

MYKOLO ROMERIO UNIVERSITETAS
VIEŠOJO VALDYMO IR VERSLO FAKULTETAS
Verslo ir ekonomikos institutas

VALERIJA MIRIBIAN

**DIRBTINIŲ INTELEKTU GRĮSTŲ METODŲ TAIKYMO AKCIJŲ
KAINŲ PROGNOZEI TYRIMAS**

MAGISTRO BAIGIAMASIS DARBAS

Vadovas:
Doc. Dr. Inga Žilinskienė

VILNIUS, 2020 m.

TURINYS

1.	ĮVADAS.....	6
2.	SANTRUMPŲ SĄRAŠAS	9
3.	ŽODYNĖLIS	10
4.	DIRBTINIO INTELEKTO SAMPRATA BEI VAIDMUO AKCIJŲ RINKŲ KAINŲ PROGNOZĖJE	12
4.1.	PAGRINDINĖS SĄVOKOS, YPATUMAI BEI PRITAIKOMUMAS	12
4.1.1.	Dirbtinis intelektas - kas tai?.....	12
4.1.2.	Mašininis mokymasis bei jo atmainos	14
4.1.3.	Vaidmuo finansų rinkose	19
4.2.	DIRBTINIO INTELEKTO METODAI AKCIJŲ KAINŲ PROGNOZEI	21
4.2.1.	Mašininio mokymosi metodų apžvalga	21
4.2.2.	Mašininio mokymosi modelių modeliai bei jų klasifikacija.....	22
4.2.3.	Prognozavimu grįstų technikų apžvalga	26
4.2.4.	Klasifikacija grįstų technikų apžvalga	31
4.3.	DIRBTINIO INTELEKTO METODŲ TAIKYMO AKCIJŲ KAINAI PROGNOZEI YPATUMAI.....	34
4.3.1.	Dirbtiniai Neuroniniai Tinklai (ANN) bei modelio ypatumai	35
4.3.2.	Atraminių Vektorių Mašina (SVM) bei modelio ypatumai	40
5.	TYRIMO METODOLOGIJA.....	45
5.1.	TYRIMO STRATEGIJA	45
5.2.	TYRIMO EIGA	48
5.2.1.	Dirbtiniai neuroniniai tinklai (ANN)	48
5.2.2.	Atraminių vektorių mašinos (SVM)	53
6.	EMPIRINIS TYRIMAS IR JO REZULTATAI.....	59
6.1.	ANN MODELIU GRĪSTAS S&P 500 INDEKSO PROGNOZAVIMAS	59
6.1.1.	Duomenų surinkimas ir apdorojimas	59
6.1.2.	ANN Modelio mokymasis ir tobulinimas.....	60
6.2.	SVM MODELIU GRĪSTAS S&P 500 INDEKSO PROGNOZAVIMAS	65
6.2.1.	Duomenų surinkimas ir apdorojimas	65
6.2.2.	SVM Modelių mokymasis ir tobulinimas.....	65
6.3.	TYRIMO REZULTATAI.....	68
7.	IŠVADOS IR PASIŪLYMAI	70
8.	LITERATŪROS SĄRAŠAS.....	72
9.	SANTRAUKA.....	75
10.	PRIEDAI	77

Paveikslų sąrašas

<i>Pav. 1</i> Dirbtinio intelekto tiriamos sritys	13
<i>Pav. 2</i> Akcijų kainų prognozavimo technikos funkcionalumas	24
<i>Pav. 3</i> Akcijų prognozavimo technikų klasifikacija	25
<i>Pav. 4</i> Bendrinė ANN modelio veikimo schema	26
<i>Pav. 5</i> Bendrinė HMM modelio veikimo schema	28
<i>Pav. 6</i> Bendrinė RNN modelio veikimo schema	29
<i>Pav. 7</i> Bendrinė SVM modelio veikimo schema	30
<i>Pav. 8</i> Dirbtiniai Neuroniniai tinklai (ANN)	36
<i>Pav. 9</i> Atraminių Vektorių Mašina (SVM)	41
<i>Pav. 10</i> Penki pagrindiniai ML proceso žingsniai	45
<i>Pav. 11</i> Tyrimo strategija	46
<i>Pav. 12</i> Šio mokslinio darbo detalus tyrimo modelis	47
<i>Pav. 13</i> ANN suformuoto modelio architektūra	51
<i>Pav. 14</i> NARX modelis ANN duomenimis klasifikuoti	60
<i>Pav. 15</i> ANN išvesties elemento atsakas	61
<i>Pav. 16</i> ANN modelio geriausias mokymosi atlikimas	62
<i>Pav. 17</i> ANN patobulinto suformuoto modelio architektūra	62
<i>Pav. 18</i> ANN patobulinto modelio geriausias mokymosi atlikimas	63
<i>Pav. 19</i> ANN modelio išvesties elemento atsakas	64
<i>Pav. 20</i> SVM optimizuoto modelio Numatyty/Faktinių reikšmių grafikas	67
<i>Pav. 21</i> SVM Optimizuoto modelio Atsako Reikšmių grafikas	68

Lentelių sąrašas

<i>1 lentelė.</i> Dirbtinio intelekto (AI) kategorijos.....	12
<i>2 lentelė.</i> Mašinų mokymasis: tipai bei paaiškinimai	15
<i>3 lentelė.</i> Akcijų kainų prognozavimo technikos	25
<i>4 lentelė.</i> ANN modeliui naudojamų kintamųjų sąrašas	48
<i>5 lentelė.</i> ANN kintamųjų grupavimas.....	49
<i>6 lentelė.</i> ANN modelio mokymuisi įvertinti naudojami parametrai.....	52
<i>7 lentelė.</i> SVM modeliui naudoti kintamieji	54
<i>8 lentelė.</i> SVM modeliams taikyti regresijos tipai bei jų ypatybės.....	56
<i>9 lentelė.</i> SVM modelio įvertinimui naudojami parametrai.....	57
<i>10 lentelė.</i> Pirminių duomenų kintamųjų lentelė (ANN)	59
<i>11 lentelė.</i> ANN mokymosi rezultatai.....	60
<i>12 lentelė.</i> ANN patobulinto modelio rezultatai	62
<i>13 lentelė.</i> S&P uždarymo kainų kintamojo aprašomoji statistika.....	65
<i>14 lentelė.</i> SVM šešių pirminių modelių rezultatai.....	66

Priedų sąrašas

1. SVM modelių atsako reikšmių grafikai.....	77
2. SVM modelių Numatytu/Faktinių reikšmių grafikai.....	78
3. SVM modelių atsako reikšmių grafikai.....	79
4. Formulų sąrašas.....	80
5. Patvirtinimas apie darbo savarankiškumą.....	81

1. Įvadas

Temos aktualumas

Akcijų graža arba akcijų rinkos prognozė šiuo metu yra labai svarbi finansų pasaulyje nagrinėjama tema, kuri jau daugelį metų traukia tyrėjų dėmesį. Dėl to galima daryti prielaidą, kad finansų rinkose viešai prieinama informacija turi tam tikrą numatomą ryšį su būsimomis akcijų gražomis. Tokios informacijos pavyzdžiai apima ekonominius kintamuosius, tokius kaip palūkanų normas ir valiutų kursas, specifinę pramonės informaciją, tokią kaip pramonės produkcijos augimo tempus ir vartotojų kainas, taip pat konkrečią informaciją apie įmonę, tokią kaip pelno (nuostolių) ataskaitas bei pagrindinius finansinius rodiklius. Kita vertus, bandymas numatyti akcijų gražą prieštarauja bendram rinkos efektyvumo suvokimui.

Šiuo metu bandomi įvairūs rinkos vertės nustatymo bei kainų prognozės modeliai, dauguma jų paremti dirbtiniu intelektu, tokie kaip neuroniniai tinklai, atraminiai vektoriai bei hibridiniai modeliai. Stebint dabartinę akcijų vertę rinkoje, siekiama kad visuomenė, konkrečiau investuotojai, galėtų sudaryti sandorius pagal tam tikrus kriterijus ir loginius dėsningumus. Rinkos efektyvumo teorija teigia, jog iš esmės neįmanoma prognozuoti būsimų akcijų kainų bei jų svyravimų, nes jos atspindi viską, kas šiuo metu žinoma ir prieinama apie turtą. Taip pat manoma, kad efektyvi rinka akimirksniu pakoreguos akcijų kainas, remdamasi naujienomis, kurios į rinką pateks atsitiktinai. Panašūs samprotavimai palaiko vadinamojo „atsitiktinio pasivaikščiojimo“ modelio pagrindimą, pagal kurį geriausia prognozuoti kito laikotarpio akcijų kainą yra skaičiuojant dabartinę turto vertę. Be to, vis dar tirama bei pateikiama įrodymų, kad akcijų rinkos nėra efektyvios, tačiau vis dėlto galima numatyti būsimas akcijų kainas ar indeksus naujais metodais, kurių rezultatai bus vis tikslesni.

Temos iširtumas

Šiuo metu sparčiai populiarėjant informacinėms technologijoms, dirbtinio intelekto samprata bei jo pritaikomumas tiriamas beveik visose mokslinėse srityse. Dirbtinio intelekto panaudojimas finansų rinkose pradėtas tirti prieš kelis dešimtmečius, bandant kurti bei pritaikyti paprastus algoritmus, iš dalies paremtus technine analize, tačiau pastaruoju dešimtmečiu (maždaug nuo 2008 metų) ši tema tapo vis labiau plėtojama. Sukurta nemažai dirbtinio intelekto metodų akcijų kainoms prognozuoti, tokie kaip ARIMA, SVM (Support Vector Machine), ANN, DSS (Decision Support Systems) bei daugelis kitų, taip pat nemažai sukurta hibridinių modelių, apjungiančių skirtingus prognozavimo metodus. Nors tyrimams skiriama labai daug pastangų, dabartinė akcijų rinkos prognozavimo technika vis dar turi daug ribotumo bei neištirtumo. Atlikus mokslinės literatūros analizę matoma, jog akcijų rinkos prognozavimas vis dar išlieka labai sudėtingas uždavinys, todėl norint tiksliau ir efektyviau prognozuoti akcijos kainų svyravimus, reikėtų giliau tirti dirbtinio intelekto modelius, atsižvelgti į įvairius galimus įtakos turinčius veiksnius bei juos įtraukti.

Darbo mokslinis naujumas ir teorinis reikšmingumas

Vertybinių popierių rinkos pokyčių tyrimai sutelkti į dvi labai plačias sritis - akcijų rinkos efektyvumo tikrinimą ir akcijų kainų ar gražos modeliavimą. Akcijų rinkos efektyvumas daro įtaką akcijų kainų modeliavimui ir yra aiškiai užfiksuotas sąvokoje, vadinamoje efektyvios rinkos hipoteze (EMH). Akcijų rinkos indekso kainoms modeliuoti įprastai naudojami įvairūs modeliavimo būdai. Šie metodai buvo sutelkti į dvi prognozavimo sritis, būtent techninę analizę ir fundamentalią analizę (Marwala, 2008). Techninė analizė rodo, kad rinkos veikla atskleidžia reikšmingą naują informaciją ir supratimą apie psichologinius veiksnius, didinančius akcijų kainą, bandant numatyti būsimas kainas ir tendencijas. Techninis požiūris grindžiamas teorija, kad kaina yra masinės psichologijos („minios“) atspindys veikiant, ja bandoma numatyti būsimus kainų pokyčius, darant prielaidą, kad minios psichologija juda tarp panikos, baimės ir pesimizmo, taip pat pasitikėjimo savimi, per didelio optimizmo ir godumo kitam. Šiai analizės kategorijai priskiriama daugybė dirbtinio intelekto metodų (Bahrammirzaee, 2010), iš kurių labiausiai žinomi yra slenkamasis vidurkis (MA), autoregresyvusis integruotasis slenkamasis vidurkis (ARIMA) ir kiti dirbtinio intelekto metodai. Fundamentalią analizę daugiausia dėmesio skiriama pinigų politikai, vyriausybės politikai ir ekonominiams rodikliams, pavyzdžiui, BVP, eksportui, importui ir kitiems rodikliams verslo ciklo kontekste. Fundamentali analizė yra labai efektyvus būdas prognozuoti ekonominius dėsningumus, tačiau nebūtinai tikslią rinkos kainą. Matematiniai metodai, kurie naudojami fundamentalią analizę, apima vektorinę autoregresiją (VAR), kuris yra daugialypis kintamojo modeliavimo būdas. Šiame darbe bus nagrinėjami du labiausiai paplitę dirbtinio intelekto modeliai akcijų kainoms prognozuoti - ANN (Dirbtiniai neuroniniai tinklai) bei SVM (Atraminių vektorių mašinos), atliktas empirinis tyrimas dviejų minėtų metodų tikslumui įvertinti bei pateikti pasiūlymai, kaip modeliavimas galėtų būti papildomai patobulintas.

Darbo problema

Moksliniuose darbuose daug nagrinėtas būsimų akcijų kainų judėjimas naudojant dirbtinio intelekto metodus. Pagrindinis šio tyrimo dėmesys bus sutelktas į ANN ir SVM metodų tinkamumą akcijų prognozavimo kontekste, vertinant modelių pagalba gautų rezultatų tikslumą. Mokslinėje literatūroje pasigendama aiškių, dirbtiniu intelektu grįstų metodų pasirinkimo akcijų kainoms prognozuoti, kriterijų analizės, kodėl pasirinktas toks metodas, ko reikia, kad jį galima būtų pradėti taikyti, kaip turėtų būti paruošiami duomenys, kaip įvertinamas metodo tikslumas, ar visiems metodams galioja tokios pačios metodikos, jei ne, tai kuo jos skiriasi, o kuo panašios.

Tyrimo problema - ar dirbtiniu intelektu grįstais metodais galima tiksliai nustatyti akcijų kainų svyravimus.

Tyrimo objektas

Dirbtiniu intelektu grįsti metodai akcijų kainų svyravimui prognozuoti.

Tyrimo tikslas

Šio tyrimo tikslas - ištirti pasirinktų dirbtiniu intelektu grįstų metodų pritaikomumą akcijų kainoms prognozuoti, atliekant reikalingų prielaidų, apribojimų, privalumų bei trūkumų juos taikant, analizę.

Tyrimo uždaviniai

1. Išanalizuoti dirbtinio intelekto sampratą atliekant jo panaudojimo akcijų kainoms prognozuoti mokslinės literatūros analizę.
2. Remiantis atlikta mokslinės literatūros analize, išanalizuoti dažniausiai taikomus dirbtinio intelekto metodus, atliekant reikalingų prielaidų, apribojimų, privalumų bei trūkumų juos taikant, analizę.
3. Sukurti pasirinktų dirbtinio intelekto metodų taikymo akcijų kainoms prognozuoti palyginimo metodologiją.
4. Empiriškai atlikti pasirinktų dirbtinio intelekto metodų taikymo akcijų kainoms prognozuoti palyginimą.
5. Susisteminti gautus rezultatus, pateikti išvadas bei rekomendacijas apie tiriamų modelių tikslumą bei praktinį pritaikomumą.

Praktinis taikomumas

Sėkmingai prognozuojant akcijų kainas, galima gauti patrauklų pelną, tiek iš investuotojo, tiek iš analitiko perspektyvos. Tačiau dėl rinkos neapibrėžtumo bei begalės įtakos darančių veiksnių, mokslinių šaltinių analizė rodo, kad kol kas nėra atrasto vieno tobulo algoritmo, gebančio tiksliai prognozuoti tendencijas bei generuoti laukiamą grąžą. Atliktas tyrimas papildys mokslinių tyrimų perspektyvą ir sukurs prielaidas tolesnei tyrimų plėtotei.

Tyrimui reikalingi empiriniai duomenys ir jų rinkimo bei analizės metodai

Teorinei analizei atlikti buvo atlikta sisteminė literatūros šaltinių analizė, nagrinėjant daugiausiai pastarųjų dešimties metų tyrimus, „ScienceDirect“ duomenų bazėje (<https://www.sciencedirect.com>). Empiriniam tyrimui atlikti S&P indekso bei kintamųjų duomenys buvo renkami iš Yahoo Finance duomenų bazės (<https://finance.yahoo.com>). ANN modeliui formuoti surinkta 20 kintamųjų, SVM modeliui – 6, dešimties metų laikotarpyje nuo 2010.01.01 iki 2020.01.01, remiantis ankstesniais moksliniais tyrimais.

Duomenų analizei atlikti buvo naudojami du mašininio mokymosi metodai - dirbtiniai neuroniniai tinklai (*angl. Artificial Neural Networks*) ir atraminių vektorių mašinos (*angl. Support Vector Machines*). Tyrimai atlikti MATLAB programinės įrangos pagalba (versija R2020b), „Statistics and Deep learning Toolbox“ papildinio. Darbe taip pat taikomi aprašomosios statistikos metodai bei lyginamoji šaltinių analizė.

2. Santrumpų sąrašas

AI - Artificial Intelligence (liet. Dirbtinis intelektas)

AN - Artificial Neuron (liet. Dirbtinis neuronas)

ANN - Artificial Neural Networks (liet. Neuroniniai tinklai)

ARCH - Autoregressive Conditional Heteroskedasticity (liet. Autoregresinis sąlyginis heterostadiškumas)

ARMA - Autoregressive Moving Average (liet. Autoregresinis slenkamasis vidurkis)

CNN - Convolutional Neural Network (liet. Konvoliucinis neuroninis tinklas)

DSS - Decision Support Systems (liet. Sprendimų palaikymo sistemos)

DT - Decision Tree (liet. Sprendimų medis)

GA - Genetic Algorithms (liet. Genetiniai algoritmai)

GARCH - Generalized Autoregressive Conditional Heteroskedasticity (liet. Generalizuotas autoregresinis sąlyginis heterostadiškumas)

HMM – Hidden Markov Model (liet. Paslėptas Markovo modelis)

MAE – Measure of Errors (liet. Vidutinė absoliuti paklaida)

ML - Machine Learning (liet. Mašininis mokymasis)

MSE – Mean Square Error (liet. Kvadratų skirtumo vidurkis)

MSE – Mean Square Error (liet. Kvadratų skirtumo vidurkis)

NB – Naïve Bayes (liet. Naivusis Bajeso)

NN - Neural Networks (liet. Neuroniniai tinklai)

PSO – Particle Swarm Optimization (liet. Dalelių būrio optimizavimas)

R – Mean (liet. Vidurkis)

RMSE – Root-mean-square deviation (liet. Vidurkio kvadratų nuokrypis)

RNN – Recurrent Neural Network (liet. Pasikartojantis neuroninis tinklas)

SVM - Support Vector Machine (liet. Atraminių vektorių mašina)

SVR – Support Vector Regression (liet. Atraminių vektorių regresija)

3. Žodynėlis

ANN - dirbtiniai neuroniniai tinklai, paprastai tiesiog vadinami neuroniniais tinklais, yra skaičiavimo sistemos, kurias miglotai įkvėpė biologiniai neuroniniai tinklai, kurie sudaro gyvūnų smegenis. ANN yra pagrįstas sujungtų vienetų arba mazgų, vadinamų dirbtiniais neuronais, rinkiniu, kuris laisvai modeliuoja biologinių smegenų neuronus.

Dirbtinis intelektas (Artificial Intelligence) - kompiuterinių sistemų, galinčių atlikti užduotis, kurioms paprastai reikia žmogaus intelekto, teoriją ir plėtrą, tokias kaip regimasis suvokimas, kalbos atpažinimas, sprendimų priėmimas ir vertimas į kalbas.

Dirbtinis Neuronas (AN) - Dirbtinis neuronas yra matematinė funkcija, sukurta kaip biologinių neuronų modelis, neuroninis tinklas. Dirbtiniai neuronai yra elementiniai dirbtinio neuronų tinklo vienetai. Dirbtinis neuronas gauna vieną ar daugiau įėjimų ir juos sumina, kad gautų išėjimą.

Decision Tree (DT) - Sprendimų medis yra sprendimų palaikymo įrankis, kuris naudoja į medį panašų sprendimų modelį ir jų galimas pasekmes, įskaitant atsitiktinių įvykių rezultatus, išteklių sąnaudas ir naudingumą. Tai yra vienas iš būdų rodyti algoritmą, kuriame yra tik sąlyginiai valdymo teiginiai.

Klasifikacija - Kategorizavimas yra procesas, kurio metu idėjos ir objektai yra atpažįstami, diferencijuojami, klasifikuojami ir suprantami. Žodis „kategorizavimas“ reiškia, kad objektai yra rūšiuojami į kategorijas, dažniausiai tam tikru tikslu.

Mašinu mokymasis (Machine Learning) - kompiuterinių sistemų, galinčių mokytis ir prisitaikyti nesilaikant aiškių nurodymų, naudojimas ir tobulinimas, naudojant algoritmus ir statistinius modelius analizuojant ir darant išvadas iš duomenų modelių.

Nepaženklinti duomenys - Nepaženklinti duomenys susideda iš duomenų, kurie yra paimti iš gamtos arba sukurti žmogaus, norint ištirti už jo esančius mokslinius modelius. Kai kurie nepažymėtų duomenų pavyzdžiai gali būti nuotraukos, garso įrašai, vaizdo įrašai, naujienų straipsniai, tweetai, rentgeno nuotraukos ir kt.

Paženklinti duomenys - Paženklinti duomenys yra pavyzdžių grupė, pažymėta viena ar daugiau etikečių. Paprastai etiketėms priskiriamas nepažymėtų duomenų rinkinys ir kiekvienas jų gabalas papildomas informacinėmis žymomis.

Prižiūrimas mokymas - Prižiūrimas mokymasis yra mašininio mokymosi užduotis, išmokstanti funkciją, kuria įvestis susiejama su išvestimi, remiantis įvesties ir išvesties porų pavyzdžiais. Ji daro išvadą apie funkciją iš pažymėtų mokymo duomenų, sudarytų iš mokymo pavyzdžių rinkinio.

Regresija - regresijos analizė yra statistinių procesų rinkinys, skirtas sąryšiams tarp priklausomo kintamojo (dažnai vadinamo „rezultato kintamuoju“) ir vieno ar daugiau nepriklausomų kintamųjų (dažnai vadinamų „prediktoriais“, „kovariatais“ arba „ypatybėmis“) įvertinti.

SVM - Mašininio mokymosi metu palaikomojo vektoriaus mašinos yra prižiūrimi mokymosi modeliai su susijusiais mokymosi algoritmais, analizuojančiais klasifikacijai ir regresijos analizei naudojamus duomenis.

Duomenų Triukšmas - Triukšmingi duomenys yra sugadinti, iškraipyti arba turintys mažą signalo ir triukšmo santykį. Netinkamos procedūros triukšmui iš duomenų išskaičiuoti gali sukelti klaidingą tikslumo jausmą arba klaidingas išvadas.

4. Dirbtinio intelekto samprata bei vaidmuo akcijų rinkų kainų prognozėje

4.1. Pagrindinės sąvokos, ypatumai bei pritaikomumas

4.1.1. Dirbtinis intelektas - kas tai?

Dirbtinis intelektas (toliau AI) - paprastai reiškia žmogaus intelekto modeliavimą mašinos, įrangose, kurios užprogramuotos taip, kad sugebėtų mąstyti kaip žmonės, tik kad sparčiau - galėtų imituoti žmogaus protą bei nuspėtų tolimesnius veiksmus. Šis terminas taip pat gali būti taikomas bet kuriai mašinai, turinčiai bruožų, susijusių su žmogaus protu, tokių kaip mokymasis ir problemų sprendimas. Ideali, galutinė dirbtinio intelekto savybė yra jo sugebėjimas racionaliau ir efektyviau atlikti veiksmus, veiksmų seką, kurie turi geriausias galimybes pasiekti konkretų tikslą ar atlikti užduotį. Dirbtinis intelektas remiasi principu, kad žmogaus intelektas gali būti apibrėžtas taip, kad mašina galėtų lengvai jį kopijuoti ir vykdyti užduotis, pradedant nuo pačių paprasčiausių ir baigiant dar sudėtingesnėmis. Kadangi dirbtinis intelektas, kaip sąvoka, pradėta vartoti jau labai seniai, XX a. antroje pusėje, kai kurios sampratos, anksčiau apibrėžusios dirbtinį intelektą, tapo nebeaktualios. Ryšium su tuo, dirbtinis intelektas vis populiarėja daugelyje pramonės šakų, jo pagalba yra reikalinga įgyvendinti efektyvumo tikslus tiek mokslo srityse, tiek finansuose, politikoje, ekonomikoje, medicinoje, psichologijoje, gamyboje ir panašiai. Natūralu, jog atsirandant inovacijoms bei naujovėms, visi algoritmai bei modeliai turi būti testuojami, pagrindžiami, įvertinti bei ištirti, todėl mokslinė literatūra taip pat labai plačiai nagrinėja su juo susijusias problemas. Esant sparčiam kintamumui bei neapibrėžtumui akcijų rinkose, prognozavimas bei modeliavimas dirbtinio intelekto pagalba tapo vis labiau populiarus bei mokslininkų dėmesį pritraukiantis objektas ir finansų srityje, kur jis dažnai pritaikomas bankininkystės ir finansų veiklai automatizuoti. AI programos taip pat naudojamos siekiant supaprastinti ir palengvinti prekybą finansų rinkose. Tai atliekama sukuriant algoritmus taip palengvinant vertybinių popierių pasiūlos, paklausos ir kainų nustatymą bei prognozavimą.

