei SP&500 portfelio 1.193. Pramoninių prekių Šarpo koeficientas 1.677, paslaugų sektoriaus portfelio 1.257, technologijų sektoriaus portfelio 1.554.
5. Sukurtas įmonių atrankos algoritmas pasitelkiant SOM gali būti rekomenduojamas, kaip įrankis investuotojams. Vis dėl to, reikalinga ilgesnė ir nuodugnesnė šio metodo analizė.

Literatūros sąrašas

1. *Artificial intelligence and machine learning in financial services. Market developments and financial stability implications* [interaktyvus]. 2017 [žiūrėta 2019-04-20]. Prieiga per: <http://www.fsb.org/wp-content/uploads/P011117.pdf>
2. ZAVADSKAYA, Alexandra. Artificial Intelligence in Finance: Forecasting Stock Market Returns Using Artificial Neural Networks. *Ekonomikos mokslai* [interaktyvus]. Helsinki, Henken School of Economics, 2017 [žiūrėta 2019-05-12]. Prieiga per: <https://helda.helsinki.fi/dhanken/bitstream/handle/123456789/170154/zavadskaya.pdf?sequence=1>
3. MYERS, David. *Psichologija*. 2008. Kaunas: Poligrafija ir informatika. ISBN 978-9986-850-61-8
4. BERNOTAS, Edvinas. Finansinių ir fizinių procesų prognozavimo modelių tyrimas ir taikymas. [interaktyvi]. Kaunas, Kauno technologijos universitetas, 2009., Prieiga per: <https://core.ac.uk/download/pdf/51686277.pdf>
5. VERIKAS, Antanas ir GELŽINIS, Adas. (2008) *Neuroniniai tinklai ir neuroniniai skaičiavimai*. Kaunas: Kauno technologijos universitetas. ISBN 978-9955-591-53-5.
6. ANDRIUŠKA, Julius. Dirbtinio intelekto metodų taikymo automatiniam prekybos robotui galimybių tyrimas [atmintukas]. Kaunas, Kauno technologijos universitetas, 2016 [2019-05-17]
7. ZHEREBTSOV, A. ir KUPERIN Yu.A. Application of Self-Organizing Maps for clustering DJIA and NASDAQ100 portfolios. *Division of Computational Physics* [interaktyvus]. St.Petersburg 198904,Russia,2003.Prieigaper:https://www.researchgate.net/publication/1936071_Application_of_Self-Organizing_Maps_for_clustering_DJIA_and_NASDAQ100_portfolios
8. SMITH, Kate ir Jatinder GUPTA. *Neural Networks in Business: Techniques and Applications*. JAV ir Didžioji Britanija: Idea Group Publishing, 2002. ISBN 1-930708-31-9.
9. KORZENIEWSKI, Jerzy. Efficient Stock Portfolio Construction by Means of Clustering. *Folia Oeconomica*. Lodziensis: Acta Universitatis Lodziensis, 2018. 85-92. ISSN 0208-6018 e-ISSN 2353-7663. Prieiga per: DOI: <http://dx.doi.org/10.18778/0208-6018.333.06>
10. EKLUND, Tomas ir kiti. Using the self-organizing map as a visualization tool in financial benchmarking. *Information Visualization* [interaktyvus]. Turku, Suomija: Turku Centre for Computer Science. 2003, 171–181. Prieiga per: doi:10.1057/palgrave.ivs.9500048
11. JOSEPH, Joel ir INDRATEMO. Visualizing Stock Market Data with Self-Organizing Map. *Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference*. Association for the Advancement of Artificial Intelligence. Department of Computer Science, Grant MacEwan University. 2013. Prieiga per: https://www.researchgate.net/publication/286741983_Visualizing_stock_market_data_with_self-organizing_map
12. EKLUND, Tomas ir Hannu VANHARANTA. A Face Validation of a SOM-Based Financial Benchmarking Model. *Journal of Emerging Technologies in Accounting*. Turku, Suomija. 2008. Prieiga per: DOI: 10.2308/jeta.2008.5.1.109
13. DEBOECK, Guido J. Financial Applications of Self-Organizing Maps. *Neural Network World*. 2000.