Pagal (Kumar, 2018) AI galima sugrupuoti į dvi kategorijas:

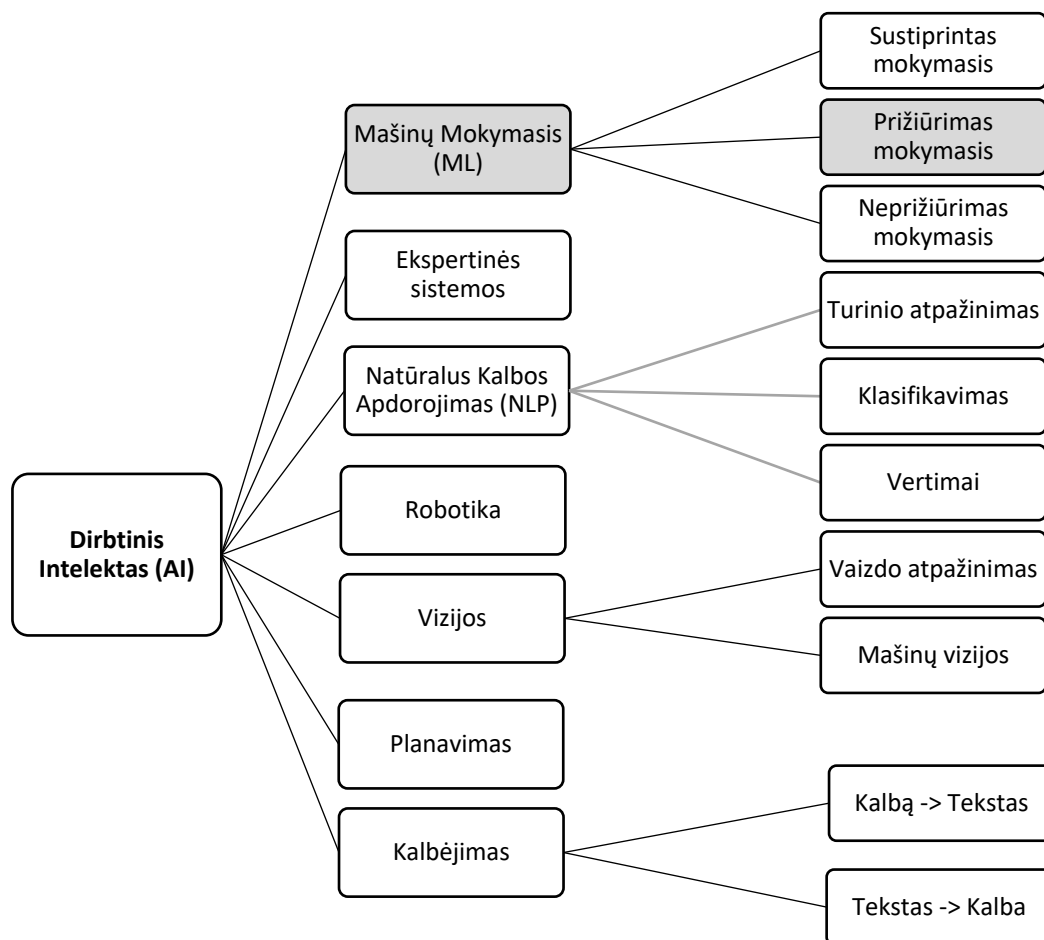
1 lentelė. Dirbtinio intelekto (AI) kategorijos

<u>Silpnas AI (angl. Weak AI)</u>	<u>Stiprus AI (angl. Strong AI)</u>
<ul style="list-style-type: none">• Sistema, skirta atlikti vieną konkretų darbą, veiksmą, procesą• Silpnam AI priskiriami kompiuteriniai žaidimai, taip pat virtualūs asistentai, tokie kaip „Siri“ - žmogus užduoda klausimą, kompiuteris atsako.	<ul style="list-style-type: none">• Sistemos, prilyginamos žmogaus gebėjimams, atliekančios sudėtingesnius veiksmus, didesnius procesus• Stiprus AI užprogramuotas spręsti situacijas, kuriuose gali reikėti išspręsti problemą neįsikišus asmeniui

	<ul style="list-style-type: none"> • Populiariausi pavyzdžiai - savarankiškai važiuojantys automobiliai, ligonių operacinės.
--	---

Šaltinis: sudaryta darbo autoriaus pagal Kumar, 2018.

Didžioji dalis algoritmų bei modelių, taikomų finansuose, konkrečiai akcijų kainų prognozavimui atlikti, priklauso silpno AI kategorijai. Dirbtinis intelektas šiuo metu turi keletą skirtingų atmainų, o jų aibė vis didėja, atsirandant naujoms technologijoms ir modeliams. Žemiau esantis paveikslas rodo, kaip dar AI gali būti klasifikuojamas (Kumar, 2018):



Pav. 1 Dirbtinio intelekto tiriamos sritys

Šaltinis: Sudaryta darbo autoriaus pagal Kumar, 2018

Svarbiausiomis atmainomis laikomos: Mašinių Mokymasis, Natūralus Kalbos Apdorojimas (NLP) bei Robotika (Robotics). Toliau šiame darbe bus plačiau nagrinėjama viena iš didžiausių dirbtinio intelekto taikymo formų - mašininis mokymasis (*angl. Machine Learning, santr. ML*).

4.1.2. Mašininis mokymasis bei jo atmainos

Terminas mašininis mokymasis pirmą kartą buvo paminėtas 1959 m. amerikiečio Arthuro Samuelio, kuris laikomas IBM ir kompiuterinių žaidimų bei dirbtinio intelekto pradininku. Septintajame dešimtmetyje buvo išleista mašininio mokymosi tyrimų reprezentacinė knyga - Nilsono „Mokymosi mašinos“, daugiausiai nagrinėjanti mašininį mokymąsi klasifikuojant modelius (Nilson, 1965). Susidomėjimas, susijęs su modelio atpažinimu, tęsėsi aštuntajame dešimtmetyje (Duda, Hart, 1973). 1981 m. buvo pateikta ataskaita apie mokymo strategijų naudojimą, kad neuroninis tinklas išmokytų atpažinti 40 ženklų (26 raidės, 10 skaitmenų ir 4 specialūs simboliai) iš kompiuterio terminalo (Bozinovski, 1981). Tomas M. Mitchellas pateikė plačiai cituojamą, formalesnį algoritmą, nagrinėjamų mašininio mokymosi srityje, apibrėžimą: „Sakoma, kad kompiuterio programa mokosi iš E patirties kai kurių užduočių klasei T ir atlikimo matui P, jei jos atlikimas atliekant užduotis T, matuojant P, tobulėja kartu su E patirtimi.“ (Mitchell, 1997). Šis užduočių, susijusių su mašininio mokymu, apibrėžimas pateikia esminį operatyvinį apibrėžimą, o ne srities apibrėžimą pažintine prasme. Tai seka Alano Turingo pasiūlymu savo dokumente „Kompiuterių technika ir intelektas“, kuriame klausimas „Ar mašinos gali mąstyti?“ pakeičiamas klausimu „Ar mašinos gali padaryti tai, ką mes (kaip mąstantys subjektai) galime padaryti?“ (Harnad, 2008).

Mašininis mokymasis yra kategorizuojamas į keturis tipus, atsižvelgiant į įvesties signalų bei išvesties grįžtamojo ryšio pobūdį:

- Prižiūrimas mokymasis (*angl. Supervised learning*)
- Neprižiūrimas mokymasis (*angl. Unsupervised learning*)
- Pusiau prižiūrimas mokymasis (*angl. Semi-supervised learning*)
- Sustiprintas mokymasis (*angl. Reinforcement learning*)

Kartais praktikoje išskiriamos tik trys kategorijos, pusiau prižiūrimas mokymasis nėra traktuojamas kaip atskiras mašininio mokymosi tipas. Taip pat, buvo pasiūlyti kiti požiūriai, kurie skiriasi nuo prieš tai minėto kategorizavimo. Svarbu paminėti, jog kartais ta pati mašininio mokymosi sistema gali būti priskiriama daugiau nei prie vieno iš tipų. (Brazdil, Carrier, Soares, Vilalta, 2009). Nuo 2020 m. giluminis mokymasis (*angl. Deep learning*) tapo vyraujančiu metodu atliekant daug nuolatinio, nepertraukiamo darbo mašinų mokymosi srityje (Alpaydin, 2020).

Mašininis mokymasis yra dirbtinio intelekto (AI) taikymo forma, suteikianti sistemoms galimybę automatiškai mokytis ir tobulėti remiantis patirtimi, prieš tai vykusiais žingsniais, be pilno programavimo. Mašininis mokymasis yra susijęs su kompiuterinių programų, kurios gali priėti prie duomenų ir taikyti juos analizei bei problemų sprendimui, tobulinimu. Mokymosi procesas prasideda stebėjimais ar duomenimis, tokiais kaip pavyzdžiai, tiesioginė patirtis ar nurodymai, siekiant ieškoti duomenų modelių ir ateityje priimti geresnius sprendimus remiantis jau prieš tai pateiktais pavyzdžiais.

Pagrindinis tikslas yra leisti kompiuteriams mokytis automatiškai be žmogaus pagalbos ir atitinkamai pakoreguoti veiksmus, esant reikalui. Kitaip tariant, tai informacinių technologijų pagalba suprogramuoti modeliai, gebantys palaikyti didelį kiekį kintamųjų bei duomenų srautų, algoritmų pagalba pateikiantis reikalingus rezultatus.

Mašininis mokymasis yra glaudžiai susijęs su matematika, statistika, kurioje pagrindinis dėmesys skiriamas prognozavimui kompiuterių pagalba. Matematinio optimizavimo tyrimas pateikia metodus, teorijas ir taikymo sritis mašininio mokymosi srityje. Duomenų gavyba yra susijusi tyrimų sritis, kurioje pagrindinis dėmesys skiriamas tiriamųjų duomenų analizei per neprižiūrimąjį mokymąsi (*angl. unsupervised learning*). Sprendžiant verslo problemas, mašininis mokymasis taip pat vadinamas prognozuojančioji analitika (*angl. predictive analytics*).

Žemiau esančioje lentelėje aprašomos keturios pagrindinės mašininio mokymosi kategorijos:

2 lentelė. Mašinų mokymasis: tipai bei paaishkinimai

<u>Mašininio mokymosi kategorija</u>	<u>Aprašymas</u>	<u>Pavyzdžiai</u>
Prižiūrimas mokymasis	<p>Tokie algoritmai gali pritaikyti tai, kas buvo išmokta praeityje, naujiems duomenims generuoti, naudojant konkrečius pavyzdžius būsimiems įvykiams numatyti. Pradedama nuo duomenų rinkinio analizės, mokymosi algoritmas sukuria numanomą funkciją numatyti išvesties reikšmes. Sistema gali pateikti tikslus bet kuriai naujai įvestai informacijai atlikus reikiamą mokymą. Mokymosi algoritmas taip pat gali palyginti savo išvestį su teisinga, numatoma išvestimi ir rasti klaidų, kad atitinkamai modifikuotų modelį.</p> <p>Prižiūrimų mokymosi algoritmų tipai apima aktyvų mokymąsi, klasifikavimą ir regresiją. (Alpaydin, 2010).</p> <p>Klasifikavimo algoritmai naudojami, kai išėjimai yra ribojami ribotam verčių rinkiniui, ir regresijos algoritmai naudojami, kai išėjimai gali turėti bet kokią skaitinę reikšmę diapazone. Kaip pavyzdį, klasifikavimo algoritmui, kuris filtruoja el. laiškus, įvestis būtų gaunamas el. laiškas, o išvestis būtų aplanko, į kurį reikia siųsti el. laišką, pavadinimas. Panašumo mokymasis yra</p>	Regresija, klasifikacija, Neuroniniai Tinklai, Atraminiai vektoriai (SVM)

	<p>prižiūrimo mašininio mokymosi sritis, glaudžiai susijusi su regresija ir klasifikacija, tačiau tikslas yra išmokti iš pavyzdžių, naudojant panašumo funkciją, matuojančią, kaip panašūs ar susiję du objektai. Jis turi programas reitingo, rekomendacijų sistemų, vaizdinės tapatybės stebėjimo, veido patikrinimo ir garsiakalbių tikrinimo srityse.</p>	
<p>Neprižiūrimas mokymasis</p>	<p>Priešingai, neprižiūrimi mašininio mokymosi algoritmai yra naudojami, kai mokomoji informacija nėra nei klasifikuojama, nei paženklinama. Neprižiūrimas mokymasis tiria, kaip sistemos gali pateikti funkciją ir padaryti išvadą apie ją iš neapdorotų ir nepaženklintų duomenų. Sistema nesuvokia kokia turi būti išvestis, tačiau tiria duomenis ir gali daryti išvadas iš duomenų rinkinių, kad apibūdintų paslėptas struktūras iš nepaženklintų duomenų.</p> <p>Pagrindinis neprižiūrimo mokymosi taikymas yra tankio įvertinimas statistikoje, pavyzdžiui, tikimybės tankio funkcijos nustatymas. (Jordan, Bishop, 2004)</p> <p>Neprižiūrimas mokymasis apima ir kitas sritis, apimančias duomenų savybių apibendrinimą ir paaiškinimą. Klasterio analizė yra stebėjimų aibės priskyrimas pogrupiams (vadinamiems klasteriais), kad to paties klasterio stebėjimai yra panašūs pagal vieną ar kelis iš anksto nustatytus kriterijus, o stebėjimai iš skirtingų klasterių skiriasi. Skirtingi grupavimo būdai daro skirtingas prielaidas dėl duomenų struktūros, dažnai apibrėžiami tam tikru panašumo metru ir vertinami, pavyzdžiui, vidiniu kompaktiškumu ar panašumu tarp tos pačios grupės grupių ir atskyrimu, skirtumų tarp grupių. Kiti metodai yra pagrįsti įvertintu tankiu ir grafiko jungiamumu.</p>	<p>Klasterizavimas, Markovo Modelis, Duomenų analizė, Pagrindinio Komponento Analizė, Asociacijos taisyklė</p>
<p>Pusiau prižiūrimas mokymasis</p>	<p>Pusiau prižiūrimi mašininio mokymosi algoritmai atsiduria kažkur tarp prižiūrimo ir neprižiūrimo mokymosi, nes mokymui jie naudoja ir paženklintus, ir nepaženklintus duomenis - paprastai tai yra nedidelis kiekis duomenų</p>	<p>Nepažymėtų duomenų rinkinių tyrimas, Paprastasis</p>

	etikečių ir didelis nepažymėtų duomenų kiekis. Sistemos, kurios naudoja šį metodą, gali žymiai pagerinti mokymosi tikslumą. Šis mokymasis dažnai naudojamas, kai neturima pakankamai prieigos ir išteklių prie paženklintų duomenų.	ženklėjimas, Įvesties – Išvesties simetrija
Sustiprintas mokymasis	Tai mokymosi metodas, sąveikaujantis su aplinkos faktoriais atliekant veiksmus ir nustatant klaidas ar naudą. Bandymų ir klaidų paieška bei atidėtas atlygis yra svarbiausios stiprinimo mokymosi savybės. Šis metodas leidžia mašinoms ir programinės įrangos agentams automatiškai nustatyti idealų elgesį konkrečiame kontekste, siekiant maksimaliai padidinti jo efektyvumą. Kad agentas sužinotų, kuris veiksmas yra geriausias, reikalingas paprastas atlygis už atlygį; tai vadinama stiprinimo signalu. Mašinų mokyme duomenų aplinka paprastai vaizduojama kaip Markovo sprendimų procesas (MDP). Daugelyje sustiprinto mokymosi algoritmų naudojami dinaminio programavimo būdai. (Otterlo, Wiering, 2012). Stiprinimo mokymosi algoritmai nereikalauja žinių apie tikslų MDP matematinį modelį ir yra naudojami, kai tikslų modelių neįmanoma rasti.	Neuroniniai tinklai, garso bei vaizdo atpažinimas, Rekomendacijų sistemos, Autopilotas

Šaltinis: sudaryta autoriaus

Taip pat, be prieš tai minėtų keturių pagrindinių mašininio mokymosi metodų, išskiriami kiti galimi tipai:

Savarankiškas mokymasis (*angl. Self Learning*) - Savarankiškas mokymasis kaip mašininio mokymosi paradigma buvo įvestas 1982 m. kartu su neuroniniu tinklu, galinčiu savarankiškai mokytis, pavadinimu Crossbar Adaptive Array (toliau, CAA) (Bozinovski, 1982). Tai mokymasis, neturintis išorinio atlygio ir neturintis išorės mokytojų patarimų. CAA savarankiško mokymosi algoritmas kryžminiu būdu apskaičiuoja sprendimus dėl veiksmų ir emocijų dėl pasekmių situacijų. Sistemą skatina pažinimo ir emocijų sąveika. (Bozinovski, 2014) CAA egzistuoja dviejose aplinkose: viena yra elgesio aplinka, kurioje ji elgiasi, o kita - genetinė aplinka, iš kurios ji iš pradžių ir tik kartą gauna pirmines emocijas apie situacijas, su kuriomis susidurs elgesio aplinkoje. Iš genetinės aplinkos gavusi genomo (rūšies) vektorių, CAA sužino tikslą, kurio siekiama elgesiu aplinkoje, kurioje yra ir pageidaujamų, ir nepageidaujamų situacijų. (Bozinovski, 2001)

Funkcijų mokymasis (*angl. Feature Learning*) - Funkcijų mokymosi algoritmai, dar vadinami reprezentaciniais mokymosi algoritmais, dažnai bando išsaugoti įvestą informaciją, bet taip pat ją transformuoja taip, kad ji būtų naudinga, dažnai kaip išankstinio apdorojimo žingsnį prieš atliekant klasifikaciją ar prognozes. Funkcijų mokymasis gali būti prižiūrimas arba neprižiūrimas. Mokant prižiūrimų funkcijų, funkcijos mokomosiose etiketėse įvedami duomenys. Pavyzdžiai: dirbtiniai neuroniniai tinklai, daugiasluoksniai suvokimai ir prižiūrimas žodyno mokymasis. Mokantis neprižiūrimų funkcijų, funkcijos mokosi naudojant nepaženklintus įvesties duomenis. Pavyzdžiai: žodyno mokymasis, nepriklausoma komponentų analizė ir įvairios klasterizacijos formos (Coates, Honglak, Andrew, 2011). Funkcijų mokymasis motyvuojamas tuo, kad mašininio mokymosi užduotims, tokioms kaip klasifikavimas, dažnai reikalingas įvestis, kurią patogiau apdoroti matematiškai ir skaičiavimo būdu. Tačiau realiojo pasaulio duomenys, tokie kaip vaizdai, vaizdo įrašai ir jutimo duomenys, nedavė bandymų algoritmiškai apibrėžti specifines savybes. Alternatyva yra aptikti tokias savybes ar reprezentacijas tiriant, nepamokant aiškiais algoritmais. (Jurafsky, James, 2009).

Retas žodyno mokymasis (*angl. Sparse dictionary learning*) – Retas žodyno mokymasis yra būdingas mokymosi metodas, kai mokymo pavyzdys pateikiamas kaip linijinis bazinių funkcijų derinys ir laikoma, kad tai yra negausi matrica. Metodas yra labai sunkus, sunkiai išsprendžiamas. (Tillmann, 2015). Populiarus euristinis retų žodynų mokymosi metodas yra K-SVD algoritmas. Retas žodyno mokymasis buvo pritaikytas keliuose kontekstuose. Klasifikuojant problema tampa nustatyti klasę, kuriai priklauso anksčiau nematytas mokymo pavyzdys. Žodyne, kuriame kiekviena klasė jau buvo sukurta, naujas mokymo pavyzdys yra susijęs su klase, kurią geriausiai reprezentuoja atitinkamas žodynas. Retas žodyno mokymasis taip pat buvo pritaikytas vaizdų triukšmui mažinti. Pagrindinė loginė mintis yra ta, kad švarų vaizdų pataisą gali nedaug pavaizduoti vaizdų žodynas, tačiau triukšmas negali. (Aharon, Elad, Bruckstein, 2006).

Anomalijos nustatymas (*angl. Anomaly detection*) – atliekant duomenų gavybą, anomalijos nustatymas, dar žinomas kaip išorinis aptikimas, yra retų daiktų, įvykių ar stebėjimų, kurie sukelia įtarimus, žymiai skiriasi nuo daugumos duomenų nustatymas (Zimer, Schubert, 2017). Paprastai anomalūs elementai atspindi tokią problemą kaip banko apgaulė, struktūrinis trūkumas, medicininės problemos ar klaidos tekste. Anomalijos vadinamos nuokrypiais, naujovėmis, triukšmu, nuokrypiais ir išimtimis. (Hodge, Austin, 2004). Egzistuoja trys plačios anomalijų nustatymo metodų kategorijos. (Chandola, Kumar, 2009). Neprižiūrimi anomalijų aptikimo būdai aptinka nepaženklintų bandymo duomenų rinkinio anomalijas, darant prielaidą, kad dauguma duomenų rinkinio egzempliorių yra normalūs, ieškant atvejų, kurie atrodo mažiausiai tinkami prie likusios duomenų rinkinio dalies. Prižiūrimi anomalijų aptikimo būdams reikia duomenų rinkinio, kuris buvo paženklintas kaip „normalus“ ir „nenormalus“ ir apima klasifikatoriaus mokymą (pagrindinis skirtumas nuo daugelio kitų statistinės klasifikavimo problemų yra savaime nesubalansuotas išorinio aptikimo pobūdis). Iš dalies

prižiūrimi anomalijų aptikimo būdai sukuria modelį, parodantį normalų elgesį iš tam tikro įprasto treniruočių duomenų rinkinio, ir tada patikrina bandymo pavyzdžio tikimybę, kad jį sugeneruos modelis.

Robotų mokymasis (*angl. Robot learning*) - vystomojoje robotikoje robotų mokymosi algoritmai sukuria savo mokymosi patirties sekas, dar žinomas kaip mokymo programa, kad būtų galima kaupiti naujus įgūdžius savarankiškai tiriant ir bendradarbiaujant su žmonėmis. Šie robotai naudoja tokius orientavimo mechanizmus kaip aktyvus mokymasis, brendimas, motorinė sinergija ir imitacija.

4.1.3. Vaidmuo finansų rinkose

Kaip ir buvo minėta darbo pradžioje, dirbtinio intelekto pritaikomumas finansų rinkose pradėtas tirti prieš kelis dešimtmečius, nuo 2010-ųjų metų - vis aktyviau. Mašininio mokymosi ir dirbtinio intelekto metodai teikia didelę naudą finansinių sprendimų priėmimo, atsižvelgiant į naujai atsirandančius modeliavimo ir prognozavimo metodus. Jų svarba taip pat buvo pripažinta finansų įmonių, kurios teigė, jog iki 2021 planuoja investuoti net apie 28 mlrd. USD dirbtinio intelekto integracijai įmonėse (IDC, 2018). Šiuo metu jiems daugiausiai dėmesio sutelkiama tokiose finansų srityse kaip algoritminė prekyba (*angl. algorithmic trading*), rizikos valdymas, procesų automatizavimas bei akcijų kainų prognozavimas (Aziz, Dowling, Hammami, 2019). Tačiau, kol kas moksliniai tyrimai bei jų kiekis atsilieka nuo praktinio dirbtinio intelekto panaudojimo, todėl ši tema tampa vis labiau aktuali tarp šios srities mokslininkų.

ML potencialas priimant finansinius sprendimus pirmiausia buvo ištirtas Hawley ir kt. (1990), iš tyrimų perspektyvos, sutelkiant dėmesį į neuroninius tinklus, kaip pagalbą priimant finansinius sprendimus. Dešimtajame dešimtmetyje „Journal of Banking & Finance“ taip pat pasirodė keletas žinomų bankų pavyzdžių, kuriuose buvo tiriama ML galimybė pagerinti skolinimo sprendimus ir kredito rizikos valdymą. Altman ir kt. (1994) pritaikė neuroninius tinklus, kad klasifikuotų Italijos firmas pagal finansinių sunkumų tikimybę, tuo tarpu Varetto (1998), remdamasis šiuo tyrimu, ta pačia tema pritaikė genetinio mokymosi algoritmus. Minėtuose tyrimuose dėmesys buvo sutelktas į prognozavimą, tačiau buvo pereita prie gilaus mokymosi ir kitų pažangių ML metodų. Šios naujausios programos apima: numatytųjų atkūrimo rodiklių supratimą (Cheng ir Cirillo, 2018); optimalių variantų apsidraudimo normų mokymąsi (Nian ir kt., 2018); investuotojų nuotaikos modeliavimą („Renault“, 2017 m.); ir akcijų kainų raidos nustatymą remiantis užsakymų knygomis (Kercheval ir Zhang, 2015).

Kadangi finansų žurnalai ėmėsi preliminarinių žingsnių pripažindami naujų ML metodų galimybes finansuose, kitos disciplinos įdėjo griežtesnes pastangas taikyti ML metodus finansiniams duomenims apdoroti. Iš dalies taip yra dėl išsamių, struktūrizuotų ir lengvai prieinamų finansų duomenų patrauklumo. Iš tiesų ML ir finansų tyrimai ne finansų žurnaluose gerokai viršija ML ir finansų tyrimus finansų žurnaluose. Keletas naujausių šio nefinansinio ML korpuso pavyzdžių atliekant finansų tyrimus apima vaizdo atpažinimo metodų taikymą vertinant atsargų techninius modelius (Sezer ir Ozbayoglu,

2018), akcijų kainų prognozavimas, remiantis internetinės paieškos ansamblio modeliais ir nuotaikos duomenimis (Weng ir kt., 2018), ir banko rizikos modeliavimas (Cerchiello ir kt., 2017). Ši plačios perspektyvos ML ir finansų tyrimų plėtra kelia iššūkius tyrėjams, norintiems suprasti šios srities tyrimų spektrą.

Tiriant ML pritaikomumą, pateikta pirmoji finansinių temų modeliavimo programa - tikimybinę metodiką panašioms moksliniams tyrimams suderinti (Aziz, Dowling, Hammami, 2019). Įvairi literatūra apie ML taikymą finansuose rodo šios technikos stiprybę holistiškai identifikuojant ir grupuojant atitinkamus tyrimus tema. Finansų žurnaluose skelbiama tik nedidelė ML taikymo finansinėms problemoms dalis, nepaisant įgimto finansų poreikio nuolat tobulinti prognozavimo ir modeliavimo metodikas. Šios tematikos aktualumas paremtas tuo, jog finansuose ML yra gan paplitęs, tačiau ne iki galo iširtas.

Atlikus literatūros analizę ML tematika, vyraujančios temos išlieka tokios: investicijų analizė, turto modeliavimas ir prognozavimas bei rizikos valdymas, taip pat 14 temų, susijusių su šiomis temomis. Šios trys kategorijos, nors ir dalijasi kai kuriomis technikomis, progresuoja skirtingu greičiu. Rizikos valdymas dėl ankstyvo mokslinių tyrimų prasidėjimo yra labai pažengęs taikant mašininį mokymąsi rizikos modeliavimo srityje. Tai matoma praktinėje kredito rizikos modeliavimo srityje, siekiant paremti banko ir kitų finansų įstaigų skolinimo sprendimus. Tačiau šiuo metu sritis kenčia nuo padidėjusio duomenų saugumo reguliavimo, nes svarbiausius duomenis įmonės laiko privačiai savo vidaus naudojimui. Turto modeliavimas ir prognozavimas taip pat yra pažengęs dėl ankstyvo supratimo apie neuroninių tinklų tinkamumą finansinių kartų eilučių modeliavimui. Mažiau pažengusi, bet galbūt perspektyvesnė yra investicijų analizės kategorija. Šioje kategorijoje slypi naujų duomenų generavimo pažadai, pavyzdžiui, apie investuotojų nuotaikas, elgesį ir pagrindinių finansų kintamųjų prognozavimą. Temų tyrimų analizė rodo, kad investuotojų nuotaikos yra viena iš pagrindinių augimo kategorijų. Ši kategorija atveria finansų ateities potencialą taikydama tokias programas kaip roboto konsultavimas ir kitos finansinės konsultavimo paslaugos. Jis taip pat remiasi tekstiniu statistiniu mokymusi, kurio analizės metodai šiuo metu yra greitai tobulinami.

Taip pat vertėtų išskirti sritis, kurių dar nesprenžia mašininis mokymasis, ypač susijusios su įmonių finansais ir investicinių bankininkų vaidmenimis, tokiais kaip susijungimai ir įsigijimai, bei tvirtais finansavimo sprendimais. Tai nereiškia, kad tyrimai nėra atliekami šiose srityse, o tik tai, kad tyrimai nėra pakankami, kad būtų galima jais remtis praktikoje. Viena svarbiausių problemų čia yra duomenų prieinamumas problemoms tirti, panašus į ML problemą rizikos valdymo tyrimuose. Finansų tyrimai apskritai turi gana ribotą prieigą prie duomenų tiek įmonėse, tiek investicijose. Norint plėtoti šią sritį, reikia atlikti tyrimus, kurie galėtų pasiekti šiuos privačių įmonių duomenis. Taip pat trūksta finansinių tinklų analizės, kuri yra pagrindinis platesnio intelekto tyrimų judėjimo objektas. Taip pat yra galimybių išplėsti tekstinę analizę ne tik rinkos požiūriu. Finansinė veikla yra labai dokumentais pagrįsta

veikla, o tobulėjant ML tekstinės analizės metodams, vertėtų daugiau dėmesio skirti žinių gavimui iš šių duomenų saugyklų. Kadangi finansinių tyrimų sritis yra grindžiama praktiniu pritaikomumu, ML yra ne tik aktualus, bet ir patrauklus tyrimo objektas, dėl pilnai praktinio pobūdžio bei plataus panaudojimo spektro.

Taigi, išanalizavus jau atliktus ML tematika finansuose tyrimus bei sunkumus juos atliekant, matome, jog kol kas plačiausiai nagrinėjamos temos yra investicinių sprendimų priėmimas bei prognozavimas (tiek finansinių rodiklių, tiek akcijų kainų). Pagrindiniais sunkumais išlieka duomenų prieiga bei ribotumas, taip pat pačių algoritmų pažinimas bei jų potencialas, galimybės. Toliau šiame darbe bus plačiau plėtojama akcijų kainų prognozavimo tema, ML pritaikomumas joje, esamų metodų apžvalga, įvertinant jų privalumus bei trūkumus.

4.2. Dirbtinio intelekto metodai akcijų kainų prognozei

4.2.1. Mašininio mokymosi metodų apžvalga

Nepaisant didelių susijusių iššūkių, modelių paieška finansų rinkų kainoms prognozuoti vis dar nėra iki galo ištirta. Finansinio turto kainos yra nelinijinės, dinamiškos ir chaotiškos; literatūroje tai vadinama finansinėmis laiko eilutėmis, kurias sunku nuspėti. Tarp naujausių metodų mašininio mokymosi modeliai yra vieni iš labiausiai ištirtų, atsižvelgiant į jų galimybes atpažinti sudėtingus modelius įvairiose programose. Mašininio mokymosi paremtų metodų egzistuoja gana daug, ir jų vis atsiranda, nes vieni metodai keičia kitus, yra tobulinami, jungiami į hibridinius metodus, o tai yra susiję su gana sparčiu informacinių technologijų tobulėjimu.

Akcijų kainos prognozavimo pažanga įgijo didelę reikšmę tarp ekspertų, analitikų ir investuotojų. Akcijų rinkos prognozavimas yra sudėtingas dėl įvairių aplinkos faktorių ir didelio nepastovumo rinkos tendencijų atžvilgiu. Akcijų kainų sudėtingumas prisitaiko prie tam tikrų veiksnių, susijusių su ketvirčio pajamų ataskaitomis, rinkos naujienomis ir skirtingu besikeičiančiu elgesiu. Prekybininkai priklauso nuo įvairių techninių rodiklių, kurie yra pagrįsti kasdien kaupiamomis atsargomis. Nors šie rodikliai naudojami akcijų grąžai analizuoti, sudėtinga prognozuoti dienos ir savaitės tendencijas rinkoje (Ticknor, 2013). Tikslus tendencijų numatymas yra įdomus ir sudėtingas uždavinys vis besikeičiančiame pasaulyje. Keli aspektai, į kuriuos atsižvelgiama prognozuojant, turintys įtakos akcijų tendencijoms yra neekonominiai ir ekonominiai veiksniai. Taigi, vertybinių popierių rinkos prognozavimas laikomas pagrindiniu iššūkiu siekiant ekonominės grąžos. Tradiciniai metodai atskleidžia, kad akcijų rinkos pajamos yra prognozuojamos pagal ankstesnę akcijų grąžą ir kitus finansinius kintamuosius bei makroekonomiką. Akcijų rinkos pajamų prognozavimas nukreipė investuotojus į nuspėjamumo priešasčių tyrimą. Akcijų tendencijų prognozavimas yra sunkus procesas, nes tam įtakos turi keli aspektai, susiję su prekybininko lūkesčiais, finansinėmis aplinkybėmis, administraciniais įvykiais ir tam tikrais aspektais, susijusiais su rinkos tendencijomis. Be to, akcijų kainų

sąrašas paprastai yra dinamiškas, sudėtingas, triukšmingas, neparаметrinis ir netiesinis (Boyacioglu M, 2010). Finansinių laiko eilučių prognozavimas tampa problema dėl tam tikrų sudėtingų ypatybių, tokių kaip nepastovumas, pažeidimai, triukšmas ir kintančios tendencijos (Araujo R, 2013).

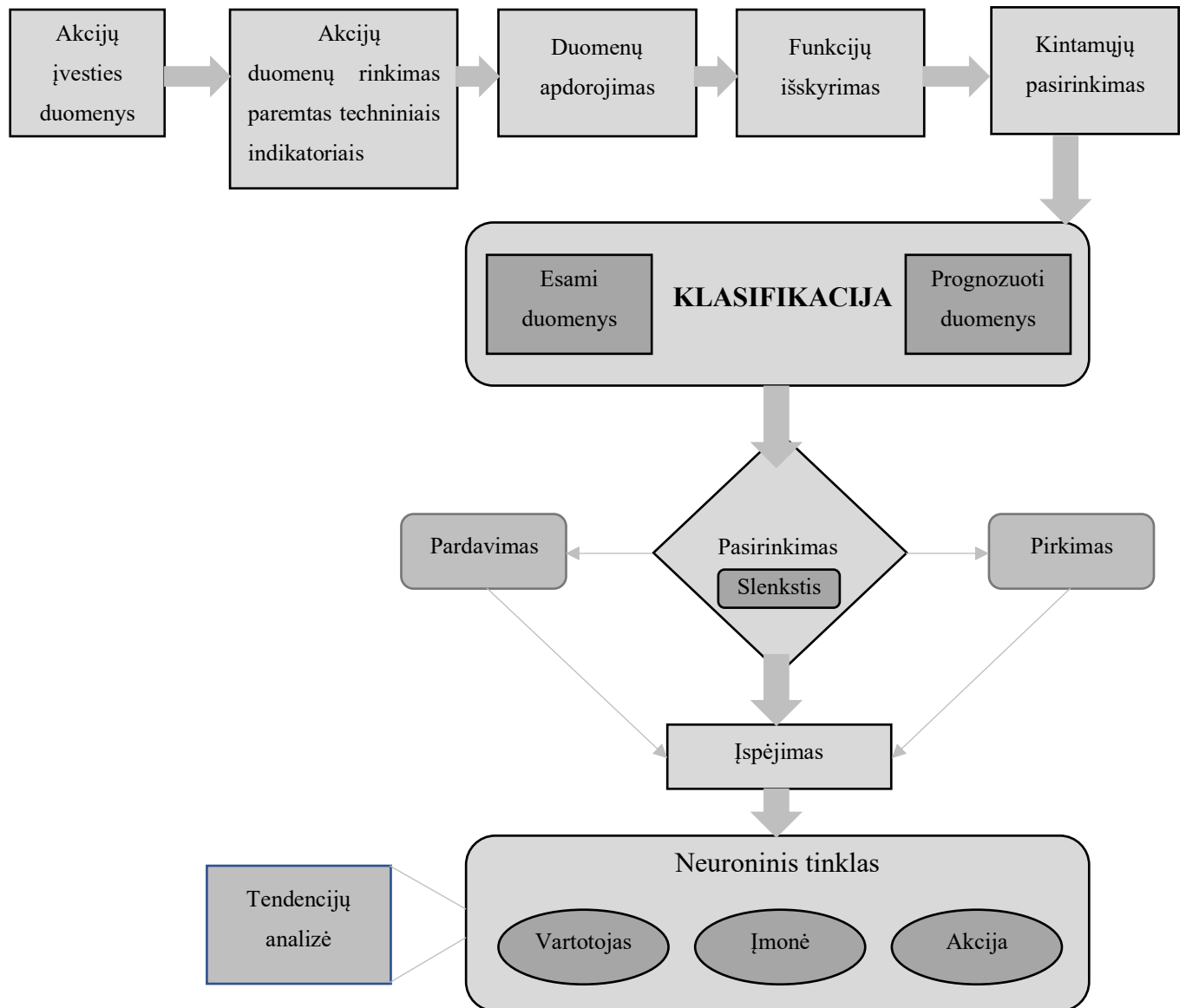
Įvairūs modeliai, taikomi prognozuojant akcijų kainas, valdomi naudojant laiko eilučių modelius, apimančius automatinio regresinio sąlyginio heteroskedastinio (ARCH) modelį, apibendrintą automatiškai regresyvu slenkamu vidurkiu (GARCH) ir automatiškai regresyvu slenkamu vidurkiu (ARMA). Tačiau šie modeliai apima istorinius duomenis ir hipotezes. Keli metodai, naudojami vertybinių popierių rinkos prognozavimui, yra pagrįsti įprastomis laiko eilutėmis, pavyzdžiui, neryškių laiko eilučių duomenys, realieji skaičiai ir neryškių rinkinių dizainas. Apytiksliai laiko eilučių duomenys yra naudojami vertybinių popierių rinkos prognozavimui, apdorojant konkrečios matematinės reikšmės duomenis, kad būtų galima gauti tikslius prognozavimo rezultatus. Šie metodai yra plačiai naudojami prognozuojant netiesinius ir dinامينius duomenų rinkinius kintančiose srityse, tokiose kaip turizmo paklausa ir akcijų rinkos (Chen, M, 2015). Prognozuojant vertybinių popierių rinkos grąžą, naudojama daugybė AI grįstų metodų, tokių kaip minkšto skaičiavimo algoritmai, neuroniniai tinklai (ANN), impulsų sklaidimo algoritmas ir genetinis algoritmas (GA). GA ir ANN yra integruotos projektuojant hibridines ekspertų sistemas, kad būtų galima priimti investavimo sprendimus. GA sukurta technika skirta bruožams atskleisti ir nustatyti dirbtinių neuroninių tinklų (ANN) svorį, siekiant numatyti akcijų kainos indeksą. Laiko eilučių modelis ir NN buvo derinami siekiant numatyti akcijų kainos kintamumą. Dirbtinio intelekto (AI) metodai, kaip ir ANN, buvo sukurti akcijų rinkos kainų prognozavimui. Daugelis tinklų, norėdami numatyti akcijų tendencijas, naudojo „feed forward“ neuroninius tinklus ir vertino kelis parametrinius ir neparаметrinius modelius, kad prognozuotų akcijų rinkos grąžą (Wei L., 2011). Sprendimams priimti taikomi pelno ir nuostolio kriterijai, taikomi intelektualiosios kompiuterijos metodai. Naudojami neryškios logikos metodai, dalelių būrio optimizavimas (*angl. Particle Swarm Optimization, santr. PSO*), ANN ir SVM. Keli tyrėjai bandė naudoti neapibrėžtus metodus ir atsitiktinumą, kad optimizuotų kainodaros modelius.

Toliau šiame darbe bus atlikta išsami modelių apžvalga, naudojant įvairias akcijų rinkos prognozavimo metodikas, kuriomis remiantis prognozuojamos ateities tendencijos ir akcijų grąža. Esami vertybinių popierių rinkos prognozavimo metodai bus suskirstyti į atskirus klasifikavimo ir grupavimo metodus, be to, tyrimas atliekamas siekiant identifikuoti galimus trukdžius, tokius kaip duomenų prieinamumas, algoritmų veikimo įvertinimo kriterijai bei jų tikslumas ir praktinio pritaikomumo įvertinimas.

4.2.2. Mašininio mokymosi modelių modeliai bei jų klasifikacija

Akcijų rinka įgijo investuotojų patrauklumą dėl pažangių programų, kuriose prognozavimas gali padėti sėkmingai prognozuoti rinką. Akcijų tendencijų prognozavimas tiesiogiai priklauso nuo

investavimo ir prekybos akcijų duomenimis. Vertybinių popierių rinkos prognozavimui naudojamos priemonės gali stebėti, numatyti ir reguliuoti rinką, kuri gali būti naudojama priimant teisingus sprendimus (Majumder M., 2009). Akcijų rinka turi susidurti su daugeliu skirtingo formato akcijų duomenų, kurie apima visą finansų rinką. Atsižvelgdami į rinkos būklę, investuotojai veiksmus pritaiko atsižvelgdami į pardavimą ir pirkimą. Keletas veiksnių, turinčių įtakos rinkos būklei, yra būsimų pajamų įvertinimas, pelno indikatoriai, dividendų deklarasavimas, valdymo pokyčiai ir panašiai. Tiriant vertybinių popierių prekybos klausimus, turi būti numatyti labai įvairūs aspektai, kurie laikomi įtakojantys akcijų kainą (Bharambe M., 2017). Prognozavimo mechanizmai, pagrįsti akcijų rinka, vaidina svarbų vaidmenį pritraukiant daugiau žmonių ir esamų investuotojų į bendrą vietą. Tikslūs vertybinių popierių rinkos prognozavimo rezultatai padeda investuotojams priimti geresnius sprendimus. Duomenų gavybos priemonės gali padėti investuotojams numatyti ateities tendencijas ir elgesį, ir padėti institucijoms rasti aktyvių sprendimų priimti žiniomis pagrįstus sprendimus. Kruopšti bei argumentuota duomenų analizė padeda algoritmams bei kompiuteriniams įrankiams lengvai gauti rezultatus. Naudingos informacijos gavimas yra veiksmingas duomenų gavybos būdas (Kumar A., 2018). Duomenų gavyba yra būtina norint prognozuoti akcijų rinką, kuri gali ieškoti paslėptų dalių ir padidinti tikslumo lygį analizuojant rinkos tendencijas naudojant metodus, tokius kaip regresijos metodas, žinių atradimas duomenų bazėse (*angl. Knowledge discovery in Databases, santr. KDD*) bei kiti metodai, kurie bus aprašyti toliau. Kad būtų aiškiau suprasti pagrindinius principus, žemiau esančiame paveiksle pavaizduotas akcijų rinkos prognozavimo sistemos veikimas (Pav. 2). 2 paveiksle atvaizduota akcijų rinkos prognozavimo sistemos funkcinė schema. Iš pradžių istoriniai atsargų duomenys yra renkami iš įvairių duomenų rinkinių, dažniausiai iš tokių kaip didžiausios finansų biržos ir kt., bei remiantis techniniu indeksu. Čia techninis indeksas apibrėžiamas kaip produkto, kuris bus naudojamas konkrečiame algoritme, tinkamumas. Tada surinkti duomenys apdorojami ir paruošiami, kad būtų pašalintas triukšmas ir artefaktai iš duomenų rinkinių. Iš pradžių reikia duomenų atributų tinkamumo analizės, kad būtų galima pašalinti nepageidaujamus atributus. Tada svarbiems bruožams parinkti naudojami iš anksto apdoroti duomenys, kurie gali būti naudingi prognozuojant akcijų tendencijas. Tada analizuojama sumodeliuota funkcija, siekiant apdoroti žinias naudojant didelį duomenų kiekį. Duomenų analizė suteikia patobulintą analizę ir patogią sąsają (Woolridge J., 1990). Čia duomenų analizatorius gautas detales skirsto į dvi klases, tai yra prognozavimo detales ir dabartines detales, kurios laikomos sprendimų priėmimo pagrindu. Prognozavimo detalės, esamos detalės ir ribinės vertės priklauso nuo sprendimų priėmimo sistemos. Naudojant slenkstinę vertę, atliekamas išankstinis apdorojimas deklaruoti pelną ar nuostolius. Tada pagal būseną sukuriama įspėjimo pranešimas. Jei investuotojas turi pelno, tada akcijos gali būti naudojamos pardavimų didinimui, o jei nuostolingos, tai šiai daliai skiriama daugiau dėmesio jos plėtrai, o efektyviems sprendimams priimti taikomas neuroninis tinklas (Gandhman D., Kumar K., 2019).



Pav. 2 Akcijų kainų prognozavimo technikos funkcionalumas

Šaltinis: Pagal Gandhmal D. and Kumar K., 2019

Toliau bus atlikta įvairių akcijų rinkos prognozavimo metodų apžvalga. 3 paveiksle parodytas skirtingų mašininio mokymosi akcijų rinkos prognozavimo metodų skirstymas į kategorijas. Vertybinių popierių rinkos prognozavimo metodai iš esmės skirstomi į du tipus: prognozėmis pagrįstus metodus ir klasteriais pagrįstus metodus. Būdai, pagrįsti dirbtiniais neuroniniais tinklais (angl. *Artificial Neural Network*, santr. *ANN*), konvoliuciniu neuroniniu tinklu (angl. *Convolutional Neural Networks*, santr. *CNN*), pasikartojančiu neuroniniu tinklu (angl. *Recurrent Neural network*, santr. *RNN*), sprendimų palaikymo sistema (angl. *Decision Support Systems*, santr. *DSS*), paslėptu Markovo modeliu (angl. *Hidden Markov Model*, santr. *HMM*), naiviaja Bajeso (angl. *Naive Bayes*, santr. *NB*), neuroninio tinklo (angl. *Neural Networks*, santr. *NN*), atraminių vektorių regresija (angl. *Support Vector Regression*,

santr. SVR) ir atraminių vektorių mašina (angl. Support Vector Machines, santr. SVM) yra suskirstyti į prognozėmis pagrįstus metodus. Filtravimu, neryškiais, k-vidurkiais ir optimizavimu pagrįstos technikos taip pat sugrupuotos į grupėmis pagrįstas technikas. Toks metodų klasifikavimas yra dažniausiai naudojamas mokslinėje literatūroje, ir yra pagrįstai apibendrintas Gandhman D ir Kumar K atliktame tyrime.



Pav. 3 Akcijų prognozavimo technikų klasifikacija

Šaltinis: parengta autoriaus pagal Gandhman ir Kumar, 2019.

Analizė, pagrįsta prognozavimo metodais - apima ANN, CNN, DSS, HMM, NN, RNN, SVM, SVR ir NB.

Analizė, pagrįsta klasterizacijos metodais atliekama remiantis literatūroje sukurtais klasterizavimo metodais ir apima tokias technikas kaip filtravimas (angl. *filtering*), neapibrėžtas metodus (angl. *fuzzy based*), optimizavimas (angl. *optimization*) ir K-vidurkis (angl. *K-means*).

Šis pasiskirstymas pavaizduotas žemiau esančioje lentelėje:

3 lentelė. Akcijų kainų prognozavimo technikos

Akcijų kainų prognozavimo technikos	
<u>Prognozavimu grįstos</u>	<u>Klasifikavimu grįstos</u>
Dirbtiniai neuroniniai tinklai (ANN)	Filtravimas
Konvoliucinis neuroninis tinklas (CNN)	
Sprendimų palaikymo Sistema (DSS)	Neapibrėžtas metodas
Naivusis Bajeso (NB)	
Neuroninis tinklas (NN)	Optimizavimas
Pasikartojantis neuroninis tinklas (RNN)	
Atraminių vektorių mašina (SVM)	K-vidurkis
Paslėptas Markovo modelis (HMM)	

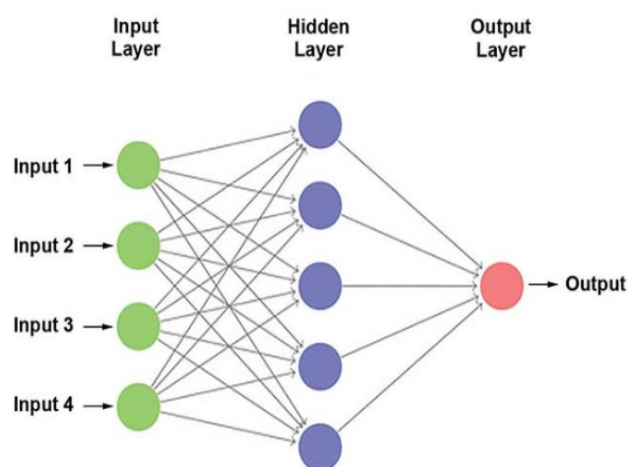
Šaltinis: parengta autoriaus pagal Gandhman ir Kumar, 2019

Toliau šiame skyriuje bus glaustai apžvelgti kiekvienas iš paminėtų metodų.

4.2.3. Prognozavimu grįstų technikų apžvalga

ANN prognozavimo technikos

ANN tiksliau už daugelį kitų statistinių metodų užfiksuoja struktūrinę akcijų rezultatų ir lemiančių veiksnių ryšį. Literatūroje atsargų grąžai prognozuoti naudojami įvairūs įvesties kintamųjų rinkiniai. Kai kurie tyrėjai iš anksto apdorojo įvesties kintamuosius, prieš pritaikydami juos ANN prognozuoti. Šiame poskyryje aprašomi įvairūs tyrimų darbai, kuriuose buvo naudojama ANN pagrįsta vertybinių popierių rinkos prognozė.



Pav. 4 Bendrinė ANN modelio veikimo schema

Šaltinis: Pagal Ralbovsky N., 2018

Ticknoras, J. L., 2013 metais sukūrė modelį, pavadintą Bayesian legalizuotu ANN, kad būtų galima prognozuoti finansų rinkos elgseną. Rinkos tendencijos ir techniniai rodikliai naudojami kaip pagrindiniai kintamieji prognozuojant kiekvienos akcijos kainą. Tikslus akcijų kainų prognozavimas yra svarbus tam, kad investuotojai galėtų padidinti akcijų grąžą. Modelis sumažina galimybes pagerinti prognozavimo ir tinklo apibendrinimo kokybę, tačiau neatsižvelgė į techninius rodiklius, kurie pagerina būsimų programų modelio kokybę. Routas, A.K ir kt. 2014, remdamasis adaptyviuoju modeliu, leidžiančiu prognozuoti finansinių laiko eilučių duomenis naudojant Indijos vertybinių popierių rinkos indeksus, sukurtas skaičiavimais efektyvus funkcinio ryšio dirbtinis neuroninis tinklas (CEFLANN). Prognozės kokybė buvo įvertinta atsižvelgiant į vidutinę absoliučią procentinę paklaidą (MAPE), naudojant akcijų indeksus. CEFLANN sąnaudos parenkamos iš ankstesnių akcijų rodiklių, naudojant skirtingus rinkos sektorius ir techninius rodiklius, siekiant nustatyti optimalias akcijų prognozavimo tendencijas. Be to, optimalius svorius parenka adaptyvus diferencinės evoliucijos (DE) mechanizmas.

Zhongas, X. ir Enke'as, D, 2017 metais sukūrė efektyvų mechanizmą, pagrįstą duomenų gavybos strategijomis, kad būtų galima prognozuoti „S&P 500“ indekso grąžos dienos tendencijas, naudojant ekonomines ypatybes. Prognozavimui naudojami trys metodai, įskaitant apytikslę tvirtą pagrindinių komponentų analizę (FRPCA), branduoliu pagrįstą pagrindinių komponentų analizę (KPCA) ir pagrindinių komponentų analizę (PCA), pagrįstus matmenų mažinimu. Norint numatyti, reikalingos pažangios branduolio funkcijos ir atitinkami branduolio parametrai.

Hadavandi, E. ir kt. 2018 metais sukūrė integruotą metodą, pagrįstą „Genetic Fuzzy Systems“ (GFS) ir ANN, kad sukonstruotų įgudusią sistemą, pagrįstą akcijų kainų prognozavimu. Tuo metu jau

pradėjo plisti hibridinių modelių pritaikomumo tyrimai (kelių metodų sujungimas į vieną). Iš pradžių, norint nustatyti veiksnius, kurie turi turėti įtakos akcijų kainoms, buvo pritaikyta laipsniškos regresijos analizė (SRA). Be to, pirminiai duomenys buvo suskirstyti į grupes, naudojant „Self-Organizing Map“ (SOM) NN. Pagaliau klasteriams taikomi GFS modeliai, naudojant taisyklių bazės išskyrimą ir duomenų bazių derinimą.

Patel, J. ir kt. 2015 metais prognozių modeliuose, skirtame atsargoms analizuoti, sukūrė sluoksnį (*angl. layer*), pavadintą „Trend Deterministic Data Preparation Layer“. Metodas apskaičiuoja dešimt techninių parametrų, naudojant akcijų prekybos duomenis, ir pateikia gautus techninius rodiklius kaip tendenciją lemiančius duomenis. Metodas buvo sutelktas į trumpą laiko prognozavimą vertinant akcijų gražos vertes. Be to, metodui trūksta ilgalaikio prognozavimo, kad būtų galima analizuoti ketvirčio atsargų prognozes, pelno gražą, įmonės organizacinį stabilumą ir pajamas. Šis metodas prognozuoja ankstesnes akcijų kainas. Metodas atsižvelgia į NASDAQ dienos akcijų pokyčius, kad būtų galima patikrinti modelio prognozavimo galimybes. Guresenas, E ir kt. 2011 metais sukūrė modelį, pavadintą „Generalized Autoregressive Conditional Heteroscedasticity“ (GARCH), skirtą vertinti NN vertinant akcijų rinką. Modelis įvertina hibridinius neuroninius tinklus, dinaminį dirbtinį neuroninį tinklą (DAN2) ir daugiasluoksnį suvokimą (*angl. multilayer perception*) (MLP), kad išgautų įvesties kintamuosius. Tačiau, kol kas nepavyko patvirtinti, ar GARCH, E-GARCH pateikia teisingas prognozes pagal koreliacinius kintamuosius.

CNN prognozavimo technikos (Convolutional neural network)

CNN yra „feed-forward“ nervinis tinklas. CNN paslėptų sluoksnių yra daugiau nei įprastame neuronų tinkle. CNN yra labiau žinomi kaip gilus mokymosi algoritmai, naudojami akcijų rinkoms prognozuoti. Nors mokslinėje literatūroje yra mažiau tyrimų apie CNN nei apie ANN, tačiau keli jų visgi buvo rasti. Vargas, M. R. ir kt. naudojo gilus mokymosi mechanizmą, kad būtų galima numatyti akcijų kainos krypties judėjimą, naudojant „Standard & Poor’s 500“ indeksą su techniniais rodikliais. Gilūs mokymosi metodai naudojami nustatant ir analizuojant sudėtingus duomenų modelius ir leidžiančius paspartinti prekybos procesą. Metodui nepavyko pritaikyti sustiprinimo mokymosi algoritmų mokyti modelį apie rinkos modeliavimą. Zhou, X ir kt. 2018 metais sukūrė bendrą sistemą, pritaikydama LSTM ir CNN mokymui, kad būtų galima numatyti dažnas akcijų rinkos tendencijas. Ši technika imituoja investuotojo prekybos būdą ir naudoja mokymų ir testavimo rinkinius, kad būtų galima analizuoti atnaujinto modelio ciklo poveikį prognozuojant rezultatus. Be to, metodui trūksta kitų nuspėjamųjų modelių daugialypėmis sąlygomis. Xu, B ir kt. sukūrė pasikartojantį konvoliucinį neuronų tinklą akcijų rinkos tendencijoms prognozuoti. Tinklas užfiksavo esminę vertybinių popierių rinkos informaciją naudodamas reikšmingą funkciją. Pirmajame sluoksnyje mokymosi procesui automatizuoti buvo

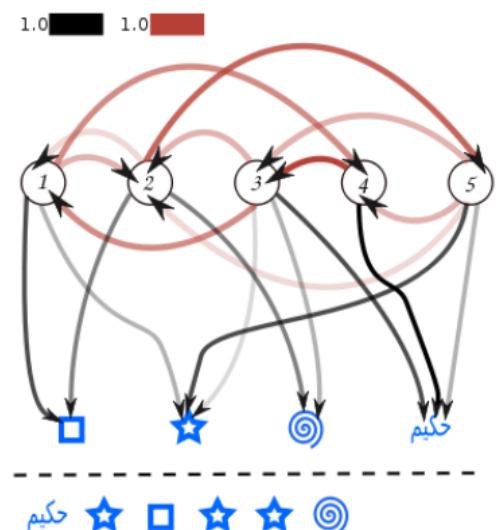
naudojamas objekto įterpimo sluoksnis. Atliekant šį metodą nebuvo atsižvelgta į finansines žinias, siekiant optimizuoti akcijų tendencijų prognozavimo modelį.