- Prieigaper:<https://www.researchgate.net/publication/2321799> Financial Applications of Self-Organizing Maps [Žiūrėta 2019-05-17]
14. MARVIN, Karina. Creating Diversified Portfolios Using Cluster Analysis. *Independent Work Report Fall*, 2015. [žiūrėta 2019-05-18]
Prieiga per: https://www.cs.princeton.edu/sites/default/files/uploads/karina_marvin.pdf
 15. SILVA, Bruno ir Nuno C. MARQUES. *Feature Clustering with Self-organizing Maps and an Application to Financial Time-series for Portfolio Selection*. ICNC 2010 – International Conference on Neural Computation. 2010, 301-309. 2010. Prieiga per: <https://www.researchgate.net/publication/221616457> Feature Clustering with Self-organizing Maps and an Application to Financial Time-series for Portfolio Selection
 16. The new approach to VISCOVER your data. [interaktyvi]. [žiūrėta 2019-05-19] <https://www.viscovery.net> . Prieiga per: <https://www.viscovery.net/somine/>
 17. Visual data mining with self-organizing maps (SOM). [interaktyvi]. [žiūrėta 2019-05-19] <https://www.viscovery.net>. Prieiga per: <https://www.viscovery.net/visual-clusters>
 18. ŽEMAITYTĖ, Raminta. *Savitvarkių neuroninių tinklų taikymas vertybinių popierių klaserizavimui*. [atmintukas]. Kaunas, Vilniaus universitetas, 2011. [2019-05-19]
 19. WALKER, Blake. Finviz Review: Are They As Good As They Want You To Believe? [interaktyvi]. 2017 [žiūrėta 2019-05-19] Prieiga per: <https://tradingreview.net/finviz-review/>
 20. Earnings per Share Growth.[interaktyvi][žiūrėta 2019-05-19].
Prieiga per: <https://www.stockopedia.com/ratios/earnings-per-share-growth-last-year-587/>
 21. SIMUTIS, Rimvydas. *Paskaitų „Valdymo metodai finansiniuose procesuose“ medžiaga*, [2016 m.];
 22. What Is Earnings Per Share – EPS? [interaktyvi]. 2019 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/terms/e/eps.asp>
 23. Return on Equity – ROE [interaktyvi]. 2019 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/terms/r/returnonequity.asp>
 24. Price-to-Earnings Ratio (P/E Ratio) Definition [interaktyvi]. 2019 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/terms/p/price-earningsratio.asp>
 25. Finansiniai rodikliai. [interaktyvi]. [žiūrėta 2019-05-19]. Prieiga per: <https://www.finansistas.net/finansiniai-rodikliai.html>
 26. Long-Term Debt. [interaktyvi]. 2017 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/terms/l/longtermdebt.asp>
 27. Long Term Debt to Equity [interaktyvi]. [žiūrėta 2019-05-19]. Prieiga per: <https://www.stockopedia.com/ratios/long-term-debt-to-equity-468/>
 28. Current Ratio Definition [interaktyvi]. 2019 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/terms/p/price-earningsratio.asp>
 29. How to use insider and institutional ownership [interaktyvi]. 2018 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/articles/stocks/05/042605.asp>
 30. Institutional Ownership [interaktyvi]. 2018 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/terms/i/institutional-ownership.asp>
 31. Šarpo rodiklis [interaktyvi].[žiūrėta 2019-05-19]. Prieiga per: <https://www.finansistas.net/sharp-koeficientas.html>

32. RUTKAUSKAS, Gintaras. Investicijų grąža ir rizika – ką reikia žinoti? 2016 [interaktyvi]. [žiūrėta 2019-05-19]. Prieiga per: <http://www.investavimas.lt/nuo-ko-priklauso-investiciju-graza/>
33. Sharpe Ratio Definition [interaktyvi]. 2019 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/terms/s/sharperatio.asp>
34. What Is a Good Sharpe Ratio? [interaktyvi]. 2019 [žiūrėta 2019-05-19]. Prieiga per: <https://www.investopedia.com/ask/answers/010815/what-good-sharpe-ratio.asp>