DSS prognozavimo technikos (Decision Support Systems)

DSS gali numatyti akcijų kainų pokyčius, kurių reikalauja investuotojai akcijų rinkoje. Wen, Q ir kt. 2010 metais sukūrė pažangią intelektualią prekybos sistemą, naudodama svyravimo dėžės prognozavimą, integruodama SVM algoritmą su akcijų dėžutės teorija. Dėžutės teorija reiškia, kad akcijų pirkimas ar mažmeninė prekyba yra sėkminga, jei kaina sutrikdo pradinę ribinę vertę, palyginti su kitomis dėžėmis. Prekybos mechanizmas, naudojant dvi privalomas prognozes, yra sukurtas efektyviems sprendimams priimti. Metodas nesugebėjo sukurti pažangių patikimų įvertinimų, kurie padėtų pagerinti prognozių tikslumą taikant papildomus minkštojo skaičiavimo metodus.

HMM pagrįstos prognozavimo technikos (Hidden Markov Model)

HMM pradėtas neseniai taikyti prognozuojant akcijų rinką. HMM sėkmingai analizuoja ir prognozuoja laiką, priklausantį nuo reiškinų ar laiko eilučių. Badge, J., sukūrė įvairius makroekonominis veiksniai Indijos akcijų rinkai su skirtingais makroekonominiais veiksniais, pavyzdžiui, techninius rodiklius. Šie techniniai rodikliai naudojami nustatant rinkos modelius konkrečiu laiku. Yra daugybė techninių rodiklių, leidžiančių gauti efektyvią prognozę taikant pagrindinio komponento analizę (PBS). Vertybinių popierių rinkos prognozavimui naudojami duomenys yra pasirinkti techniniai rodikliai. HMM pritaikytas būsimoms kainoms nustatyti ir laikomas dominuojančiu stochastiniu modeliu. Gupta, A. ir Dhingra, B. 2012 metais sukūrė „Posteriori HMM“ metodą, skirtą prognozuoti akcijų etiką, panaudojant ankstesnius duomenis. Šis metodas atsižvelgia į dalinius akcijų indeksų svyravimus HMM mokymui. Taip pat HMM naudojamas maksimaliam „Posteriori“ sprendimui priimti pagal akcijų vertes. Metodas neatsižvelgė į koreliacijas kuriant modelį. Našumas pagerinamas naudojant kvantavimą, atsižvelgiant į akcijų vertę valandą po valandos ir minutę.



Pav. 5 Bendrinė HMM modelio veikimo schema

Šaltinis: Pagal Gupta A., 2012

Naivioji Bayeso prognozavimo technika

Naivusis Bayeso algoritmas yra klasifikavimo metodas, generuojantis Bayeso tinklus tam tikram duomenų rinkiniui, remiantis Bayeso teorema. Manoma, kad nurodytame duomenų rinkinyje yra tam

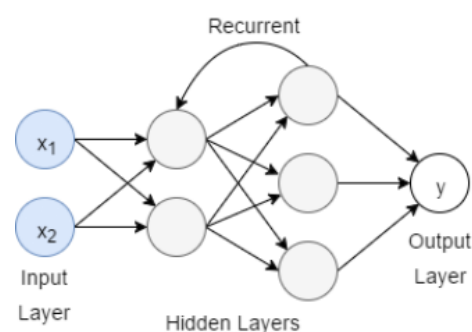
tikras klasės bruožas, kuris nėra susijęs su jokia kita funkcija. Naivų Bayeso algoritmą lengva sukurti, jis naudingas labai dideliems duomenų rinkiniams ir lenkia labai sudėtingus klasifikavimo metodus. More, A.M ir kt. 2018 metais informacijai apie atsargas gauti ir akcijų grafikams nustatyti naudojo neurolingvistinio programavimo (NLP) metodą, kuris gali padėti vartotojams nustatyti tinkamas investicijas su dideliu pelnu. „Hadoop“ sistema buvo naudojama norint pasiekti daugybę akcijų duomenų lygiagrečiai, o „Naive Bayes“ algoritmas buvo naudojamas priimant sprendimus, susijusius su tikimybės trukdžiais, ir metodą, kuris naudojo išankstines žinias ateities tendencijoms prognozuoti.

NN pagrįstos prognozavimo technikos (Nearest Neighbour)

NN yra algoritmų serija, atpažįstanti pirminius duomenų rinkinio ryšius per procesą, imituojanti žmogaus smegenų veikimą. Chang, P.C 2012 metais sukūrė modelį, pavadintą „Evolving Partially Connected Neural Network“ (EPCNN), kad prognozuotų akcijų tendencijas, pagal techninius rodiklius. Modelio architektūra turi pažangias funkcijas, naudojant skirtingas ANN funkcijas, kuriose ryšys tarp neuronų yra atsitiktinis ir gali būti modeliuojami keli sluoksniai. Evoliucinis algoritmas pritaikytas mokymosi algoritmui ir treniruočių svoriui tobulinti. EPCNN naudoja atsitiktinius ryšius tarp neuronų, kad sužinotų ankstesnių laiko eilučių duomenis, kad pagerintų nervų tinklus. Metodas neįtraukė pažangių minkštojo skaičiavimo metodų, kaip tvarkyti kitus laiko eilučių duomenis. Pang, X ir kt. sukūrė novatorišką NN metodą, kad pasiektų patobulintas vertybinių popierių rinkos prognozes. Asadi, S ir kt. 2012 metais sukūrė hibridinį intelektualųjį modelį vertybinių popierių biržoms prognozuoti. Hibridinis modelis yra GA, duomenų perdirbimo mechanizmų ir Levenbergo – Marquardo (LM) algoritmo integravimas norint sužinoti NN. Pradinis NN svoris naudojamas sureguliuojant LM algoritmą pritaikant GA. Išankstinio duomenų apdorojimo mechanizmuose yra įvesties kintamųjų pasirinkimas ir duomenų transformavimas, siekiant pagerinti bendrą modelio tikslumą. Metodas išbandomas taikant akcijų biržos indeksus akcijų tendencijoms prognozuoti.

RNN pagrįstos prognozavimo technikos (Recurrent neural network)

RNN yra ANN klasė, kurioje ryšiai tarp mazgų sudaro nukreiptą grafiką pagal laiko seką. Tai leidžia parodyti dinamišką laiko elgesį. Hsieh, T.J ir kt. 2012 metais sukūrė vieningą sistemą, kurioje RNN pagrįstas dirbtinės bičių kolonijos (ABC-RNN) algoritmu, kuris yra integruotas akcijų kainų prognozavimui. Modelyje yra trys etapai, kuriuose iš pradžių „Haar“ bangos yra naudojamos akcijų kainos laiko eilutės duomenims skaidyti ir artefaktams bei triukšmams pašalinti. Antra, RNN buvo naudojamas kuriant įvesties ypatybes



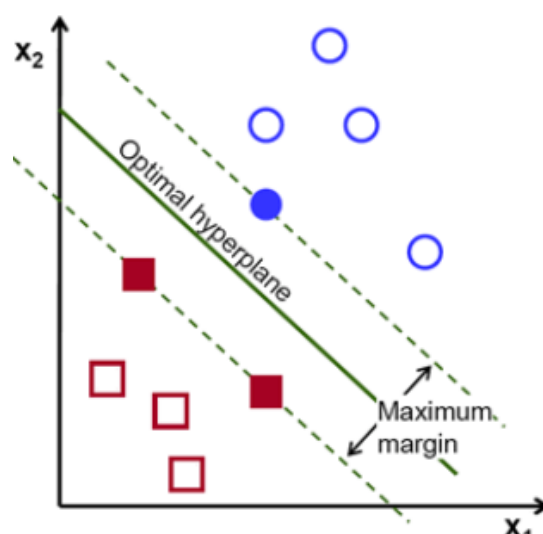
Pav. 6 Bendrinė RNN modelio veikimo schema

Šaltinis: Pagal Hsieh T., 2011

naudojant laipsniško regresijos ir koreliacijos pasirinkimą (SRCS). Trečia, ABC buvo pritaikytas optimizuoti RNN svorį ir poslinkius projektuojant parametrus. Tačiau metodui trūksta pažangaus modelio pasirinkimo mechanizmo, kad būtų galima gauti esminius modelius iš duomenų. Xie, X. K. ir Wangas, H 2017 metais sukūrė RNN laiko eilučių duomenims analizuoti. Duomenų rinkinį sudaro didžiulis dienos duomenų kiekis iš Kinijos Šanchajaus „Shenzhen 300“ indekso. RNN buvo naudojamas kasdieninėms funkcijoms klasifikuoti naudojant dienos duomenis. Modelio našumas buvo analizuojamas naudojant tikslumą ir vidutinį pelną.

SVM pagrįstos prognozavimo technikos (Support Vector Machine)

SVM yra prižiūrimas mokymosi modelis su susijusiais mokymosi algoritmais, analizuojančiais klasifikacijai ir regresijos analizei naudojamus duomenis. SVM yra galinga prognozavimo priemonė vertybinių popierių prognozėms finansų rinkoje. Oztekinas, A 2016 metais pristatė bendrą prognozavimo modelį, kuris leistų prognozuoti kasdienio atsargų judėjimo kainą, derinant tris prognozavimo modelius, apimančius adaptyvias neurofuzzy išvadų sistemas, ANN ir SVM, pagrįstus duomenų analize. Modelis buvo patikrintas naudojant dešimterioją stratifikuotą kryžminį patvirtinimą, siekiant kuo tiksliau sumažinti atsitiktinių imčių šališkumą. Tačiau metodas neatsižvelgė į kitas akcijų rinkas, kad būtų galima



Pav. 7 Bendrinė SVM modelio veikimo schema

Šaltinis: Gandhi R., 2018

patikrinti metodą skirtingose šalyse. Poršnevas, A ir kt. sukūrė metodą, pavadintą leksikos principu, siekiant pagerinti akcijų rinkos rodiklių tikslumą, naudojant „Twitter“ vartotojų mąstyseną. Mąstysenos analizė atskleidžia aštuonių svarbių emocijų egzistavimą daugiau nei 755 milijonuose „Twitter“ pranešimų. Be to, SVM pagrįsti metodai taikomi numatant DJIA ir S & P500 rodiklius. Tačiau metodas nesugebėjo maksimaliai padidinti mokymo laikotarpį ir patobulinti nuotaikos analizės algoritmų. Zhang, X ir kt. 2018 metais pristatė kelių šaltinių bei egzempliorių modelį, kuris gali integruoti nuotaikas, įvykius ir kiekybiškai įvertinamus duomenis į įtraukiančią sistemų visumą. Naujienų įvykių fiksavimui naudojami ištraukimo ir pateikimo metodai. Naujienų įvykiai ir kiekybiniai duomenys daro įtaką akcijų svyravimams, tuo pačiu prognozuodami akcijų rinkos tendencijas. Dėl tos priežasties šis modelis laikomas vienu iš stipriausių, nes geba įtraukti labai skirtingus kintamuosius.

SVR pagrįstos prognozavimo technikos (Support Vector Regression)

SVR klasifikavimui naudojami tie patys principai, kaip ir SVM, turint tik keletą nedidelių skirtumų. (Ghughe C., 2018). Kazem, A ir kt. 2013 metais, remdamasis „Firefly“ algoritmu, SVR ir chaotišku žemėlapiu, sukūrė prognozavimo modelį, leidžiantį prognozuoti akcijų rinkos kainas. Prognozavimo modelis susideda iš trijų fazių. Pirmajame etape nagrinėjamas uždelsimo koordinacių įterpimo metodas bendrajai dinamikai rekonstruoti. Antrame etape, siekiant optimizuoti SVR parametrus, buvo pritaikytas chaotiškas „Firefly“ algoritmas. Galiausiai, optimizuotas SVR yra pritaikytas prognozuoti akcijų rinkų kainas. Be to, metodas naudojo struktūrinės rizikos mažinimą (SRM) mokant SVR procesą, kad metodas būtų patikimesnis nei esami modeliai. Kao, L.J ir kt. sukūrė atsargų prognozavimo modelį, kuris panaudojo netiesinę nepriklausomą komponentų analizę (NLICA) kaip išankstinį apdorojimą funkcijoms iš kintamųjų išgauti. Savybės, pavadintos „Independent Components“ (IC), yra naudojamos kaip SVR įvestis kuriant prognozavimo modelius. Šis metodas gali nustatyti paslėptą informaciją iš pradinių duomenų, naudojant funkcijų išgavimo procesą. Taigi NLICA padeda pateikti vertingos informacijos finansiniam apdorojimui. Xiong, T ir kt. sukūrė algoritmą, pavadintą „Firefly“ algoritmu pagrįstą daugialypės išvesties palaikymo vektorinę regresiją (FA-MSVR), kad nustatytų MSVR parametrus. Vertinimas atliekamas naudojant ekonominius kriterijus, skaičiavimo išlaidas ir statistinius kriterijus. Remiantis ekonominiais kriterijais, rezultatai vertinami remiantis paprastu prekybos mechanizmu, remiantis santykiine prognoze. Remiantis statistiniais kriterijais, prognozavimas atliekamas naudojant prognozavimo priemones ir bandymo metodus. Tačiau metodas neatsižvelgė į kitus vertinamus laiko eilučių duomenis, kad prognozuotų užduotis, įskaitant valiutos kursą.

4.2.4. Klasifikacija grįstų technikų apžvalga

Filtravimo pagrįstos prognozavimo technikos (Filtering based)

Filtravimas pritraukė daug dėmesio ir turėjo įvairias galimas prognozavimo programas. Arévalo, R ir kt. 2017 metais sukūrė prekybos taisykles, naudodamas vėliavos modelio atpažinimą, ir įtraukė esmines naujoves, remdamasis ankstesniais tyrimais. Iš pradžių sukurta dinaminė langų schema, leidžianti sustabdyti nuostolius ir periodiškai gauti atnaujintą pelną. Be to, vėliavos modelis atitinka dabartines tendencijas, taigi, norint filtruoti sandorius, buvo pridėtas rodiklis, pavadintas EMA. Techniniai rodikliai apskaičiuojami naudojant 15 minučių ir 1 dienos laikotarpius, palengvinančius vienu metu apdorojamus trumpus ir vidutinius terminus. Tačiau, parengus tam tikras prekybos taisykles, modelis nesugebėjo išspręsti duomenų analizės problemų. Ariyo, A.A ir kt. 2014 metais pristatė prognozavimo modelį, naudodamasis ARIMA modeliu akcijų rinkos prognozavimui. Skelbiami akcijų duomenys yra gauti iš Nigerijos vertybinių popierių biržos (NSE) ir Niujorko vertybinių popierių biržos (NYSE). ARIMA modelių rezultatai naudojami su pažangiomis prognozavimo metodikomis trumpam

laiko prognozavimui. Srinivasanas, P. ir Ibrahimas, P 2010 metais sukūrė prognozavimo modelį, paremtą GARCH modeliu. Šis modelis buvo pritaikytas prognozuoti tam tikro laikotarpio SENSEX indekso nepastovumą. Naudojant našumo metriką, GARCH modelis rodo geriausius rezultatus, numatant SENSEX indekso grąžą.

Apytikslio prognozavimo metodikos (Fuzzy based)

Apytikslė logika yra daugelio vertinamos logikos forma, kurioje kintamųjų tiesos reikšmės gali būti bet kokie realūs skaičiai nuo 0 iki 1. Esfahanipouras, A. ir Aghamiri, W sukūrė „Neuro-Fuzzy Inference System“ akcijų kainoms prognozuoti. Apytiksliaame modelyje kaip įvestis buvo naudojami techniniai rodikliai, o gaunama dalis yra tiesinis įvesties derinys. „Fuzzy C-Means“ (FCM) grupavimas buvo pritaikytas siekiant nustatyti taisyklių rinkinį. Narystės funkcija buvo aprašyta naudojant Gauso funkciją. ANFIS tikslas yra numatyti akcijų kainų kitimą naudojant kelis aspektus, tokius kaip politinės priežastys, makroekonominiai pokyčiai, techninis indeksas ir fundamentali analizė. Enke, D ir kt. sukūrė trijų fazių vertybinių popierių rinkos prognozavimo sistemą, kurioje pirmasis etapas apima daugialypę regresijos analizę, kad būtų galima apibrėžti finansinius kintamuosius. Antrajame etape diferencinė evoliucija pagrįsta 2 tipo neryškia klasteriais pritaikytą prognozavimo modelį sukurti. Trečiajame etape „Fuzzy 2“ tipo neuroninis tinklas pritaikytas būsimų akcijų kainų prognozavimo argumentams atlikti. Boyacioglu, M. A. ir Avci, D 2010 metais pristatė gan paplitusį modelį - akcijų rinkos grąžos prognozavimo modelį, pavadintą „Adaptive Network-Based Fuzzy Inference System“ (ANFIS).

Chenas, M.Y. ir Chenas, B. T. sukūrė pažangų neaiškių laiko eilučių modelį akcijų rinkos kainoms prognozuoti. Modelis buvo sukurtas remiantis granuliuotu skaičiavimo metodu, naudojant entropialinį diskretizavimą ir susiejimo mechanizmus. Modelis pritaikė neapibrėžtų laiko eilučių modelius akcijų indeksams prognozuoti. Metodo metu nebuvo atsižvelgta į kitą esminį veiksni, įskaitant finansines ataskaitas, techninius rodiklius ir prekybos apimtį. Wei, L.Y ir kt. 2011 metais sukūrė sistemą, pavadintą ANFIS, kurios akcijų kainos tendencijoms prognozuoti buvo naudojami keli techniniai rodikliai. Iš pradžių reikalingi techniniai rodikliai parenkami naudojant pažangius rodiklius, atsižvelgiant į koreliacijos matricą. Antra, grupavimo mechanizmas pritaikytas techniniams rodikliams suskirstyti naudojant duomenų diskretizavimą. Trečia, „Fuzzy Inference System“ (FIS) naudojama kalbinių terminų taisyklėms išgauti iš techninių rodiklių ir optimizuoti FIS parametrus naudojant adaptyvius tinklus, kad būtų galima tiksliai prognozuoti. Tačiau metodas nebuvo taikomas prognozuojant kitus akcijų indeksus.

K-vidurkio klasterizavimo metodas (K-means)

K-klasterių klasterizavimas yra vektorių kvantavimo technika, madinga atliekant klasterinę analizę duomenų gavyboje. Nanda, S.R. ir kt. 2010 metais sukūrė duomenų gavybos mechanizmą, skirtą

atsargoms klasifikuoti į grupes. Baigus klasifikuoti, iš kiekvienos grupės portfeliui sudaryti buvo pasirinktos akcijos. Rizikos mažinimo kriterijus pasiekiamas diversifikuojant portfelį. Grupavimo mechanizmas dalija akcijas pagal konkrečius investavimo kriterijus. Akcijų graža naudojama kartu su GSE duomenų rinkinio vertinimo rodikliais. Grupių sudarymo metodas buvo įvestas valdant portfelį ir renkantis akcijas, kad būtų pasiekta efektyvi riba.

Optimizavimu pagrįstos grupavimo technikos (Optimization)

Optimizavimas - tai sprendimas, kuriuo siekiama kuo tobulesnio, funkcionalesnio ar efektyvesnio rezultato. Tyrimai, naudojant optimizavimo akcijų rinkos prognozavimo sistemą, paaiškinti šiame poskyryje. Chengas, C.H. ir kt. 2010 metais sukūrė hibridinį prognozavimo modelį, kuris naudojo kelis techninius rodiklius akcijų kainų tendencijoms prognozuoti. Modelyje buvo naudojami keturi hibridinio modelio algoritmai, skirti veiksmingoms taisyklėms prognozuoti ir naudoti išgautas taisykles kartu su palaikomosiomis vertėmis, naudojant neapibrėžtų rinkinių teoriją. Pirmasis buvo techniniai rodikliai, kurie naudojami prognozuojant būsimas akcijų kainas pagal koreliacijos matricas. Antrasis naudojamas metodas buvo kaupiamasis tikimybės paskirstymo metodas (CDPA), skirtas techniniams rodikliams suskirstyti, o trečiasis - neapdorotų rinkinių teorijos (RST) algoritmas, skirtas kalbos taisyklėms išgauti, ir ketvirtasis - GA, siekiant patobulinti ištrauktas taisykles, kad būtų užtikrintas geresnis prognozavimo tikslumas. Tačiau metodas neatsižvelgė į kitus duomenų diskretizavimo metodus ir dirbtinio intelekto algoritmus, kad gautų patobulintą prognozavimo procesą. Remiantis metodu, yra sugalvoti įvairių hiperparametrų rinkinių pranašumai, siekiant pagerinti bendrą sistemos veikimą. Be to, bandymai ir klaidos buvo naudojami norint nustatyti tinkamus hiperparametrų parametrus.

Atlikus sistemine literatūros analizę, galima teigti, jog, nors ir egzistuoja gan daug modelių, visi turi tam tikrų trūkumų ir nėra iki galo išdirbti. Todėl vis dar vyksta tyrimai, bandymai apjungti ir integruoti skirtingas technikas, kad pasiektume kuo tikslesnius prognozavimo rezultatus. Prognozuojant vertybinių popierių rinkos metodus, naudojami ANN, SVM, SVR, HMM, NN, neryškūs metodai, K-priemonės ir kiti. Efektyviam vertybinių popierių rinkos prognozavimui pasiekti dažniausiai naudojama technika yra „ANN“ ir „fuzzy“ pagrįsta technika. Šiuos metodus galima efektyviai naudoti kontroliuojant ir stebint visą akcijų rinką. Pagrindinis iššūkis, su kuriuo susiduria akcijų kainų prognozavimo sistemos, yra tai, kad daugelio esamų metodų neįmanoma aptikti naudojant istorinius akcijų duomenis, nes jiems įtakos turi tam tikri veiksniai, susiję su vyriausybės politikos sprendimais, rinkos nuostatomis ir pan. Taigi sprendimams priimti reikalingi duomenys iš skirtingų šaltinių, o duomenų išankstinis apdorojimas yra sudėtinga duomenų gavybos užduotis. Tai yra pagrindiniai apribojimai, kuriuos reikia spręsti ateityje pritaikant pažangias vertybinių popierių rinkos prognozavimo

metodikas. Atlikta analizė taip pat parodė, jog daugiausiai tyrimų yra atliekama prognozavimui pritaikant ANN modelius, po jų seka SVM, kitų modelių taikymas yra gana retesnis. Be to, ankstesni tyrimai rodo, kad tiksliausiai prognozuoti geba būtent šie du modeliai, kadangi pasižymi gan sudėtingu ir ilgu algoritmu. Kiti minėti metodai taip pat yra tinkami akcijų kainai prognozuoti, tačiau tam, kad pasiekti modelio tikslumo, esamose tyrimuose jie naudojami ne kaip pavieniai metodai, o jungiami tarpusavyje. Kadangi sisteminė mokslinių šaltinių analizė parodė, jog daugiausiai potencialo turi ANN ir SVM metodai, jie ir bus nagrinėjami bei taikomi šiame tyrime.

4.3. Dirbtinio intelekto metodų taikymo akcijų kainai prognozei ypatumai

Numatyti akcijų kainas ir akcijų indeksus ir jų pokyčius vis dar išlieka viena iš sudėtingiausių laiko eilučių prognozavimo taikymo sričių. Nors buvo atlikta daug empirinių tyrimų, kurie nagrinėja akcijų kainų prognozavimo klausimus, dauguma empirinių išvadų siejamos su išsivysčiusiomis finansų rinkomis. Tačiau literatūroje yra nedaug tyrimų, leidžiančių numatyti akcijų kainų judėjimo kryptį besivystančiose rinkose, pagrindinė to priežastis yra ribotas duomenų prieinamumas. Tikslios akcijų kainų bei jų judėjimo prognozės yra labai svarbios kuriant efektyvias rinkos prekybos strategijas (Leung, Daouk ir Chen, 2000). Taigi investuotojai gali apsidrausti nuo galimos rinkos rizikos, o spekuliantai ir arbitrai turi galimybių gauti pelno prekiaudami akcijų indeksu (Manish & Thenmozhi, 2005). Akcijų rinkos prognozavimas laikomas sudėtingu finansinių laiko eilučių prognozavimo proceso uždaviniu, nes akcijų rinka iš esmės yra dinamiška, netiesinė, komplikuota, neparimetrinė ir chaotiška (Abu-Mostafa ir Atiya, 1996). Be to, akcijų rinką veikia daugybė makroekonominių veiksnių, tokių kaip politiniai įvykiai, firmų politika, bendros ekonominės sąlygos, investuotojų lūkesčiai, institucinių investuotojų pasirinkimai, kitos akcijų rinkos judėjimas, investuotojų psichologija ir kt. (Tan, Quek, 2007).

ANN ir SVM buvo sėkmingai naudojamos finansinių laiko eilučių modeliavimui ir prognozavimui (Leung, Daouk ir Chen, 2000). Nors ANN gali būti viena iš labai naudingų įrankių prognozuojant laiko eilutes, keli tyrimai parodė, kad ANN turėjo tam tikrų apribojimų mokytis modelių, nes akcijų rinkos duomenys turi didžiulį triukšmą, nestacionarias savybes ir sudėtingą matmenį. ANN dažnai rodo nenuoseklų ir nenuspėjamą triukšmingų duomenų pasirodymą (Kim, 2003; Kim & Han, 2000; Manish & Thenmozhi, 2005). Todėl prognozuoti akcijų kainų pokyčius yra gana sunku. Palaikomosios vektorinės mašinos (SVM) ir dirbtiniai neuroniniai tinklai (ANN) yra mašininio mokymosi metodai, kurie gali plačiai pritaikyti tiek inžinerijos, tiek socialinių mokslų srityje. Finansuose dirbtiniai neuroniniai tinklai (ANN) ir palaikomojo vektoriaus mašinos (SVM) yra naudojami prognozuoti akcijų rinkos tendencijas: pakilimus ir nuosmukius. Tikslas yra iširti tam tikrų parametrų, pvz. abiejų klasifikatorių (ANN ir SVM), panašių duomenų rinkinių, skirtumų prognozuojant akcijų rinkos tendencijas, kitimą. Kadangi minėtuose tyrimuose ANN ir SVM metodai rodo gan tikslus

ir reikšmingus prognozavimo rezultatus, toliau tyrime bus pateikti skirtingi SVM ir ANN akcijų prognozavimo problemos modeliai, siekiant gauti kuo tikslesnius rezultatus, skirtingiems duomenų rinkinių deriniams, gautiems iš įvairių įmonių istorinių duomenų ir lyginamosios analizės.

Atlikus mokslinių šaltinių sisteminę analizę, galima daryti išvadą, jog šie du mašininio mokymosi metodai – ANN ir SVM – gali būti vadinami populiariausi bei daugiausiai ištirti. Toliau šiame moksliniame darbe šie modeliai bus smulkiau nagrinėjami bei tiriami, siekiant nustatyti abiejų pranašumus bei trūkumus prognozuojant akcijų kainų svyravimą.

4.3.1. Dirbtiniai Neuroniniai Tinklai (ANN) bei modelio ypatumai

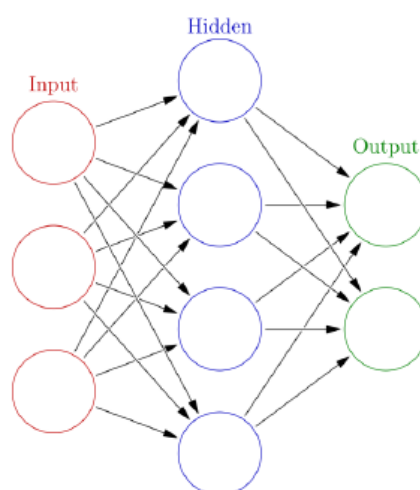
Egzistuoja didžiulis literatūros kiekis, orientuotas į akcijų rinkos nuspėjamumą bei prognozavimą. Tyrimuose naudojami įvairūs ANN tipai, siekiant tiksliai numatyti akcijų kainos grąžą ir jos judėjimo kryptį. Įrodyta, kad ANN teikia perspektyvius rezultatus prognozuojant akcijų kainos grąžą (Avci, 2007; Egeli, Ozturan ir Badur, 2003; Karaatli, Gungor, Demir ir Kalayci, 2005; Kimoto, Asakawa, Yoda ir Takeoka, 1990; Olson & Mossman, 2003; White, 1988; Yoon & Swales, 1991). Leung ir kt. (2000) išnagrinėjo įvairius prognozavimo modelius, pagrįstus daugiamatės klasifikacijos metodais, ir palygino juos su daugeliu parametrinių ir neparimetrinių modelių, kurie prognozuoja indekso grąžos kryptį. Empirinis eksperimentas parodė, kad klasifikavimo modeliai (diskriminantų analizė ir tikimybinis neuroninis tinklas) lenkia lygio vertinimo modelius (adaptyvus eksponentinis išlyginimas, vektorinė automatinė regresija su Kalmano filtro atnaujinimu, daugiamatė perdavimo funkcija ir daugiasluoksnis perdavimo neuroninis tinklas) numatant vertybinių popierių rinkos judėjimo kryptį ir maksimaliai padidinant investicijų prekybos grąžą. Chen ir kt. (2003) bandė numatyti Taivano vertybinių popierių biržos indekso grąžos kryptį. Tikimybinis neuroninis tinklas (PNN) naudojamas indekso grąžos kryptį prognozuoti. Statistiniai PNN prognozių rezultatai lyginami su apibendrintais momentų metodais (GMM) su Kalmano filtru ir atsitiktiniu ėjimu. Empiriniai rezultatai parodė, kad PNN rodo stipresnę prognozavimo galią nei GMM – Kalmano filtras ir atsitiktinio ėjimo prognozavimo modeliai. Diler (2003), remdamasis įvairiais techniniais rodikliais, mokė neuroninius tinklus įvertinti ISE 100 indekso kryptį. Techninis naudojami rodikliai: MA, impulsas, RSI, stochastika K%, slenkamasis vidutinis konvergencijos-divergencijos (MACD) rodiklis. Tyrimo rezultatai parodė, kad ISE 100 indekso kryptis gali būti prognozuojama 60,81% norma. Altay ir Satman (2005) palygino prognozuojamus neuroninių tinklų modelių rodiklius su įprastos mažiausios aikštės (OLS) regresijos modeliu ISE-30 ir ISE-All indeksams. Nors neuroninių tinklų modelių prognozavimo efektyvumas dienos ir mėnesio duomenims nepralenkė linijinės regresijos modelio, šie modeliai geba tiksliau numatyti indeksų kryptį. Cao, Leggio ir Schniederjans (2005) siekė parodyti ANN tikslumą, numatant Šanchajaus vertybinių popierių biržoje (SHSE) prekyjamų įmonių akcijų kainų pokyčius. Jie palygino kapitalo turto kainodaros modelį (CAPM) ir „Fama“ bei „French“ trijų veiksnių modelį su vienmačio ir

daugiamačio neuroninio tinklo modelių prognozuojamąją galia. Jų rezultatai parodė, kad neuroniniai tinklai lenkia palygintus linijinius modelius.

Kai kurie tyrimai yra linkę hibridizuoti keletą dirbtinio intelekto (AI) metodų, kad būtų galima numatyti akcijų rinkos gražą (Baba & Kozaki, 1992; Chu, Chen, Cheng ir Huang, 2009; Hiemstra, 1995; Kim & Chun, 1998; Leigh, Purvis ir Ragusa, 2002; Oh & Kim, 2002; Pai & Lin, 2005; Saad, Prokhorov ir Wunsch, 1998; Takahashi, Tamada ir Nagasaka, 1998; Tan ir kt., 2007; Yudong & Lenan, 2009). Tsaih, Hsu ir Lai (1998) taikė hibridinį AI metodą, norėdami numatyti S&P 500 akcijų indekso ateities sandorių dienos kainų pokyčių kryptį. Hibridinis AI metodas integravo taisyklėmis pagrįstas sistemas ir neuroninių tinklų techniką. Empiriniai rezultatai parodė, kad samprotaujantys neuroniniai tinklai (RN) lenkia kitus du ANN modelius (nugaros sklidimo tinklai ir perceptronas). Empiriniai rezultatai taip pat patvirtino, kad integruota ateities prekybos sistema (IFTS) lenkia pasyvią pirkimo ir laikymo investavimo strategiją.

Sąvoka „neuroninis tinklas“ kilo iš bandymų surasti matematinius informacijos apdorojimo vaizdus biologinėse sistemose (Bishop, 1995).

Dirbtinis neuronas (AN) yra biologinio neurono (BN) modelis. Kiekvienas AN gauna signalus iš aplinkos, arba kiti AN, surenka šiuos signalus ir, šaudydamas, perduoda signalą visiems prijungtiems AN. 10 paveiksle pavaizduotas dirbtinis neuroninis tinklas. Įvesties signalai yra slopinami arba sužadunami naudojant neigiamą ir teigiamą skaitinį svorį, susietą su kiekvienu ryšiu su AN. AN veikimas ir išeinančio signalo stiprumas valdomi per funkciją, vadinamą įjungimo funkcija. AN surenka visus gaunamus signalus ir apskaičiuoja grynąjį įvesties signalą kaip atitinkamų svorių funkciją. Grynasis įvesties signalas yra įjungimo funkcijos įvestis, kuri apskaičiuoja AN išėjimo signalą (EngelBrecht, 2007).



Pav. 8 Dirbtiniai Neuroniniai tinklai (ANN)

Šaltinis: Pateikta Shrialve H., 2018

ANN komponentai

Paženklintas neuronas j , gaunantis įvestį $p_j(t)$ iš pirminių neuronų, susideda iš šių komponentų:

- įjungimo $a_j(t)$, priklausomai nuo atskiro laiko parametro,
- galbūt slenkstis (riba) Θ_j , kuris lieka fiksuotas, nebent tai pakeičia mokymosi funkcija
- aktyvinimo funkcija f , kuri apskaičiuoja naują aktyvumą tam tikru laiku $t + 1$ iš $a_j(t)$, Θ_j ir grynosios įvesties $p_j(t)$, dėl kurios atsiranda ryšys:

$$a_j(t + 1) = f(a_j(t), p_j(t), \Theta_j)$$

Formulė 1 ANN aktyvinimo funkcija

- ir išvesties funkciją, apskaičiuojančią aktyvacijos išvestį:

$$o_j(t) = f_{out}(a_j(t))$$

Formulė 2 ANN aktyvacijos išvesties funkcija

Dažnai išvesties funkcija yra tiesiog tapatybės funkcija.

- Įvesties neuronas neturi pirmtako, bet yra viso tinklo įvesties sąsaja. Panašiai išvesties neuronas neturi įpėdinio, todėl tarnauja kaip viso tinklo išvesties sąsaja.
- Jungtys ir svoriai:

Tinklas susideda iš jungčių, kiekviena jungtis perkelia neurono i išvestį į j neurono įvestį. Šia prasme i yra j pirmtakas, o j yra i įpėdinis. Kiekvienam ryšiui priskiriamas svoris w_{ij} .

- Dauginimo funkcija:

Sklidimo funkcija apskaičiuoja neurono j įvestį $p_j(t)$ iš pirmtakų neuronų išėjimų $o_i(t)$ ir paprastai turi formą:

$$p_j(t) = \sum_i o_i(t) w_{ij}$$

Formulė 3 ANN sklidimo funkcija

ANN Techniniai Indikatoriai:

1. Santykinio stiprumo indeksas (RSI) – angl. *Relative Strength Index*

RSI rodo prekybos signalą, pagrįstą dabartinės dienos akcijų kainos ir praėjusio laikotarpio akcijų kainos santykiu. RSI vertę galima apskaičiuoti pagal šią formulę:

$$RSI = 100 - [100 / (1 + RSI)] \quad RSI = \frac{\text{Average of 14 days closes up}}{\text{Average of 14 days closes up}}$$

Formulė 4 Santykinio Stiprumo Indekso (RSI) skaičiavimas

2. Stochastinė virpesių technika (SOT) – angl. *Stochastic Oscillatory Technique*

Stochastinio osciliatoriaus technika palygina dabartinę akcijų uždarymo kainą su jos kainų diapazonu praeityje. Prekybos signalui generuoti naudojamos dvi eilutės: % K linija ir % D linija. Šios eilutės apskaičiuojamos pagal šią formulę:

$$K\%line = 100 - \frac{(\text{Recent closes} - \text{Lowest low}(n))}{(\text{Highest High}(n) - \text{Lowest Low})} \quad D\%line = \text{3 period moving average of \%K line}$$

Formulė 5 ANN Prekybos signalo skaičiavimas (SOT)

Stochastinė technika rodo prekybos signalus, kai % K linija ir % D linija yra viena kitoje virš 80 arba žemiau 20. Pardavimo signalas generuojamas, kai %K linija ir %D linija yra viena kitai virš 80, o pirkimo signalas nurodomas, kai %K linija linija ir %D linija viena nuo kitos žemiau 20.

3. Slankioji vidutinė konvergencija ir divergencija (MACD) – angl. *Moving Average Convergence and Divergence (MACD)*

MACD techniką siūlo Geraldas Appelis. Ši technika gali parodyti prekybos signalus (pirkimą ir pardavimą), taip pat akcijų kainos tendencijas (didėjimo tendencija, žemyn tendencija ir tendencija į šoną). MACD technika rodo prekybos signalą, pagrįstą dviejų judančių vidurkių linijų skirtumu. Tai yra 12 dienų ir 26 dienų. MACD vertė apskaičiuojama pagal šią formulę:

$$MACD\ Line = EMA(12) - EMA(26)$$

Formulė 6 MACD formulė

MACD technika rodo prekybą. Kai MACD kerta per nulinę liniją, parodomas pirkimo signalas ir akcijų tendencija pradeda didėti. Nors MACD kerta pagal nulinę liniją, generuojamas pardavimo signalas ir prasideda mažėjimo tendencija.

4. Pokyčio greitis (ROC) – angl. *Rate of Change*

Pokyčio greičio indikatorius (ROC) yra impulsinis osciliatorius. Jis apskaičiuoja procentinį kainos pokytį tarp laikotarpių. ROC ima dabartinę kainą ir palygina su kaina, buvusia prieš „n“ laikotarpius.

$$ROC = \left[\frac{(\text{Current Close} - \text{Close } n \text{ periods ago})}{(\text{Close } n \text{ periods ago})} \right] \times 100$$

Formulė 7 ROC formulė

, kur n = A vartotojo nurodytas numeris

5. Eksponentinis slenkamasis vidurkis (EMA) – angl. *Exponential Moving Average*

Slenkamasis vidurkis yra kaina pagrįstas, atsilikęs (arba reaktyvus) rodiklis, rodantis vidutinę vertybinių popierių kainą per nustatytą laikotarpį. Pagrindinis skirtumas su EMA yra tas, kad seni duomenų taškai niekada nepalieka vidurkio. EMA apskaičiuoti reikia 3 etapais:

- Apskaičiuojamas SMA:

$$SMA = \text{Periodų reikšmės} / \text{Periodų skaičius}$$

Formulė 8 SMA

- Apskaičiuojamas multiplikatorius:

$$\text{Multiplikatorius} = (2 / (\text{Periodų skaičius} + 1))$$

Formulė 9 Slenkamojo vidurkio (SMA) multiplikatorius

- Apskaičiuojamas EMA:

$$EMA = \{ \text{Close} - EMA(\text{previousday}) \} \times \text{multiplier} + EMA(\text{previousday})$$

Formulė 10 Eksponentinis slenkamasis vidurkis

Šie argumentai pateisina ANN naudojimą prognozuojant finansines laiko eilutes (Eakins ir Stansell 2003; Hornik et al. 1989; Hussain et al. 2007; Lam 2004):

1. ANN yra netiesiniai. Tai reiškia, kad jie gali užfiksuoti netiesinius ryšius tarp požymio (įvesties ar nepriklausomo) ir atsako (išvesties ar priklausomo) kintamųjų.
2. ANN yra pagrįsti duomenimis. Tai reiškia, kad jiems nereikia jokių aiškių modelio prielaidų tarp įvesties ir išvesties.
3. ANN gali apibendrinti. Tai reiškia, kad po mokymosi jie gali duoti gerų rezultatų, net kai susiduria su naujais įvesties modeliais.
4. Skirtingai nuo statistikos metodų, ANN nereikia daryti prielaidų apie įvesties duomenų paskirstymą.

Nors ANN turi minėtus pranašumus, jų rezultatų patikimumas buvo suabejotas (Saad et al. 1998). Be to, ANN apribojimai yra tokie (Hussain ir kt. 2007; Lam 2004):

1. Sunku nustatyti optimalų tinklo parametrų derinį, pvz., Mokymosi greitį, impulsą, paslėptų sluoksnių skaičių ir paslėptų mazgų skaičių kiekviename sluoksnyje.
2. Atitinkamų ANN ypatybių pasirinkimas nėra lengvas darbas.
3. Norint išmokyti tinklą pasiekti tikslų rezultatą, reikalingas didelis duomenų kiekis.

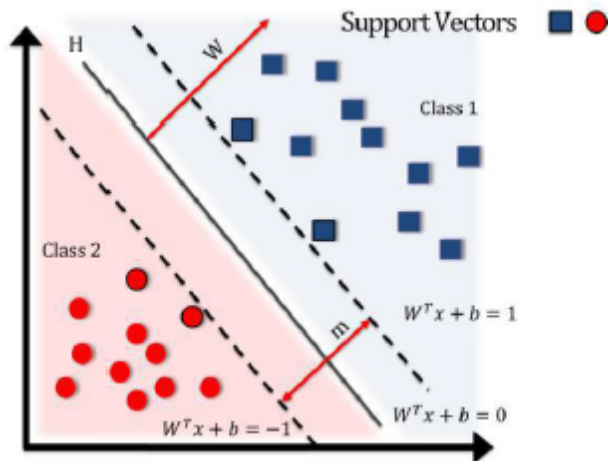
4.3.2. Atraminių Vektorių Mašina (SVM) bei modelio ypatumai

Pastaruoju metu akcijų kainų indeksui ir jo pokyčiams prognozuoti taip pat sėkmingai pritaikytos pagalbinės vektorinės mašinos (SVM). Kim (2003) naudojo SVM, norėdamas numatyti Korėjos sudėtinio akcijų kainų indekso (KOSPI) dienos akcijų kainos pokyčių kryptį. Šiame tyrime buvo atrinkta 12 techninių rodiklių, kurie sudarė pradinis požymius. Rodikliai yra stochastinis K%, stochastinis D%, stochastinis lėtas D%, impulsas, ROC, Williams % R, A / D osciliatorius, skirtumai5 (*angl. disparity*), skirtumai10, OSCP, CCI ir RSI. Be to, šiame tyrime buvo ištirta SVM taikymo finansinėje prognozėje galimybė, lyginant jį su nugaros sklidimo neuroniniu tinklu (BPN) ir atvejais pagrįstais samprotavimais (CBR). Eksperimentiniai rezultatai įrodė, kad SVM pranoksta BPN ir CBR ir yra perspektyvi alternatyva vertybinių popierių rinkos prognozavimui. Manishas ir Thenmozhi (2005) naudojo SVM ir atsitiktinį mišką, norėdami numatyti nacionalinės vertybinių popierių biržos S&P CNX NIFTY rinkos indekso dienos krypties judėjimą ir palygino rezultatus su tradicinių diskriminacinių ir logitinių modelių bei ANN rezultatais. Savo tyrime jie naudojo tuos pačius techninius rodiklius kaip ir Kim (2003) taikomi įvesties kintamieji. Eksperimentiniai rezultatai parodė, kad SVM lenkia atsitiktinius miško, nervų tinklo ir kitus tradicinius modelius. Huang, Nakamori ir Wang (2005) savo tyrime tyrė finansinio judėjimo krypties nuspėjamumą su SVM, prognozuodami savaitės NIKKEI 225 indekso judėjimo kryptį. Norėdami įvertinti SVM prognozavimo galimybes, jie palygino jo našumą su linijinės diskriminantų analizės, kvadratinės diskriminantų analizės ir Elmano neuroninių tinklų išplitimo rezultatais. Eksperimento rezultatai parodė, kad SVM pralenkia kiti klasifikavimo metodai. Manishas ir Thenmozhi (2006) ištyrė ARIMA, ANN, SVM ir atsitiktinių miškų regresijos modelių naudingumą prognozuojant ir prekiaujant S&P CNX NIFTY indekso grąža. Trijų netiesinių modelių ir linijinio modelio našumas matuojamas statistiškai ir finansiškai, atliekant prekybos eksperimentą. Empirinis rezultatas parodė, kad SVM modelis gali pranokti kitus jų tyrime naudojamus modelius. Hsu, Hsiehas, Chihai ir Hsu (2009) sukūrė dviejų pakopų architektūrą integruodami savęs organizavimo žemėlapi ir palaikydami vektorių regresiją vertindami akcijų kainą. Jie ištyrė septynis pagrindinius akcijų rinkos indeksus. Rezultatai parodė, kad dviejų pakopų architektūra yra perspektyvi alternatyva prognozuojant akcijų kainą. Suinteresuotiems skaitytojams gali būti naudingas apžvalginis Atsalakio ir Valavanio (2009) straipsnis, kuriame apibendrinama susijusi literatūra.

„Support Vector Machine“ (SVM) yra mašinų mokymosi (ML) prižiūrima mokymosi technika. SVM palaikymo vektoriai parodyti 11 paveiksle. Jis analizuoja duomenis ir atpažįsta modelius, kurie naudojami regresijai ir klasifikacijai analizuoti. SVM mokymo algoritmas sukuria modelį, kuris priskiria naujus pavyzdžius vienai ar kitai kategorijai, todėl jis nėra tikimybinis dvejetainis tiesinis klasifikatorius. SVM modelis yra pavyzdžių pavaizdavimas kaip taškai erdvėje, suskirstyti taip, kad atskirų kategorijų pavyzdžiai būtų padalyti iš aiškios, kuo platesnės spragos. Tuomet toje pačioje erdvėje atvaizduojami

nauji pavyzdžiai ir prognozuojama, kad jie priklausys kategorijai pagal tai, kurioje tarpo pusėje jie patenka.

SVM sukuria *hyper* plokštumą arba *hyper* plokštumų rinkinį, kurią galima naudoti regresijai ir klasifikacijai, kaip parodyta 11 paveiksle. Gerą atskyrimą atlieka *hyper* plokštuma, kuri taip pat turi didžiausią atstumą, jei skirtumas yra didesnis, klasifikatoriaus paklaidos apibendrinimas yra mažesnis.



Pav. 9 Atraminių Vektorių Mašina (SVM)

Šaltinis: Pateikta Shrimalve H., 2018

Modelio matematinė išraiška:

Tarkime, 2D yra duomenų rinkinys:

D yra duomenų rinkinys, kuriame yra $\{x_i, y_i\}$ rinkinių skaičius, kur:

x_i = Atributų / rinkinių rinkinys

y_i = klasių rinkinys

Tuomet, duomenų rinkinys yra 2D, $y_i = +1 / -1$

• hyper plokštumos lygtis:

$$w \cdot x + b = 0$$

Formulė 11 SVM Hyper plokštumos lygtis

Kur,

w = svorio vektorius (atributo atstumas nuo hiper plokštumos)

x = atributo skaičius

b = skaliarinis / šališkas (*angl. Scalar / Bias*)

Atributų lygtis, kuri yra tolygi hyper plokštumoje:

$$W_0 + W_1 \cdot X_1 + W_2 \cdot X_2 = 0$$

Formulė 12 SVM atributų lygtis, tolygios hyper plokštumai

Atributų lygtis, kuri yra virš hyper plokštumos:

$$H1: W_0 + W_1.X_1 + W_2.X_2 \geq 0.$$

Formulė 13 SVM atributų lygtis, virš hiper plokštumos

Atributų lygtis, kuri yra žemiau hiper plokštumos:

$$H2: W_0 + W_1.X_1 + W_2.X_2 \leq -1$$

Formulė 14 SVM atributų lygtis, žemiau hiper plokštumos

Derindami atributo lygtį virš hiper plokštumos (H1) ir žemiau hiper plokštumos (H2), gaunama ši lygtis:

$$Y_i(W_0 + W_1.X_1 + W_2.X_2) \geq 1$$

Formulė 15 SVM atributų suderinimo funkcija

Didžiausia kraštinė hiper plokštuma (MMH), kuri vaizduoja šia funkcija:

$$\frac{2}{\|W\|}, \text{ where } \|W\| \text{ Eclidean norm of } W$$

Formulė 16 SVM didžiausia kraštinė hiper plokštuma

SVM regresija laikoma nparametrine technika, nes ji remiasi branduolio funkcijomis.

„Statistics and Machine Learning Toolbox™“ įgyvendina tiesinę nejautrią SVM (ϵ -SVM) regresiją, kuri dar vadinama L1 nuostoliu. Esant ϵ -SVM regresijai, mokymo duomenų rinkinyje yra numatomieji kintamieji ir stebimos atsako vertės. Tikslas yra rasti funkciją $f(x)$, kuri kiekvienam treniruotės taškui x nukrypsta nuo y_n ne didesne kaip ϵ verte ir tuo pačiu yra kuo lygesnė.

Kaip jau žinome, prieš paplįtant ML investiciniais sprendimams priimti analitikų dažniausiai naudojamos dvi analizės: fundamentalioji analizė ir techninė analizė. Šiais laikais dirbtinis intelektas padeda investuotojams priimti sprendimus. Plačiai naudojama dirbtinio intelekto technika yra „Support Vector Machines“ (SVM). Ją pritaikyti, taip pat yra naudojami techninės analizės komponentai, tokie kaip RSI ir MACD rodikliai. Šiame tyrime bus tiriamas SVM metodas, pagrįstas technine analize - RSI ir MACD yra SVM įvestys.

Didžiausias skirtumas tarp SVM ir kitų tradicinių mokymosi metodų yra tas, kad SVM nesusitelkia ties hipotezių, darančių nedaug klaidų, kaip kiti metodai, konstravimu. SVM bando sudaryti prognozes, kuriose vartotojas galėtų būti labai tikras, kad rezultatai bus teisingi, nors tam tikram laikotarpiui gali būti daug klaidų. Tradiciškai dauguma mokymosi algoritmų orientavosi į modelio sukeltų klaidų sumažinimą. Jie pagrįsti vadinamuoju empirinio rizikos minimizavimo principu (ERM, Empirical Risk Minimization). SVM dėmesys yra skirtingas. Juo nesiekama sumažinti empirinę riziką padaryti keletą klaidų, bet daroma prielaida, kad kuria patikimą modelį. Šis principas vadinamas

struktūrinės rizikos mažinimu. SVM ieško struktūrinio modelio, kuriame yra mažai rizikos padaryti klaidų su būsimais duomenimis.

Pagrindinė SVM idėja yra sukonstruoti hiper plokštumą kaip sprendimo paviršių, kad būtų kuo labiau padidinta teigiamų ir neigiamų pavyzdžių atskyrimo riba. SVM gali būti naudojamas įvairiai: regresija arba klasifikacija. Šiame tyrime buvo naudojamas palaikomasis vektorinis mašinių klasifikatorius.

Kalbant apie ANN, norint sukurti tinkamą ANN problemą, reikėtų nuspręsti dėl tinklo topologijos, tinklo sluoksnių skaičiaus, mazgų skaičiaus kiekviename sluoksnyje, mazgų aktyvavimo funkcijos ir galiausiai dėl mokymosi algoritmo.

Remiantis topologija, ANN daugiausia skirstomas į dvi perdavimo ir pasikartojančių tinklų grupes. Norint atlikti vienkartinę prognozavimo analizę, pasikartojančios topologijos naudojimas yra dažnesnis nei perdavimo tinklai. Hussainas ir kt. (2007), Lin ir kt. (2006) ir Saad ir kt. (1998) yra tie, kurie naudojami pasikartojančia topologija. Tačiau finansinių prognozių srityje, ypač atliekant daugialypės prognozavimo analizę, „feed-forward“ topologija sulaukė daug daugiau dėmesio. Dėl to pastaroji topologija bus naudojama ir atliekant šiuos tyrimus.

Tinklo sluoksnių skaičius priklauso nuo modeliuojamos problemos sudėtingumo. Be įvesties ir išvesties sluoksnių, kurie yra būtini ANN projektui, daugelis perdavimo tinklų turi vieną ar daugiau paslėptų sluoksnių. Keli tyrėjai pasiūlė skirtingus metodus, kaip nustatyti optimalų paslėptų sluoksnių ir paslėptų mazgų skaičių. Tačiau šie metodai yra labai sudėtingi ir sunkiai pritaikomi (Zhang ir kt., 1998). Be to, nė vienas iš esamų metodų negali garantuoti, kad gauti ANN yra optimalūs. Todėl šiame tyrime laikomasi įprasto tinkamo tinklo projektavimo nustatymo praktikos, kai lyginami skirtingų dizainų įrenginių našumai ir tinklas, kurio rezultatai geriausi (Hosseini et al. 2006). Be to, nors skirtingi literatūros tyrėjai rekomendavo nenaudoti daugiau nei dviejų paslėptų sluoksnių, šiame tyrime optimalus paslėptų mazgų skaičius tik viename paslėptame sluoksnyje nustatomas naudojant eksperimentų projektą (DOE). Kitaip tariant, paslėptų mazgų skaičius laikomas galimu įtakingu DOE veiksmu, turinčiu įtakos siūlomo ANN prognozavimo galimybėms. Šio veiksmio lygiai bus aptariami vėliau šiame skyriuje.

ANN įvesties sluoksnį sudaro įvesties kintamieji (ypatybės), kurie, atrodo, turi įtakos išvesties kintamajam. Šios įtakingos savybės nustatomos ir naudojant DOE. Vėlgi, siūlomos ANN ypatybės yra galimi DOE veiksniai. Šios funkcijos ir jų lygiai bus aptariami vėliau šiame skyriuje. ANN išvesties sluoksnis susideda iš mazgų, susijusių su priklausomais kintamaisiais. Kadangi šio tyrimo tikslas yra tinkamai prognozuoti S&P 500 indekso dienos kryptį ir kryptis yra teigiama arba neigiama, siūlomo ANN išvesties sluoksnį sudaro tik vienas mazgas.

Kaip aktyvavimo funkcija, visų sluoksnių mazgams naudojama liestinė hiperbolinė sigmoidinė (Tansig) funkcija, kuri yra labiausiai paplitusi atitinkamoje literatūroje. Be to, klaidų atgalinio sklaidimo algoritmas naudojamas mokant suprojektuotą ANN.

Tiek ANN, tiek SVM metodai yra galingi ir sudėtingi ML algoritmai, gebantys atlikti laiko eilučių prognozavimą. Vienas pagrindinių jų skirtumų yra kintamųjų pasirinkimas. ANN naudoja fundamentalia analize grįstus rodiklius, todėl įvesčiai reikalauja gana didelio duomenų kiekio. SVM modelis remiasi technine analize, jam atlikti reikia labiau matematinės prieigos, o SVM regresija geba atlikti tikslų prognozavimą nereikalaujant didelio duomenų rinkinio. Kadangi vis dar esama nemažai ginčų, kuri iš dviejų gerai žinomų prieigų yra patikimesnė, o nagrinėjant ML pritaikomumą akcijų indekso kitimo prognozavime šiuos skirtumus labai gerai atspindi ANN ir SVM metodai, toliau šiame tyrime bus atliktas jų praktinis pritaikymas prognozuojant S&P indekso kainų kitimo.

5. Tyrimo metodologija

Pagrindinis šio tyrimo tikslas yra atlikti dienos „Standard & Poor's“ (S&P 500) indekso prognozę naudojant du plačiausiai naudojamus ML metodus – ANN ir SVM bei juos palyginti, nustatyti abiejų modelių pranašumus, trūkumus, įvardyti, kuris iš jų yra tinkamesnis akcijų kainų kitimams prognozuoti.

S&P 500 indekso kainos kaip pagrindinis tyrimo kintamasis buvo pasirinktas naudoti dėl jo paprastumo, nuoseklumo bei nuspėjamumo. Atliekant empirinius tyrimus bei juos testuojant, rinkos indeksas yra labiau tinkamas atrankos objektas nei konkrečių įmonių akcijos, kadangi indekso kaina dažniausiai atspindi visą rinkoje turimą informaciją ir daugeliu atveju koreliuoja su konkrečių įmonių akcijų svyravimais.

Kaip ir minėta darbo teorinėje dalyje, mašininio mokymosi metodų pritaikymas skiriamas į kelis etapus – duomenų surinkimą, paruošimą, mokymosi procesą, testavimą bei tobulinimą. Modeliams veikti, reikalingi tiek praeities duomenys, tiek algoritmo mokymuisi bei testavimui skirta duomenų imtis. Dažniausiai praktikoje laiko eilučių prognozavimas vykdomas naudojantis bent dešimties metų istoriniais duomenimis. Tuo tarpu taikant ML, algoritmų mokymosi procesui skiriama nuo dviejų iki trijų metų imtis, ir nuo vienerių metų iki kelių mėnesių imtis skiriama modelio testavimui. (Rosillo, Fuente, Brugos, 2012). Anot minėtų autorių, imtis, naudojama mokymuisi, nėra apibrėžta ir dažniausiai priklauso nuo duomenų prieinamumo, kuo didesnė mokymosi imtis, tuo algoritmas geba geriau prognozuoti tolimesnes laiko eilutes.

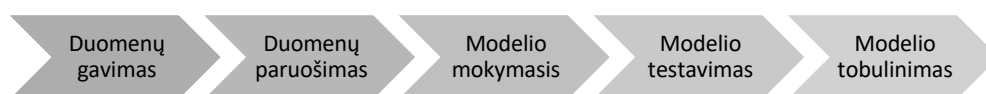
Tyrimui įvykdyti indekso duomenys pasirinkti tokiose intervaluose:

- S&P 500 indekso uždarymo kainos (*angl. closing prices*) nuo 2010.01.01 iki 2020.01.01.
- Algoritmo mokymuisi skirtas periodas nuo 2010.01.01 iki 2019.01.01.
- Algoritmo testavimui skirtas periodas nuo 2019.01.01 iki 2020.01.01.

Šiame darbe metodologija grįsta kiekybinė duomenų analizė su tikslu paruošti duomenis ir prognozuoti indekso kainos kitimą bei palyginti dviejų pasirinktų metodų veikimo principus. Šio mokslinio darbo kiekybinio tyrimo metodas yra antrinių duomenų rinkimas bei analizė, siekiant patikrinti modelių tikslumą bei įgyvendinant darbo uždavinius ir tikslus.

5.1. Tyrimo strategija

Šio tyrimo modelis sukonstruotas iš šešių etapų, remiantis Mašininio mokymosi metodika bei principu, kuris pavaizduotas žemiau esančiame paveiksle:



Pav. 10 Penki pagrindiniai ML proceso žingsniai

Šaltinis: Pagal Mleczeck A, 2019

Šio mokslinio darbo tyrimo modelis apima šešis žingsnius:



Pav. 11 Tyrimo strategija

Šaltinis: parengta autoriaus

1. Duomenų surinkimas

Pirmasis mašininio mokymosi proceso žingsnis yra duomenų surinkimas. Duomenų rinkimo žingsnis yra mašininio mokymosi proceso pagrindas. Klaidos, tokios kaip netinkamų funkcijų pasirinkimas arba sutelkimas į ribotą duomenų rinkinio įrašų tipą, gali padaryti modelį visiškai neveiksmingą. Dėl šios priežasties renkant duomenis būtina atsižvelgti į būtinas aplinkybes tvarkingam duomenų rinkiniui gauti (duomenų naujumas, duomenų patikimumas, matematinių skaičiavimų tikslumas, skirtingų formatų paisymas), nes šiame etape padarytos klaidos tik sustiprės mums pereinant į paskutinius etapus.

2. Duomenų apdorojimas

Realūs praktiniai duomenys dažnai turi neorganizuotų, trūkstančių ar triukšmingų elementų. Todėl, kad mašininis mokymasis būtų sėkmingas, surinktus duomenis reikės išvalyti, paruošti ir manipuluoti duomenimis. Šis procesas yra kritinis žingsnis, nes tik turint švarų duomenų rinkinį, modelis gali būti tikslesnis.

3. ML modelio mokymasis

Duomenų rinkinys integruojamas į paruoštą kompiuterinį algoritmą, o algoritmas pasitelkia sudėtingą matematinį modeliavimą, kad sužinotų ir plėtotų prognozes. Šie algoritmai paprastai skirstomi į vieną iš trijų kategorijų:

- Dvejetainis - skirstymas į dvi kategorijas
- Klasifikacija - klasifikavimas į daugelį kategorijų
- Regresija - rezultatų nuspėjimas

4. ML modelio testavimas

Išmokius modelį, reikia išbandyti, ar jis gerai veiktų realaus pasaulio situacijose. Štai kodėl vertinimui sukurta duomenų rinkinio dalis naudojama modelio kvalifikacijai patikrinti. Tai įtraukia modelį į scenarijų, kuriame jis susiduria su situacijomis, kurios nebuvo jo mokymo dalis.

5. ML modelio tobulinimas

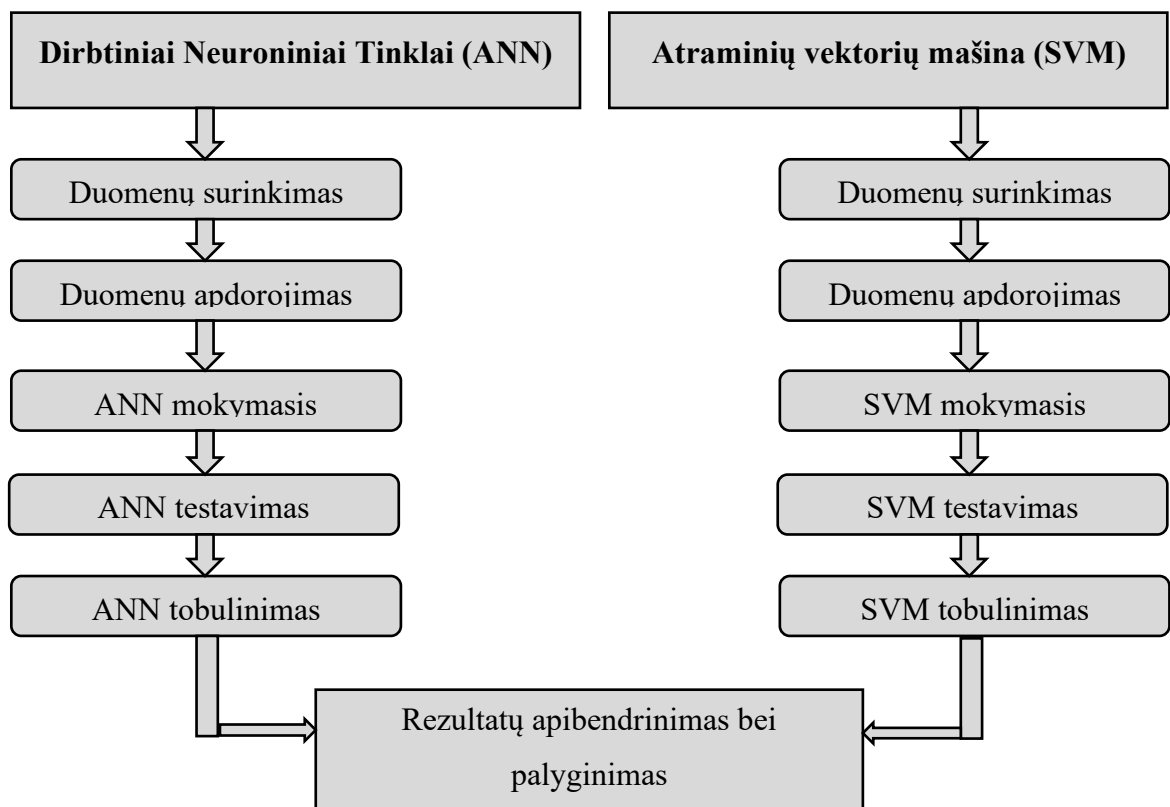
Jei įvertinimas bus sėkmingas, bus pereita prie modelio tobulinimo žingsnio. Šiuo žingsniu bandoma pagerinti teigiamus rezultatus, pasiektus vertinimo etape. Modelio tobulinimas gali būti įvairus. Vienas iš jų yra mokymo etapo peržiūra ir modelio mokymui naudojami keli treniruočių duomenų rinkiniai. Tai gali sukelti didesnę tikslumą, nes ilgesnė treniruotės trukmė suteikia daugiau ekspozicijos ir pagerina modelio kokybę. Kitas būdas tai padaryti yra patobulinti pradinės modeliui suteiktas vertes. Atsitiktinės pradinės vertės dažnai duoda prastus rezultatus, nes bandymų ir klaidų būdu jos palaipsniui tobulinamos.

6. Rezultatų apibendrinimas

Sukūrus ANN ir SVM modelius, atlikus visu penkis prieš tai minėtus žingsnius, gauti rezultatai bus interpretuojami bei apibendrinami.

Visi tyrimo modelio žingsniai bus plačiau ir konkrečiau išdėstyti tyrimo eigos poskyryje.

Žemiau esanti schema detalizuoja šio darbo tyrimo modelį:



Pav. 12 Šio mokslinio darbo detalus tyrimo modelis

Šaltinis: Parengta autoriaus

Tyrimui atlikti bus naudojama programinė įranga MATLAB (R2020b versija) su integruotais “Machine Learning and Deep Learning” papildiniais.

5.2. Tyrimo eiga

5.2.1. Dirbtiniai neuroniniai tinklai (ANN)

Duomenų rinkimas

Pirmame ir viename svarbiausių žingsnyje bus surinkti duomenys, kurie bus naudojami ANN modeliui sukonstruoti. Kaip ir jau minėta, tyrimo imtis apima S&P 500 kainas bei 20 kintamųjų vertes dešimties metų periode nuo 2010.01.01 iki 2020.01.01. 80% duomenys - Aštuonerių metų laikotarpis bus naudojamas mokymosi tikslams, kiti 10% modelio verifikacijai bei likę 10% modelio testavimui.

S&P 500 indekso uždarymo kainos dešimties metų periode išeksportuotos iš www.finance.yahoo.com.

Kadangi ANN, priešingai nei SVM, remiasi fundamentalia analize, todėl daroma prielaida, jog akcijų kainą įtakoja tokie makroekonominiai rodikliai kaip akcijų grąža, finansų bei ekonominiai indikatoriai, valiutos kursai, didžiausių rinkos indeksų grąža bei prekybos dydžiai. Remiantis Niaki, Hoseinzade, 2013 metais atlikta faktorine duomenų analize, autoriai nurodė, kokie svarbiausi indikatoriai turi būti įtraukti į modelį, prognozuojant S&P indekso kainos judėjimą. Šiame tyrime kintamieji sukonstruoti remiantis minėtų autorių atliktu tyrimu.

Žemiau esančioje lentelėje pateiktas sąrašas tyrimui naudojamų kintamųjų bei nurodyti šaltiniai, kuriais remiantis duomenys buvo surinkti:

4 lentelė. ANN modeliui naudojamų kintamųjų sąrašas

Nr.	Kintamasis	Kintamojo pavadinimas	Šaltinis
1	SPYt-1	S&P 500 indekso grąža t-1	https://finance.yahoo.com
2	SPYt-2	S&P 500 indekso grąža t-2	https://finance.yahoo.com
3	SPYt-3	S&P 500 indekso grąža t-3	https://finance.yahoo.com
4	Oil	Santykinis naftos kainos pasikeitimas (USD/barelis)	https://finance.yahoo.com
5	Gold	Santykinis aukso kainos pasikeitimas	https://finance.yahoo.com
6	CTB3M	JAV išdo vertybinių popierių rinkos grąžos pokytis per pastovų 3 mėnesių terminą	https://www.treasury.gov/resource-center/
7	AAA	“Moody” korporacijų obligacijų pajamingumo pokytis	https://fred.stlouisfed.org/series/AAA

8	AAPL	“Apple” akcijų grąža t-1	https://finance.yahoo.com
9	AMZN	„Amazon“ akcijų grąža t-1	https://finance.yahoo.com
10	MSFT	“Microsof” akcijų grąža t-1	https://finance.yahoo.com
11	FB	“Facebook Inc” akcijų grąža t-1	https://finance.yahoo.com
12	GOOGL	“Alphabet Inc.” akcijų grąža t-1	https://finance.yahoo.com
13	USD-JPY	Santykinis pasikeitimas USD/JPY valiutos kurso	https://finance.yahoo.com
14	USD-GBP	Santykinis pasikeitimas USD/GBP valiutos kurso	https://finance.yahoo.com
15	USD-CAD	Santykinis pasikeitimas USD/CAD valiutos kurso	https://finance.yahoo.com
16	HSI	Hang Sen indekso grąžą t-1	https://finance.yahoo.com
17	FCHI	CAC 40 indekso grąžą t-1	https://finance.yahoo.com
18	FTSE	FTSE 100 indekso grąžą t-1	https://finance.yahoo.com
19	DJI	Dow Jones indekso grąžą t-1	https://finance.yahoo.com
20	V	Santykinis pokytis S&P 500 indekso prekybos t-1	https://finance.yahoo.com

Šaltinis: parengta autoriaus

Vienas svarbiausių duomenų rinkimo procesų, užtikrinant sklandžią duomenų klasifikaciją bei ANN modelio veikimą yra sugrupavimas. Remiantis šia logika, mūsų tyrimo kintamieji bus išskirstyti į šešias skirtingas grupes.

Žemiau esančioje lentelėje pavaizduotos kintamųjų duomenų grupės bei kintamųjų pasiskirstymas:

5 lentelė. ANN kintamųjų grupavimas

Grupės numeris	Grupės aprašymas	Grupės pavadinimas	Įtraukti kintamieji
1	S&P 500 indekso trijų prieš tai esančių dienų grąžą	G1	SPYt-1
			SPYt-2
			SPYt-3
2	Finansiniai ir ekonominiai indikatoriai	G2	Oil
			Gold
			CTB3M
			AAA
3	Penkių didžiausių S&P 500 indekso įmonių grąža	G3	AAPL
			AMZN

			MSFT
			FB
			GOOGL
5	Valiutos kurso pokytis	G4	USD-JPY
			USD-GBP
			USD-CAD
6	Keturių didžiausių akcijų indeksų gražą	G5	HSI
			FCHI
			FTSE
			DJI
6	S&P 500 indekso prekybos kiekis	G6	V

Šaltinis: parengta autoriaus

Duomenų apdorojimas:

Šiame etape labai svarbu tinkamai apdoroti ir paruošti duomenis. Praktikoje dažniausiai pastebime, kad duomenys nesutampa, vieni rodikliai įtraukia savaitgalių rodmenis, kiti ne. Taip pat, kai kurių laikotarpių duomenys nėra prieinami.

Šio tyrimo duomenų apdorojimo etapas apima žemiau išvardintus žingsnius:

- Sudėlioti, kad visų duomenų lentelės būtų vienodo dydžio (vienodas kiekis dienų ir atitinkamų reikšmių)
- Peržiūrėti, ar nėra eilučių, kuriose duomenų nebuvo rasta. Jei tokios yra, sutvarkyti ir užpildyti reikšmes, kad modelis neturėtų spragų bei tolimesnis rezultatas nebūtų iškraipytas.
- Importuoti duomenis į MATLAB bei paruošti ANN modeliui vykdyti.

Modelio mokymasis:

Paruošus duomenis, esame pasirengę ANN Modelio vykdymui. Šiame tyrime modelis nuo pradžių nėra kuriamas programuojant, naudojamas MATLAB aplikacijoje esamas ANN papildinys. Komanda, kuri reikalinga modeliui vykdyti, vadinasi „*nnstart*“. „*Nnstart*“ atidaro langą su paleidimo mygtukais neuroninio tinklo pritaikymui, modelio atpažinimui, grupavimui ir laiko eilučių įrankiams. Čia taip pat pateikiamos nuorodos į duomenų rinkinių sąrašus, pavyzdžius ir kitą naudingą informaciją pradedant.

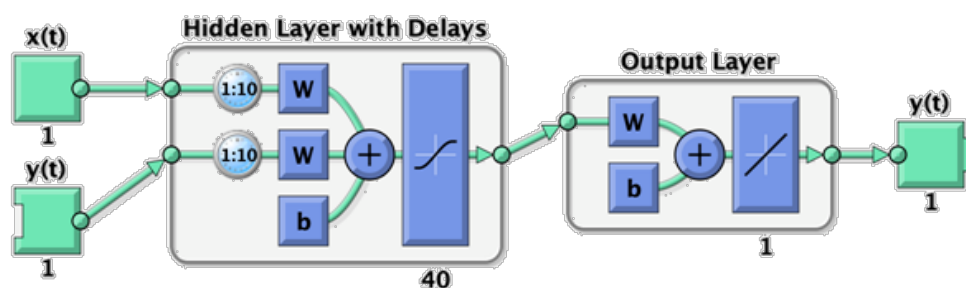
ANN modeliui pradėti vykdyti įvedami duomenys:

NARX modelis duomenų loginei klasifikacijai vykdyti - Laiko eilučių modeliavime netiesinis autoregresinis egzogeninis modelis (NARX) yra netiesinis autoregresyvus modelis, turintis egzogeninių įėjimų. Tai reiškia, kad modelis susieja dabartinę laiko eilutės vertę su abiem:

- tos pačios serijos praeities vertės; ir
- dabartinės ir praeities varomosios (egzogeninės) serijos vertės, tai yra išoriškai nustatytos serijos, turinčios įtakos dominančiai serijai.

Išvestis (*angl. Target Time series, defining desired output $y(t)$*) – nurodome, kad S&P 500 indekso uždarymo kaina yra mūsų tikėtinas rezultatas.

Prieš paleidžiant algoritmą mokymuisi, reikia apibrėžti neuroninio tinklo struktūrą, nurodant paslėptų neuronų kiekį. Neuroniniuose tinkluose paslėptas sluoksnis yra tarp algoritmo įvesties ir išvesties, kuriame funkcija taiko įvadams svorius ir nukreipia juos per aktyvinimo funkciją kaip išėjimą. Trumpai tariant, paslėpti sluoksniai atlieka netiesines įvesčių į tinklą transformacijas. Dažniausiai praktikoje vienas iš pirminių bandymų nurodoma paslėptų neuronų kiekį įvardinti kaip dvigubai didesnę, nei kintamųjų kiekis, kadangi turime 20 kintamuosius, pirminei struktūroje naudosime 40 paslėptų neuronų. Taip pat, šiame etape nurodomas uždelsimo skaičius, taip modelis įvertina galimą dienų paklaidą. MATLAB kaip pirminiam modeliui siūlo 4:1 proporciją tarp paslėptų neuronų ir uždelsimų skaičių. Apibrėžus kiekis, neuroninių tinklų modelis atrodo taip:



Pav. 13 ANN suformuoto modelio architektūra

Šaltinis: parengta autoriaus (MATLAB)

Tinklas bus sukurtas ir apmokytas atviros kilpos forma, kaip parodyta žemiau. Atviros ciklo (vieno žingsnio) treniruotės yra efektyvesnės nei uždaro ciklo (daugiapakopės) treniruotės. Atvira kilpa leidžia mums tiekti tinklą teisingomis prieš tai apmokytomis išvesties reikšmėmis, kai mes mokome jį generuoti teisingas išvesties reikšmes.

Prognozavimo funkcija apibrėžiama formule:

$$y(t) = f(x(t-1), \dots, x(t-d), y(t-1), \dots, y(t-d))$$

Formulė 17 ANN Prognozavimo funkcija

Sekančiame žingsnyje taip pat nurodomas procentinis duomenų kiekis, kiek duomenų bus skirta mokymuisi, kiek vertinimui (verifikacijai) ir kiek testavimui. Kaip ir jau minėta anksčiau, proporcijos nustatėmė kaip 80%, 10% ir 10% atitinkamai.

Mokymasis (*angl. Training*) – jie pateikiami tinklui mokymo metu ir tinklas koreguojamas pagal jo klaidą.

Verifikacija (*angl. Validation*) - jie naudojami tinklo apibendrinimui įvertinti ir sustabdyti mokymą, kai apibendrinimas nustoja gerėti.

Testavimas (*angl. Testing*) - šie duomenys neturi jokios įtakos treniruotėms, todėl teikia nepriklausomą tinklo našumo vertinimą mokymo metu ir po jo.

Mokymosi algoritmas pasirinktas Bajeso dėsningumo algoritmas (*angl. Bayesian Regularization*). Šis algoritmas užima daugiau laiko, tačiau suteikia geresnę generalizaciją kai turime didelius duomenų rinkinius, įtraukiančius daug skirtingų kintamųjų, nes įtraukia duomenų normalizacijos žingsnį. Kaip alternatyva, mokymuisi galima rinktis Levenberg-Marquardt algoritmą, jis automatiškai sustabdo mokymosi procesą, kai modelis nebėra tobulinamas, atsižvelgiant į statistinį rodiklį – šaknies vidurkio kvadrato nuokrypį (*angl. Root-mean-square deviation, santr. RMSE*). Nors ir abu algoritmai laikomi tinkamais naudoti laiko eilučių prognozavimui, tačiau mes pasirenkame naudoti Bajeso algoritmą, kadangi turime labai daug skirtingų duomenų su skirtingomis reikšmėmis, todėl šis algoritmas papildomai sugebės išspręsti normalizacijos problemą.

Modelio testavimas:

Parametrai, kuriais bus remiamasi ANN mokymuisi interpretuoti ir įvertinti:

6 lentelė. ANN modelio mokymuisi įvertinti naudojami parametrai

<u>Parametras</u>	<u>Parametro reikšmė</u>	<u>Optimali reikšmė</u>
Laikas (<i>angl. Time</i>)	Laikas, kiek modelis užtruko mokymuisi.	Neapibrėžta.
Atlikimas (<i>angl. Performance</i>)	Modelio prognozių palyginimas su (žinomomis) priklausomo kintamojo reikšmėmis duomenų rinkinyje.	Nulis. (Kuo mažesnis teigiamas skaičius)
Gradientas (<i>angl. Gradient</i>)	Pokyčio greitis, atsižvelgiant į kintamo kiekio atstumą.	Vienas. (Kuo arčiau vieneto)
MSE (<i>angl. Mean Square Error</i>)	Vidurinis kvadratų skirtumas tarp išvesties ir rezultatų.	Nulis. (Žemesnė reikšmė geresnė)

R (Vidurkis)	Išvesties ir rezultatų koreliacija.	Vienas. (artima koreliacija)
--------------	-------------------------------------	------------------------------

Šaltinis: parengta autoriaus

Taip pat mus domina laiko eilučių išvesties elemento atsakas bei koreliacija taip išvesties ir planinio rodiklio.

Modelio tobulinimas:

Jei įvertinimas bus sėkmingas, bus pereita prie modelio tobulinimo žingsnio. Šiuo žingsniu bandoma pagerinti teigiamus rezultatus, pasiektus vertinimo etape.

Modelio tobulinimui MATLAB siūlo tris variantus:

1. Pakartotinis mokymasis
2. Tinklo dydžio didinimas (paslėptų neuronų)
3. Papildomų kintamųjų įtraukimas

Reikalingas žingsnis bus numatytas tyrimo eigoje priklausomai nuo modelio pirminio mokymosi rezultato.

5.2.2. Atraminių vektorių mašinos (SVM)

Atraminių vektorių klasifikatorius - mašininio mokymosi algoritmas skirtas klasifikuoti duomenims. Tai prižiūrimo mokymosi (*angl. supervised learning*) metodas, kuomet siekiama suklasifikuoti jau pažymėtus (t. y. skirtingų, iš anksto žinomų klasių) duomenis. „Support Vector Machine“ taip pat gali būti naudojamas kaip regresijos metodas, išlaikant visas pagrindines algoritmą apibūdinančias savybes. Nors SVR yra mažiau populiarus nei SVM, įrodyta, kad tai yra veiksminga priemonė vertinant realios vertės funkciją. MATLAB „Statistics and Machine Learning Toolbox™“ įgyvendina tiesinę epsilonui nejautrią SVM (ϵ -SVM) regresiją, kuri dar vadinama L1 nuostoliu. Esant ϵ -SVM regresijai, mokymo duomenų rinkinyje yra numatomieji kintamieji ir stebimos atsako vertės. Tikslas yra rasti funkciją $f(x)$, kuri kiekvienam treniruotės taškui x nukrypsta nuo y_n ne didesne kaip ϵ verte ir tuo pačiu yra kuo lygesnė.

Mūsų tyrimo S&P indekso uždarymo akcijų kainas prognozuoja SVR. Šaknies vidurkis Kvadratinė paklaida (RMSE) ir vidutinė absoliuti procentinė paklaida (MAPE) yra naudojamos modelių tinkamumui įvertinti. Šias priemones naudojo ir Nayak ir kt. (2015), Patel ir kt. (2015b), Araujo ir kt. (2015), Manahovas ir kt. (2014). Remiantis šių autorių tyrimais, MAPE ir RMSE bus apskaičiuoti modelio remiantis šiomis formulėmis (kintamųjų reikšmės pateiktos žemiau po formule):

$$MAPE = \frac{1}{T} \sum_{i=1}^T \left| \frac{d_i - \hat{d}_i}{d_i} \right|$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (d_i - \hat{d}_i)^2}$$

Formulė 18 SVM: MAPE ir RSME apskaičiavimas

Priešingai nei ANN modeliui, čia kintamieji bus matematiniai rodikliai, apskaičiuoti pagal toliau pateiktas formules.

Šiame tyrime naudotoms kintamųjų reikšmėms buvo apibrėžti šie žymėjimai:

- N : bendras imčių skaičius;
- T : bendras testavimui skirtų imčių skaičius;
- d : tikroji imties vertė;
- d^{\wedge} : vertė, nustatyta modelio;
- P : laikotarpis (minutėmis ar dienomis);
- Cl : laikotarpio pabaigos kaina;
- Hi : didžiausia kaina laikotarpyje;
- Lo : mažiausia kaina laikotarpyje;
- Up : kainų kilimo skaičius;
- Dw : kainų mažėjimo skaičius;
- r : grąža.

Kaip ir prieš tai minėta, literatūroje dažniausiai naudojami SVR ir SVM modelių prognozavimo rodikliai yra techninės analizės rodikliai. Anot Nayak ir kt. (2015, p. 672), TA rodikli sudaro duomenys, gauti taikant tam tikrą formulę pagal buvusias akcijų kainas. Šiame tyrime šių rodiklių vertės, išsamiai aprašytos žemiau esančioje lentelėje, laikomos kintamaisiais SVR modelyje.

7 lentelė. SVM modeliui naudoti kintamieji

<u>Kintamasis</u>	<u>Formulė apskaičiuoti</u>	<u>Kintamojo aprašymas</u>
SMA (Paprastas slenkamasis vidurkis / Simple Moving Average)	$SMA = \frac{1}{T} \sum_{i=1}^T Cl_i$	Paprasčiausias pasirinktas rodiklis yra slenkamasis vidurkis, kurį lengva suprasti ir apskaičiuoti. Paprastasis slenkamasis vidurkis (SMA) yra T praeities kainų Cl_i kliento aritmetinis vidurkis.

WMA (Svertinis slenkamasis vidurkis / Weighted Moving Average)	$WMA = \frac{PCl_i + (P-1)Cl_{i-1} + \dots + Cl_{i-P}}{P + (P-1) + \dots + 1}$	Šiame tyrime naudojamas vienas slankiojo vidurkio pokytis, žinomas kaip svertinis slenkamasis vidurkis (WMA), kuris pateikia tikslesnius svetus naujausioms kainoms.
RSI (Santykinio stiprumo indeksas / Relative Strength Index)	$RSI = 100 - \frac{100}{1 + \frac{\sum_{i=0}^{p-1} Up_{t-i}/n}{\sum_{i=0}^{p-1} Dw_{t-i}/n}}$	Santykinis stiprumo indeksas (RSI) yra nuostolių ir pastarojo laikotarpio prieaugio palyginimo rodiklis, kuris nustato ar rinkoje yra perteklius ar stygius. Laiku t jis turi ekvivalento formą. Šiame tyrime apskaičiuojamas RSI pagal akcijų uždarymo kainas pasirinktu laikotarpiu.
ADO (Kaupimo pasiskirstymo osciliatorius / Accumulation Distribution Oscillator)	$ADO = \frac{Hi_t - Cl_{t-1}}{Hi_t - Lo_t}$	Kaupimo / paskirstymo osciliatorius (ADO) yra impulso rodiklis, nurodantis kainų tendencijos stiprumą
ATR (Vidutinis tikrasis diapazonas / Average True Range)	$TR = \max[Hi_t - Lo_t, Hi_t - Cl_{t-1} , Lo_t - Cl_{t-1}]$ $ATR_t = \frac{ATR_{t-1}(P-1) + TR_t}{P}$	Šiuo rodikliu siekiama išmatuoti bulių ir lokių kainų tendencijos, naudojant tikrąjį diapazoną (TR) (Wilder, 2020).

Šaltinis: sudaryta autoriaus

Kad sėkmingai pavyktų sumodeliuoti SVM MATLAB programoje, įvardinome šešis pagrindinius žingsnius, kuriuos reikės įvykdyti, tam kad galėtumėme interpretuoti modelio tinkamumą prognozavimui:

- Duomenų rinkinio paruošimas bei jo padalijimas į mokymosi bei testavimo etapus
- Validavimui skirtų duomenų paruošimas (k-fold CV)
- Kintamųjų pasirinkimas bei integravimas
- Mokymasis: geriausių parametru paieškos funkcijos pagalba (*angl. hyperplane*)
- Modelio testavimas
- *Hyperplane* vizualizavimas

SVM modeliui suformuoti bei mokyti bus naudojamas vienas iš MATLAB papildinių „Regression Learner“. SVM modelis gali būti paremtas šešiais skirtingais regresijos tipais (7 lentelė),

kiekvienas iš jų turi skirtingų ypatybių, todėl šiame tyrime mes pritaikysime SVM visus penkis regresijos tipus ir pateiksim išvadas, kuris modelis pateikė tiksliausius rezultatus.

8 lentelė. SVM modeliams taikyti regresijos tipai bei jų ypatybės

<u>Regresijos Modelio Tipas</u>	<u>Apibrėžimas</u>
Linijinis SVM (<i>Linear SVM</i>)	Naujausias itin greitas mašinų mokymosi algoritmas, leidžiantis išspręsti daugiasluoksnes klasifikavimo problemas iš itin didelių duomenų rinkinių, įgyvendinantis originalią patentuotą pjovimo plokštumos algoritmo versiją, skirtą projektuoti tiesinę atramos vektoriaus mašiną.
Kvadratinis SVM (<i>Quadratic SVM</i>)	Tai nauja nelineinio branduolio nelineinio palaikymo vektoriaus mašina (vadinama QSVM). SVM optimizavimo problemą galima išreikšti taip: geometrinės ribos padidinimas atsižvelgiant į visus treniruotės duomenis, o funkcinė riba yra didesnė nei konstanta.
Kubinis SVM (<i>Cubic SVM</i>)	Rečiau naudojamas, tačiau gana tikslus regresijos tipas, jo veikimo principas analogiškas kvadratiniam SVM, tačiau reikšmių transformavimas vykdomas kubiniu parametru.
Puikus Gauso SVM (<i>Fine Gaussian SVM</i>)	Gauso mišinio modeliai yra tikimybinis modelis, rodantis normaliai pasiskirsčiusias populiacijos grupes. Puikus Gauso SVM leidžia greitai keisti atsako funkciją. Branduolio skalė nustatyta į $\sqrt{P} / 4$, kur P yra numatančiųjų skaičius.
Vidutinis Gauso SVM (<i>Medium Gaussian SVM</i>)	Suteikia mažiau lanksčią reagavimo funkciją. Branduolio skalė nustatyta į \sqrt{P} .
Šiurkštus Gauso SVM (<i>Coarse Gaussian SVM</i>)	Suteikia griežtą atsako funkciją. Branduolio skalė nustatyta į $\sqrt{P} * 4$.

Šaltinis: parengta autoriaus pagal MATLAB „Statistics and Machine Learning Toolbox“

Kiekviename iš šešių SVM programa prašo papildomai apibrėžti keturis parametrus:

- Dėžutės apribojimo režimas (*angl. Box constraint mode*) - Dėžutės apribojimas kontroliuoja klaidą, gaunamą dėl kintamųjų pernelyg didelio pasiskirstymo. Didesnis dėžės apribojimas suteikia lankstesnį modelį. Mažesnė vertė suteikia tvirtesnį modelį, mažiau jautrų pertekliniam pritaikymui. Kai langelio apribojimo režimas nustatytas į

automatinį, programa naudoja heuristinę procedūrą, kad pasirinktų langelio apribojimą. Mūsų tyrime naudosisime automatinį pasirinkimą, darome prielaidą, jog parinkta reikšmė bus optimaliausia.

- „Epsilon“ režimas (*angl. Epsilon mode*) - Spėjamosios klaidos, kurios yra mažesnės už epsilon (ϵ) vertę, nepaisomos ir traktuojamos kaip lygios nuliui. Mažesnė „epsilon“ vertė suteikia lankstesnį modelį. Kai „Epsilon“ režimas nustatytas į „Auto“, programa naudoja euristinę procedūrą, kad pasirinktų branduolio skalę. Šioje vietoje mes taip pat naudosisime automatinį pasirinkimą.
- Branduolio mastelio režimas (*angl. Kernel scale mode*) - Branduolio skalė valdo numatančiųjų skalę, kurioje branduolys smarkiai skiriasi. Mažesnė branduolio skalė suteikia lankstesnį modelį. Kai branduolio mastelio režimas nustatytas kaip automatinis, programa branduolio mastui pasirinkti naudoja euristinę procedūrą.
- Standartizavimas (*angl. Standardize*) - Prognozuotojų standartizavimas juos transformuoja taip, kad jų vidurkis būtų 0 ir standartinis nuokrypis 1. Standartizavimas pašalina priklausomybę nuo savavališkų skalių rodiklių ir apskritai pagerina našumą. Šioje vietoje yra pasirinkimas, ar naudoti standartizavimą, ar ne. Mūsų SVM modeliams standartizavimas bus pritaikytas, nes kintamųjų reikšmės nėra tolygios, o modelis reikalauja kuo tikslesnio rezultato.

Rezultatų tikslumui įvertinti bus naudojami penki parametrai (MATLAB sistemos pateikiami duomenys):

9 lentelė. SVM modelio įvertinimui naudojami parametrai

<u>Parametras</u>	<u>Parametro reikšmė</u>	<u>Optimali reikšmė</u>
Laikas (<i>Time</i>)	Laikas, kiek modelis užtruko mokymuisi.	Neapibrėžta.
RMSE (<i>Root-mean-square deviation</i>)	RMSD parodo skirtumų tarp numatomų verčių ir stebimų verčių antrojo mėginio momento kvadratinę šaknį arba šių skirtumų kvadratinį vidurkį.	Nulis. Kuo žemesnis RMSE, tuo modelis tikslesnis.
R-squared	„R kvadratu“, yra priklausomo kintamojo dispersijos dalis, kurią galima nuspėti iš nepriklausomo kintamojo.	Vienas. (Kuo arčiau vieneto, tuo geriau).
MSE (<i>Mean Square Error</i>)	Vidurinis kvadratų skirtumas tarp išvesties ir rezultatų.	Nulis. (Žemesnė reikšmė geresnė)

MAE (<i>Measure of errors</i>)	Statistikoje vidutinė absoliuti paklaida yra klaidų matas tarp suporuotų stebėjimų, išreiškiančių tą patį reiškinį.	Nulis, tačiau taip griežtai neapibrėžiama. Mažesnis skaičius reiškia tikslesnį modelį.
----------------------------------	---	--

Šaltinis: parengta autoriaus

Modelio įvertinimui bus remiamasi grafikais, tokiais kaip standartinis atsakas (*angl. Response Plot*), kuris atspindi regresijos rezultatus. Išmokus regresijos modelį, atsako diagrama rodo numatomą atsaką, palyginti su duomenų skaičiumi.

Taip pat, bus stebimas Numatyty/Faktinių (*angl. Predicted vs. Actual Plot*) reikšmių grafikas, skirtas patikrinti modelio veikimą. Šis grafikas padės suprasti, kaip gerai regresijos modelis numato skirtingas atsako reikšmes. Tobulas regresijos modelis turi numatomą atsaką, lygų tikram atsakui, todėl visi taškai yra ant įstrižainės. Vertikalus atstumas nuo tiesės iki bet kurio taško yra to taško prognozavimo klaida. Geras modelis turi nedidelių klaidų, todėl prognozės yra išsklaidytos šalia linijos.

Be to, stebėsime Rezultato klaidos diagramą (*angl. Residuals plot*). Ši diagrama rodo skirtumą tarp suprognuotų ir tikrųjų atsakymų. Paprastai gero modelio klaidos yra išsibarsčiusios maždaug simetriškai apie 0. Jei išsibarstymas netenkina šios sąlygos, modelis bus tobulinamas.

Atlikus SVM modelio mokymosį žingsnį šešiais skirtingais regresijos tipais iš dalies įgyvendiname modelio tobulinimo etapą, kadangi vienas iš pasirinkimų, kaip SVM gali būti tobulinamas, yra regresijos tipo pasirinkimas. Tyrime taip pat bandysime modelį tobulinti naudojant PCA kintamųjų transformavimui, kad sumažinti dimensijas.

6. Empirinis tyrimas ir jo rezultatai

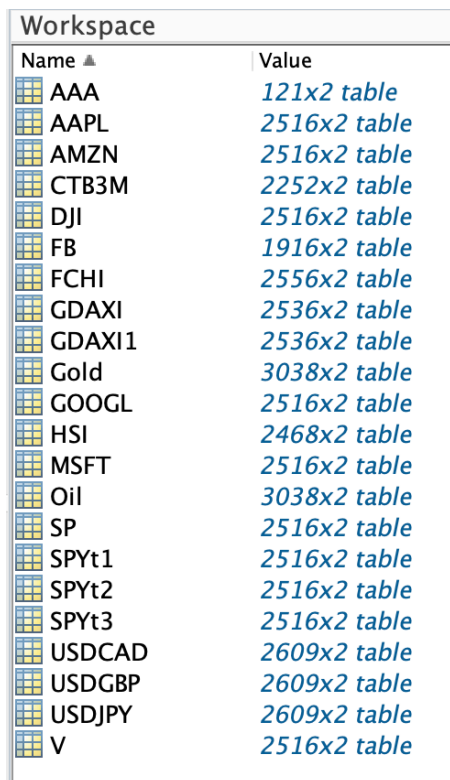
6.1. ANN modeliu grįstas S&P 500 indekso prognozavimas

6.1.1. Duomenų surinkimas ir apdorojimas

Pirmasis žingsnis atliekant indekso kainų kitimo prognozavimą ANN metodu yra duomenų surinkimas (Lentelė 4). Kadangi šis tyrimas talpina 20 kintamųjų, tikėtini reikšmių kiekio neatitikimai, nulinės reikšmės, taip pat, kadangi naudoti keli šaltiniai, svarbu atkreipti dėmesį į duomenų formatavimą bei, esant neatitikimams, juos suvienodinti. Apdorojus duomenis, duomenys sukeliama kaip algoritmo kintamieji ir ANN pradedama mokymosi procesą.

Išeksportavus reikalingus kintamųjų duomenis, pastebėjome lentelių dydžių neatitikimus:

10 lentelė. Pirminių duomenų kintamųjų lentelė (ANN)



Name	Value
AAA	121x2 table
AAPL	2516x2 table
AMZN	2516x2 table
CTB3M	2252x2 table
DJI	2516x2 table
FB	1916x2 table
FCHI	2556x2 table
GDAXI	2536x2 table
GDAXI1	2536x2 table
Gold	3038x2 table
GOOGL	2516x2 table
HSI	2468x2 table
MSFT	2516x2 table
Oil	3038x2 table
SP	2516x2 table
SPYt1	2516x2 table
SPYt2	2516x2 table
SPYt3	2516x2 table
USDCAD	2609x2 table
USDGBP	2609x2 table
USDJPY	2609x2 table
V	2516x2 table

Šaltinis: parengta autorės (MATLAB)

Skirtumai tarp duomenų išryškėjo dėl to, jog kai kurie rodikliai įtraukia tiek savaitgalių, tiek šventinių dienų reikšmes. Darome prielaidą, jog teisinga imties suma lygi 2516 reikšmės (pagal indekso parodymus, įtraukiant tik biržos darbo dienas dešimties metų intervale), todėl nereikalingi duomenys, tokie kaip savaitgaliai, buvo ištrinti.

Stiprus duomenų trūkumas pastebėtas AAA indekse. Išanalizavus duomenų rinkinį, išsiaiškinta, jog parodymai yra skelbiami pirmąją mėnesio dieną. Kad užpildyti kitų dienų duomenų eilutes ir suvienodinti duomenis, padaryta prielaida, jog indekso reikšmė likusiomis dienomis buvo vienoda kaip ir pirmąją mėnesio dieną.

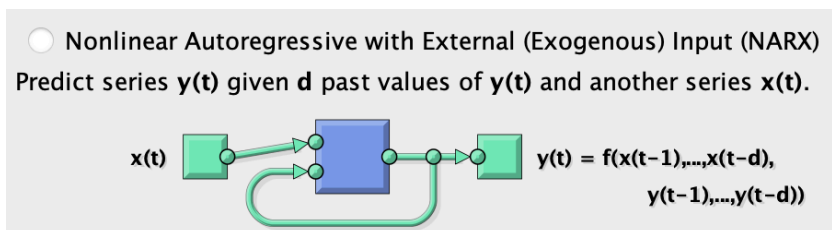
Kai kuriose vietose pastebėta, jog trūksta tam tikrų duomenų reikšmių. Tokiu atveju, reikšmė buvo apskaičiuota kaip $t-1$ ir $t+1$ dienų vidurkis.

FB indekse („Facebook“) duomenų kiekį palikome toks, kokį ir ištraukėme, nors reikšmių kiekis gautas mažesnis. To priežastis – duomenys iki 2012.05.18 nėra prieinami, nes įmonė pradėjo prekiauti savo akcijomis biržoje tik minėtą dieną.

Sutvarkius duomenis, visose lentelėse (išskyrus FB), gavome tokius pačius eilučių bei stulpelių skaičius visiems kintamiesiems. Taip užtikriname, kad duomenys yra paruošti modeliui vykdyti.

6.1.2. ANN Modelio mokymasis ir tobulinimas

Iškvietus *nneural* programą, pasirenkamas nelinejinis autoregresijos modelis (exogeninis). Kadangi mūsų kintamųjų duomenys yra skirtingi ir tarpusavyje nesusiję, nelinejinis modelis duomenų rinkiniui yra labiausiai tinkamas (NARX).



Pav. 14 NARX modelis ANN duomenimis klasifikuoti

Šaltinis: parengta autoriaus (MATLAB)

Toliau, remiantis tyrimo eigoje minėtais žingsniais, įvedėme reikalingus įvesties ir išvesties duomenis.

ANN modelis: NARX, Hidden Neurons (40), Number of delays(4), Bayesian Regularization, data structure 80%-10%-10%.

Gauti rezultatai:

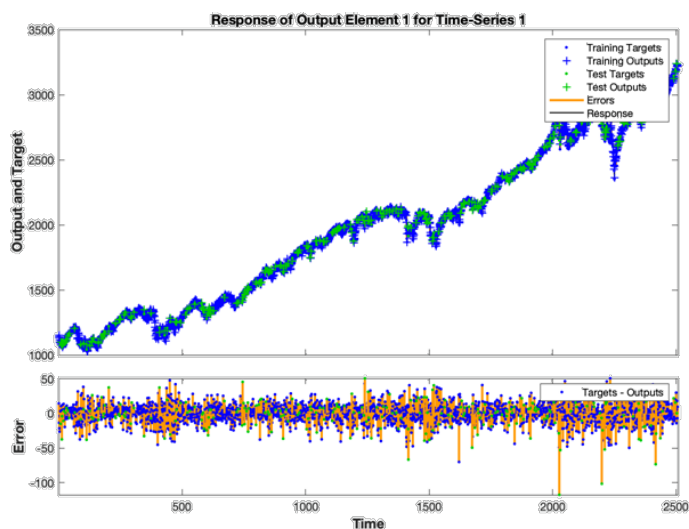
11 lentelė. ANN mokymosi rezultatai

Parametras	Gauta reikšmė	Optimali reikšmė
Laikas (angl. Time)	0:02:59	Neapibrėžta.
Atlikimas (angl. Performance)	3,79	Nulis. (Kuo mažesnis teigiamas skaičius)
Gradientas (angl. Gradient)	1,71	Vienas. (Kuo arčiau vieneto)
MSE (angl. Mean Square Error)	1,87	Nulis. (Žemesnė reikšmė geresnė)
R (Vidurkis)	0,91	Vienas. (artima koreliacija)

Šaltinis: parengta autoriaus

Paleistas ANN modelis sugeneravo išties gerus rezultatus per gan trumpą mokymosi laiką. Statistiniai rezultatai pavaizduoti 7 lentelėje. ANN modelis įvykdė mokymąsi per beveik tris minutes, įskaitant gan didelį duomenų kiekį, kurį turėjo apdoroti. Tačiau šiame tyrime mes nesiekiame modelio mokymosi greičio, mums svarbiausia pasiekti tikslių prognozavimo rezultatų. Atlikimo koeficientas yra sąlyginai žemas, šis rodiklis bus interpretuojamas tiksliau kelių modelių palyginimui. Įprastai, iš kelių modelių, mokytų naudojant skirtingus parametrų rinkinius, galiausiai bus pasirinktas tas modelis, kuris geriausiai veikia pagal pasirinktą našumo kriterijų. Gradientas yra labai arti vieneto, kas reiškia, kad kai kintamasis pasislenka vienu vienetu, išvesties rezultatas pasislenka per 1,71, kas statistiškai interpretuojama kaip tikslus rezultatas. Vidurkio reikšmė beveik lygi vienetinei koreliacijai tarp tikėtinų ir gautų reikšmių, kas dar kartą įrodo modelio tinkamumą indekso prognozavimui.

Žemiau esantis grafikas atspindi laiko eilučių judėjimą viso mokymosi bei prognozavimo laiko metu. Pastebėta, jog modelis veikė be didelių klaidų beveik visą laikotarpį, reikšmės šiek tiek padidėjo laikotarpio pabaigoje ir dėl to mūsų statistiniai parodymai teikia ne idealų modelio tikslumą. Svarbu turėti omenyje, jog indeksas kurį laiką judėjo labai stabiliai, o nuo 2019 metų vidurio pastebimi didesni svyravimai dėl makroekonominių faktorių, todėl galimai modeliui paprasčiausiai pritrūko duomenų mokymosi procesui, kurie įtraukia didesnę variaciją.



Pav. 15 ANN išvesties elemento atsakas

Šaltinis: parengta autoriaus (MATLAB)

Įdomu tai, jog ANN algoritmas puikiai geba susitvarkyti su dideliu duomenų kiekiu, įskaitant skirtingus kintamuosius. NARCH pritaikytas algoritmas modeliui išsprendė duomenų išsibarstymo problemą, kadangi naudotų kintamųjų reikšmės gerokai skyrėsi tarpusavyje, todėl nagrinėjant kintamųjų integraciją, rezultatai neparodė jokių didelių nukrypimų.

Puikus mokymosi rezultatas taip pat matomas 18 paveiksle, grafikas atspindi laiką, nuo kada modelis pasiekė geriausią galimą rezultatą. Jau antrajame trečdalyje proceso mokymasis buvo tikslingas, tačiau testavimas šiek tiek atsiliko.

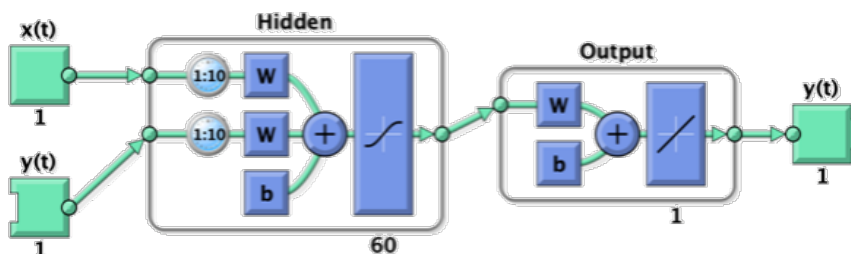


Pav. 16 ANN modelio geriausias mokymosi atlikimas

Šaltinis: parengta autoriaus (MATLAB)

Tačiau, norėdami atlikti platesnį vertinimą, ir patikrinti hipotezę, jog išplėtus neuroninių tinklų kiekį modelio veikimas gali tapti gerokai tikslesnis, atlikome modelio tobulinimo procesą, pakeisdami ANN modelio architektūrą.

Modelio tobulinimui buvo išplėstas paslėptų neuronų kiekis iki 60.



Pav. 17 ANN patobulinto suformuoto modelio architektūra

Šaltinis: parengta autoriaus (MATLAB)

ANN modelis: NARX, Hidden Neurons (60), Number of delays(10), Bayesian Regularization, data structure 80%-10%-10%.

Gauti rezultatai:

12 lentelė. ANN patobulinto modelio rezultatai

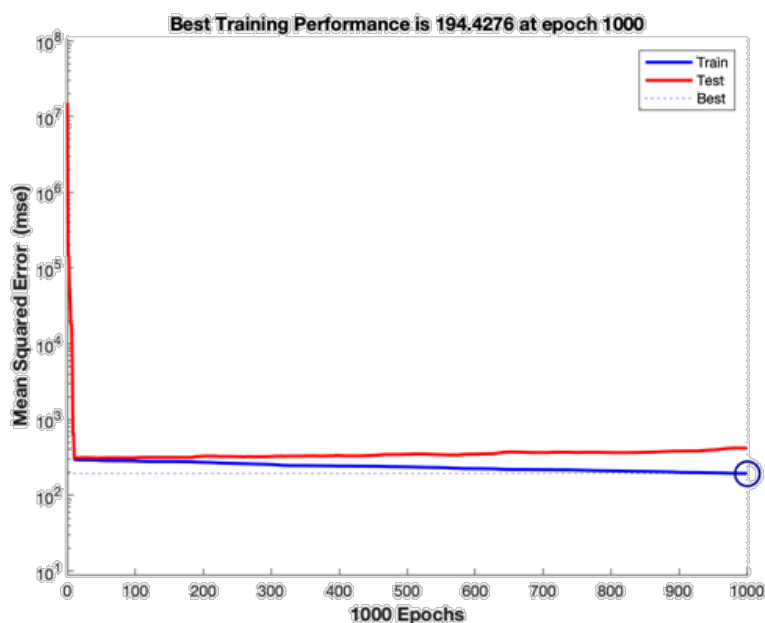
Parametras	Gauta reikšmė	Optimali reikšmė
Laikas (angl. Time)	0:07:13	Neapibrėžta.

Atlikimas (angl. Performance)	1,29	Nulis. (Kuo mažesnis teigiamas skaičius)
Gradientas (angl. Gradient)	3,62	Vienas. (Kuo arčiau vieneto)
MSE (angl. Mean Square Error)	1,48	Nulis. (Žemesnė reikšmė geresnė)
R (Vidurkis)	0,97	Vienas. (artima koreliacija)

Šaltinis: parengta autoriaus

Kaip ir buvo tikėtasi, neuroninio tinklo išplėtimas pastebimai pagerino algoritmo atlikimą. Tiek MSE, tiek R rodikliai priartėjo arčiau prie tobulo prognozavimo rezultato. Deja, bet tinklo išplėtimas sąlygojo gerokai ilgesnį modelio mokymosi laiką. Tačiau, kaip ir minėta anksčiau, tai nėra kertinis rodiklis mūsų prognozavimo tikslumui ir gebėjimams įvertinti.

ANN patobulintas modelis rodo geresnį mokymosi proceso atlikimą, tai galime matyti žemiau esančiame paveiksle. Testavimo kreivė priartėjo arčiau tikėtino rezultato, nei pirmiau sukonstruoto modelio.

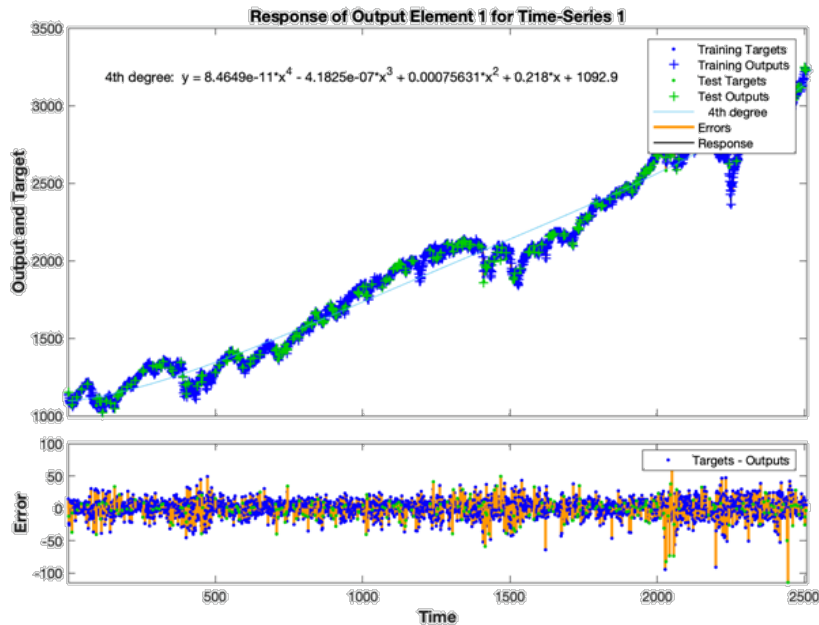


Pav. 18 ANN patobulinto modelio geriausias mokymosi atlikimas

Šaltinis: parengta autoriaus (MATLAB)

Deja, bet klaidų pasiskirstymas mokymosi periodo pabaigoje išliko panašus, gal tik sumažėjo nuokrypio amplitudė, tačiau net ir išplėtus neuroninį tinklą, mums nepavyko pašalinti problemos dėl mokymosi kiekio ir galimo duomenų nepakankamumo. (Pav.19) Ištirta, jog papildomų kintamųjų įvedimas, įtraukiant makroekonominis rodiklius, gali prisidėti prie šios problemos eliminavimo, tačiau geriausias rezultatas būtų pasiektas tik integruojant į modelį papildomus mašininio mokymosi algoritmus, gebančius apdoroti socialinių tinklų, žiniasklaidos signalus. Taip pat tinkamas sujungimas

su ANN būtų DT sprendimų medžio modelis, kuris padėtų modeliui priimti atitinkamus sprendimus pagal nustatytus parametrus bei sąlygas.



Pav. 19 ANN modelio išvesties elemento atsakas

Šaltinis: parengta autoriaus (MATLAB)

Darome išvadą, kad ANN modelis suformuotas šiame tyrime gali būti naudojamas preliminariam indekso judėjimo prognozavimui atlikti, tačiau siekiant tikslesnių rezultatų, būtina iširti papildomas tobulinimo galimybes, tinklo išplėtimą. Pastebėta, kad modeliui taip pat gali trūkti techninės analizės duomenų, nes, kaip ir matėme mokymosi laikotarpio pabaigoje pradėjo didėti klaidų skaičius nuo didesnio indekso kainos svyravimo, kurių fundamentalūs rodikliai paprasčiausiai nesugebėjo apdoroti. Todėl, kyla abejonių, jog modelį pritaikius konkrečios įmonės akcijų kainos prognozei, kur pastebimas gerokai didesni svyravimai, nei indekso atveju, ar prognozės bus tikslios.

6.2.SVM modeliu grįstas S&P 500 indekso prognozavimas

6.2.1. Duomenų surinkimas ir apdorojimas

Pirmame SVM tyrimo etape buvo paruošti duomenis. Naudojantis S&P indekso uždarymo kainomis, pagal metodologinėje dalyje pateiktas formules, buvo apskaičiuoti šešių reikalingų kintamųjų reikšmės (6 lentelė). Reikšmių apskaičiavimui papildomai reikėjo aprašomosios statistikos duomenų priklausančių išvesties kintamajam S&P500 uždarymo kainoms.

13 lentelė. S&P uždarymo kainų kintamojo aprašomoji statistika

SP	
Mean	1962,61
Standard Error	11,74
Median	1986,48
Mode	1178,10
Standard Deviation	588,91
Sample Variance	346815,28
Kurtosis	-1,10
Skewness	0,20
Range	2217,44
Minimum	1022,58
Maximum	3240,02
Sum	4937923,88
Count	2516,00
Largest(1)	3240,02
Smallest(1)	1022,58
Confidence Level(95,0%)	23,02

Šaltinis: parengta autoriaus, naudojantis Excel „Data Analysis“ papildiniu

Atlikus skaičiavimus, visi reikalingi duomenys buvo įkelti į MATLAB tolimesniam SVM mokymuisi.

6.2.2. SVM Modelių mokymasis ir tobulinimas

Pirmajame modelių mokymosi etape buvo testuojami šeši SVM metodai (pagal šešis skirtingus regresijos tipus).

Rezultatų lentelėje modeliai įvardijami atitinkamais skaičiais:

1.1 - Linear SVM

1.2 - Quadratic SVM

1.3 - Cubic SVM

1.4 - Fine Gaussian SVM

1.5 - Medium Gaussian SVM

1.6 - Coarse Gaussian SVM

Mokymosi rezultatai pateikti žemiau esančioje lentelėje:

14 lentelė. SVM šešių pirminių modelių rezultatai

	<u>1.1</u>	<u>1.2</u>	<u>1.3</u>	<u>1.4</u>	<u>1.5</u>	<u>1.6</u>
RMSE	38	48	48	89	45	28
R-Squared	1	0,99	0,99	0,98	0,99	1,00
MSE	14,59	23,66	23,09	79,64	20,39	8,3
MAE	3,2	4,4	4,1	6,1	3,4	2,2
Time (sec.)	0,292	0,206	0,214	0,145	0,0849	0,106

Šaltinis: parengta autoriaus

Visų modelių grafikai, kurie bus toliau naudojami interpretacijai, pateikti prieduose atitinkamai:

- Priedas Nr.1 - SVM Modelių Atsako reikšmių grafikai
- Priedas Nr.2 - SVM Modelių Numatytų/Faktinių reikšmių grafikai
- Priedas Nr.3 - SVM Modelių Atsako Reikšmių grafikai

Mokymosi rezultatai parodė labai tikslius rezultatus, pirmasis skirtumas su ANN modeliais identifiukuotas neįtikėtinais greitas mokymosi atlikimo laikas – sistema apdorojo duomenis per nepilną minutę, ko nepasakysi apie ANN modelius. R-squared vidurkio matas, rodantis koreliaciją tarp išvesties kintamojo bei modelio prognozuotų duomenų beveik visais atvejais rodo nepriekaištingą tikslumą, kas leidžia manyti, jog prognozuojant visiškai naujus modelius SVM veiks taip, kaip tikimasi. Vienetinė koreliacija gerai matoma grafikuose (žiūrėti Priedas Nr.2).

Deja, bet žiūrint į klaidų rodiklius (MSE, MAE), vis tiek pastebėtas prognozuotų duomenų išsibarstymas bei paklaidos. Prasčiausiu rezultatus sugeneravo SVM modelis 1.4, su puikia Gauso Regresija, o tiksliausiai suprognozavo SVM su šiurkščiu Gausu. Iš gautų rezultatų galima formuoti išvadą, jog techniniai rodikliai mašininio mokymosi metodams suteikia kur kas geresnio tikslumo. Kitą vertus, kadangi RMSE rodiklis yra vis dar nutolęs nuo nulio, kas pasako jog esama paklaidų, modelio tobulinimo etape bandyta optimizuoti modelį, tam, kad patikrinti, ar galima pasiekti dar didesnio tikslumo.

Į SVM modelį integravus pagrindinius taškų rinkinio komponentus (angl. Principal component analysis PCA), kas reiškia pagrindinių komponentų skaičiavimo ir jų naudojimo duomenų bazės keitimui procesas, buvo siekiama pagerinti SVM modelio veikimo principą transformuojant ir

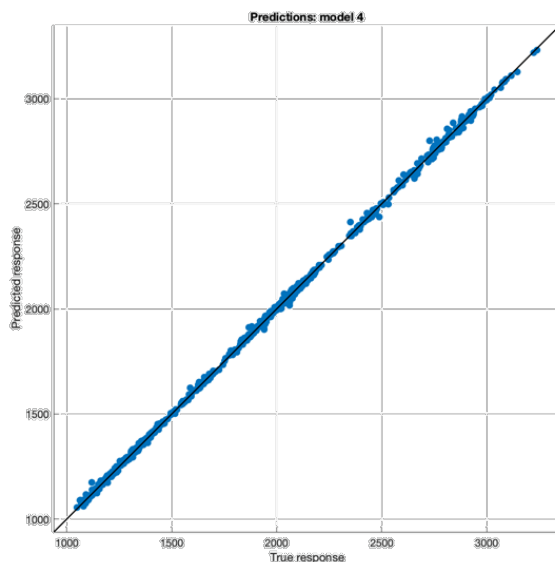
normalizuojant kintamųjų duomenis. Nors ir modelis pasiekė R-squared vieneto reikšmę, tačiau nesumažino RSME reikšmės, kuri pasiekė 31. Todėl teigiame, jog mūsų tyrime PCA integracija netiko SVM modelio tobulinimui.

Toliau, bandoma sukurti optimizuota SVM modelį keičiant pagrindinius parametrus. Optimizuotas SVM gali būti laikomas galingu įrankiu, nes vienu metu atlieka kelis žingsnius, pradedant nuo duomenų apdorojimo ir standartizavimo (išvesties reikšmes transformuoja į „true“ ir „false“), iš visų regresijos tipų pasirinktas tinkamiausias esamam duomenų rinkiniui, nustatomas tiek epsilon, tiek dežutės apribojimas (*angl. Box constraint*). Patobulintas SVM modelis turi žemiau išvardintus parametrus:

- Modelio tipas: Optimizuotas SVM
- Kernelio funkcija: Kvadratinė (*angl. Quadratic*)
- Dežutės apribojimas: 987,13 (*angl. Box Constraint*)
- Epsilon: 0,7903 (*angl. Epsilon*)
- Stantatizavimas: taikytas

Papildomai, modelis atliko Bajeso optimizavimą kintamųjų reikšmėms normalizuoti.

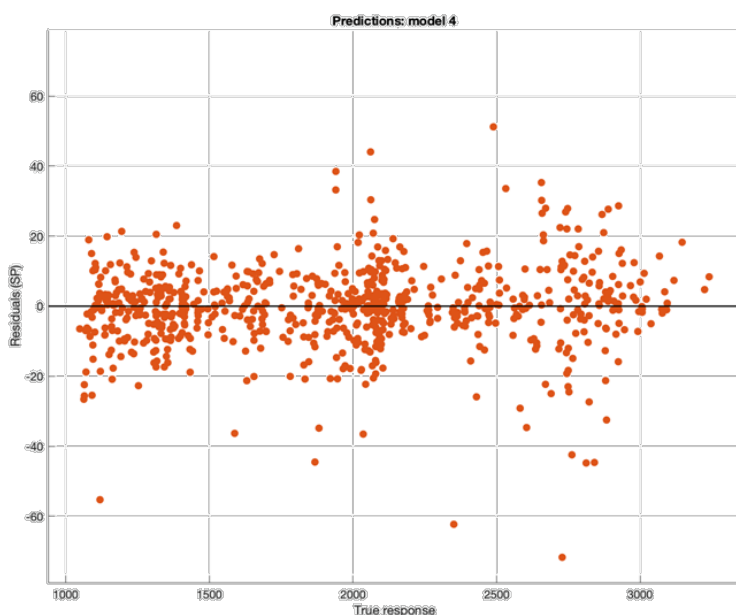
Paleidus šį modelį mokymuisi, buvo gauti geriausi ir tiksliausi rezultatai. Šį kartą modelio mokymosi laikas buvo kiek ilgesnis, tačiau priimtinas – 252 sekundės (~3 minutės). RMSE rodiklis sumažėjo iki 11, R-squared parametras išliko vienetas, MSE rodiklis beveik priartėjo prie nulio (gauta 1,35 reikšmė), o MAE – 7,96. Numatytu/Faktinių reikšmių grafikas lyginant su „hyperplane“ atrodo neįtikėtinais tiksliai, beveik visos reikšmės išsibarsčiusios ant tiesiosios kreivės:



Pav. 20 SVM optimizuoto modelio Numatytu/Faktinių reikšmių grafikas

Šaltinis: parengta autoriaus (MATLAB)

Žvelgiant į klaidos pasiskirstymo grafiką, kaip ir minėta metodologinėje dalyje, mes tikimės, jog visos reikšmės bus išsibarsčiusios aplink 0 X-ašį. Žemiau esantis grafikas tai atspindi:



Pav. 21 SVM Optimizuoto modelio Atsako Reikšmių grafikas

Šaltinis: parengta autoriaus (MATLAB)

Laikome, jog modelio tobulinimo etapas buvo sėkmingas, ir tyrimo metu išsiaiškinta, jog S&P indekso prognozavimui labiausiai tinka SVM Optimizuotas modelis su anksčiau išvardintais parametrais. Mano manymu, įtraukus į šį modelį papildomus techninius rodiklius, rezultatai gali būti dar labiau patobulinti, todėl ši mintis bus išplėta atliekant tolimesnius ML tyrimus ateityje ir bandoma pritaikyti prognozuojant konkrečių įmonių akcijų kainas.

6.3. Tyrimo rezultatai

Atlikus ANN modelio tyrimą, gauta, jog modelis yra sąlyginai tinkamas ir geba prognozuoti indekso judėjimo kryptį.

ANN modelis dažniausiai įraukia fundamentalios analizės rodiklius, mūsų tyrime mes pasirinkome net dvidešimt kintamųjų, tarp kurių buvo tiek makroekonominiai rodikliai, tiek valiutų kursai, tačiau net ir tobulinimo etape nepavyko pasiekti kuo tikslesnio modelio. Tyrimo rezultatai rodo, jog atlikus papildomus modelio tobulinimo bandymus būtų galima ANN modelį naudoti rinkos indeksų judėjimui nuspėti, tačiau kainų prognozei jis nėra pakankamas. Kaip ir pastebėjome mokslinėje literatūroje, prieš tai atlikti tyrimai taip pat iškelia šią problemą, ir rekomenduoja nenaudoti ANN kaip pavienio modelio, o konstruoti hibridinius modelius prijungiant kitus mašininio mokymosi modelius.

Atlikus SVM grįstu modelių testavimą, pagrindiniai skirtumai iš karto išryškėjo. Kalbant apie SVM, pasiekus optimizuota modelį (SVM su kvadratine regresija), pavyko sugeneruoti gana tikslių prognozavimui skirtą įrankį, su vienetine R-square reikšme, žemu RMSE rodikliu, bei greitu atlikimo laiku. Rezultatai parodė, jog techninių rodiklių naudojimas ženkliai pagerino prognozavimo tikslumą.

Atliekant tyrimą, buvo susidurta su tam tikrais veiksniais, kurie kliudė dar labiau išplėtoti tyrimo rezultatus. Pirmiausiai, buvo pritrūkta programavimo įgūdžių tam, kad savarankiškai būtų galima atlikti visus modelio kūrimo nuo pradžios iki pabaigos žingsnius, kai kuriose vietose atlikti papildomas kintamųjų reikšmių normalizacijas, bei valdyti duomenis. Išlieka klausimas, kokią tiksliai įtaką kintamųjų reikšmių išsibarstymas turėjo modelių tikslumui. Taip pat, nepakako techninių įrankių sugeneruoti ir įtraukti papildomus kintamuosius. Kita vertus, tyrimui atlikti kliudę veiksniai tapo tolimesnių tyrimų šia tematika vienu iš probleminių klausimų.

Tyrimą atlikti padėjo gan platus mokslinės literatūros prieinamumas, gebėta detaliai išnagrinėti mašininio mokymosi modelius, tam kad suprasti pagrindinius konceptus ir pritaikyti kuriamiems modeliams. Taip pat, MATLAB programinė įranga pasitarnavo kaip galingas įrankis modelių kūrimui bei testavimui. Platus papildinių pasirinkimas labai prisidėjo prie kiekvieno iš tyrimo etapų ir kai kuriuose žingsniuose išsprendė programavimo žinių trūkumo problemą.

Šio tyrimo išvada yra tokia, jog S&P 500 indekso kitimui tikslesnis yra SVM modelis, kadangi jam taikyti kintamieji yra techninės analizės rodikliai, apskaičiuoti matematinėmis formulėmis, kas užtikrino duomenų rinkinio patikimumą bei sumažino verčių išsibarstymą. Įrodyta, jog laiko eilučių prognozavimui regresijos logika paremta metodika geba tiksliau prognozuoti indekso kainų kitimą.

7. Išvados ir pasiūlymai

1. Dirbtinio intelekto pritaikomumas akcijų kainų prognozėje ypač plačiai tiriamas nuo 2010-ųjų metų. Mašininio mokymosi ir dirbtinio intelekto metodai teikia didelę naudą finansinių sprendimų priėmimo, atsižvelgiant į naujai atsirandančius modeliavimo ir prognozavimo metodus. Pagrindiniais sunkumais išlieka duomenų prieiga bei ribotumas, taip pat pačių algoritmų pažinimas bei jų potencialas, galimybės. Atlikus sisteminę literatūros analizę, išsiaiškinta jog, nors ir egzistuoja gan daug ML modelių, visi turi tam tikrų trūkumų ir nėra iki galo išdirbti. Atlikta analizė taip pat parodė, jog daugiausiai tyrimų yra atliekama prognozavimui pritaikant ANN modelius, po jų seka SVM, kitų modelių taikymas yra gana retesnis.

2. Tiek ANN, tiek SVM metodai yra galingi ir sudėtingi ML algoritmai, gebantys atlikti laiko eilučių prognozavimą. Vienas pagrindinių jų skirtumų yra kintamųjų pasirinkimas. ANN naudoja fundamentalia analize grįstus rodiklius, todėl įvesčiai reikalauja gana didelio duomenų kiekio. SVM modelis remiasi technine analize, jam atlikti reikia labiau matematinės prieigos, o SVM regresija geba atlikti tikslių prognozavimą nereikalaujant didelio duomenų rinkinio.

3. Abiejų modelių rezultatai gauti gana panašūs. Daroma išvada, kad SVM ir ANN yra efektyvūs mašininio mokymosi modeliai atliekant S&P indekso kainos kitimo prognozavimą. Kadangi ANN algoritmas mokymuisi reikalauja daugiau laiko ir duomenų, tyrimas neparodė tobulo tikslumo konkrečiai laiko eilutei prognozuoti. Kaip ir buvo nagrinėta mokslinėje literatūroje, rekomenduojama nenaudoti ANN kaip pavienio modelio, o konstuoti hibridinius algoritmus prijungiant kitus ML metos. Kitą vertus, tyrime sukonstruotas modelis geba prognozuoti indekso kitimo krypti, tačiau tikslios reikšmės nuspėti negali dėl aukštų paklaidų vertinimui naudotų indikatorių reikšmių.

4. SVM tyrimas parodė kur kas tikslesnius prognozavimo rezultatus – regresija buvo atlikta testuojant SVM šešiais skirtingais regresijos tipais - linijiniu, kvadratinu, kubiniu, puikiu Gauso, vidutiniu Gauso ir šiurkščiu Gauso. Kintamųjų normalizacijai atlikti buvo naudotas Bajeso optimizavimo algoritmas. Tiksliausias prognozavimo rezultatas pasiektas optimizuotu SVM, kvadratinio regresijos tipo funkcija, vidurkio kvadrato reikšmei pasiekus vienetą, taip pat klaidų statistiniai rodikliai MSE ir MAE rodė žemiausias reikšmes lyginant su kitais testuotais modeliais. Tyrimui pasirinkti techniniai rodikliai įrodė, kad jie yra labiau tikslūs prognozuojant indekso kainų judėjimo kryptį.

5. Išsiaiškinus, jog ML algoritmai veikia tiksliau, kai jiems suteikiamas kuo didesnis mokymosi duomenų spektras, padaryta išvada, kad tiek ANN, tiek SVM gali būti dar labiau tobulinami įvedant papildomus kintamuosius, tokius kaip tokie kaip užsienio valiutų kursai, palūkanų normos ir vartotojų kainų indeksas, išplėčiant mokymuisi skirto periodo laikotarpį, taip pat jungiant juos su kitais ML modeliais ir kuriant hibridinius algoritmus.

6. ANN modeliui tobulinti siūloma papildomai prijungti DT algoritmą, išplėsti neuroninį tinklą bet įtraukti daugiau kintamųjų, ypatingai techninės analizės rodiklių, kurių ANN trūksta. SVM tobulinimui siūloma išplėsti kintamųjų aibę įtraukiant kitus veiksnius, turinčius įtaką kainoms, nes šiuo atveju tyrimas apėmė tik S&P indekso techninę analizę, todėl tyrimo metu suformuotas optimizuotas SVM nebūtinai tiks kitų turto klasių prognozavimui.

8. Literatūros sąrašas

Aharon, M, M Elad, and A Bruckstein. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation." *Signal Processing, IEEE Transactions* (2016): 4311–4322

Altman, E. I., Marco, G., and Varetto, F. „Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)“. *Journal of Banking & Finance*, (1994)18(3):505–529.

Asadi S., E. Hadavandi, F. Mehmanpazir, M.M. Nakhostin, “Hybridization of evolutionary Levenberg–Marquardt neural networks and data preprocessing for stock market prediction”, *Knowl.-Based Syst.* 35 (2012) 245–258.

Aziz S., Dowling M., Hammami H. „Machine learning in finance: A topic modeling approach“. *Rennes School of Business*, (2019) France. JEL codes: G00, C45

Bahrammirzaee Arash. “A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems”. *Neural Comput & Applic* (2010) 19:1165–1195. DOI 10.1007/s00521-010-0362-z

Bozinovski S., "Teaching space: A representation concept for adaptive pattern classification" *COINS Technical Report No. 81-28*, Computer and Information Science Department, University of Massachusetts at Amherst, MA, (1981). <https://web.cs.umass.edu/publication/docs/1981/UM-CS-1981-028.pdf>

Bozinovski, S. "A self-learning system using secondary reinforcement" . In *Trapp, Robert (ed.). Cybernetics and Systems Research: Proceedings of the Sixth European Meeting on Cybernetics and Systems Research*. (1982) North Holland. pp. 397–402. ISBN 978-0-444-86488-8.

Bozinovski, Stevo. "Modeling mechanisms of cognition-emotion interaction in artificial neural networks, since 1981." *Procedia Computer Science*. (2014) p. 255-263

Bozinovski, S. "Self-learning agents: A connectionist theory of emotion based on crossbar value judgment." *Cybernetics and Systems* (2001) 32(6) 637-667.

Cerchiello, P., Giudici, P., and Nicola, G. „Twitter data models for bank risk contagion. *Neurocomputing*“, (2017) 264:50–56.

Chandola, V.; Banerjee, A.; Kumar, V. "Anomaly detection: A survey". *ACM Computing Surveys*. (2009) 41 (3): 1–58. doi:10.1145/1541880.1541882.

Cheng, D. and Cirillo, P. „A reinforced urn process modeling of recovery rates and recovery times“. *Journal of Banking & Finance*, (2018) 96:1–17.

Coates, Adam; Lee, Honglak; Ng, Andrew Y. “An analysis of single-layer networks in unsupervised feature learning”. *Int'l Conf. on AI and Statistics (AISTATS)*. (2011)

Daniel Jurafsky; James H. Martin (2009). *Speech and Language Processing*. Pearson Education International. pp. 145–146.

- Duda, R., Hart P. "Pattern Recognition and Scene Analysis", *Wiley Interscience*, (1973)
- Ethem A.. Introduction to Machine Learning. *MIT Press*. (2010) p. 9. ISBN 978-0-262-01243-0.
- Ethem Alpaydin. "Introduction to Machine Learning (Fourth ed.)". *MIT* (2020) pp. xix, 1–3, 13–18. ISBN 978-0262043793.
- Gupta A., B. Dhingra, "Stock market prediction using hidden markov models", *Proceedings of Students Conference on Engineering and Systems*, (2012), pp. 1–4.
- Guresen E. , G. Kayakutlu, T.U. Daim, "Using artificial neural network models in stock market index prediction", *Expert Syst. Appl.* 38 (8) (2011) 10389–10397.
- Hadavandi E., H. Shavandi, A. Ghanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting", *Knowl.-Based Syst.* 23 (8) (2010) 800–808.
- Harnad, Stevan, "The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence", in Epstein, Robert; Peters, Grace (eds.), *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, (2008) pp. 23 66, ISBN 9781402067082
- Hodge, V. J.; Austin, J. "A Survey of Outlier Detection Methodologies". *Artificial Intelligence Review*. (2004) 22 (2): 85 126. CiteSeerX 10.1.1.318.4023. doi:10.1007/s10462-004-4304-y.
- Jordan, Michael I.; Bishop, Christopher M. "Neural Networks". In *Allen B. Tucker (ed.). Computer Science Handbook, Second Edition (Section VII: Intelligent Systems)*. (2004) ISBN 978-1-58488-360-9.
- Kercheval, A. N. and Zhang, Y. „Modelling high-frequency limit order book dynamics with support vector machines“. *Quantitative Finance*, (2015) 15(8):1315–1329.
- Mitchell, T. Machine Learning. *McGraw Hill*. (1997) p. 2. ISBN 978-0-07-042807-2.
- More A., P.U. Rathod, R.H. Patil, D.R. Sarode, B.E. Student, "Stock market prediction system using hadoop", *Internat. J. Engrg. Sci.* (2018) 16138.
- Nian, K., Coleman, T. F., and Li, Y. „Learning minimum variance discrete hedging directly from the market“. *Quantitative Finance*, (2018) 18(7):1115–1128.
- Nilsson N. *Learning Machines*, McGraw Hill, 1965.
- Otterlo, M.; Wiering, M. "Reinforcement learning and markov decision processes". *Reinforcement Learning. Adaptation, Learning, and Optimization*. (2012) pp. 3–42. doi:10.1007/978-3-642-27645-3_1. ISBN 978-3-642-27644-6.
- Oztekin A., R. Kizilaslan, S. Freund, A. Iseri, "A data analytic approach to forecasting daily stock returns in an emerging market", *European J. Oper. Res.* 253 (3) (2016) 697–710.

Patel J., S. Shah, P. Thakkar, K. Kotecha, “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques”, *Expert Syst. Appl.* 42 (2015) 259–268.

Pavel Brazdil, Christophe Giraud Carrier, Carlos Soares, Ricardo Vilalta. „Metalearning: Applications to Data Mining (Fourth ed.) “. *Springer Science+Business Media.* (2009) pp. 10–14, passim. ISBN 978-3540732624.

Rout A.K, B. Biswal, P.K. Dash, “A hybrid FLANN and adaptive differential evolution model for forecasting of stock market indices”, *Int. J. Knowl.-Based Intell. Eng. Syst.* 18 (1) (2014) 23–41.

Sezer, O. B. and Ozbayoglu, A. M. „Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach“. *Applied Soft Computing*, (2018) 70:525–538.

Ticknor J.L., “A Bayesian regularized artificial neural network for stock market forecasting, *Expert Syst*”. *Appl.* 40 (14) (2013) 5501–5506.

Tillmann, A. M. "On the Computational Intractability of Exact and Approximate Dictionary Learning". *IEEE Signal Processing Letters.* (2015) 45–49. doi:10.1109/LSP.2014.2345761.

Varetto, F. „Genetic algorithms applications in the analysis of insolvency risk“. *Journal of Banking & Finance*, (1998) 22(10-11):1421–1439.

Wen Q., Z. Yang, Y. Song, P. Jia, “Automatic stock decision support system based on box theory and SVM algorithm”, *Expert Syst. Appl.* 37 (2) (2010) 1015–1022.

Weng, B., Lu, L., Wang, X., Megahed, F. M., and Martinez, W. „Predicting short-term stock prices using ensemble methods and online data sources“. *Expert Systems with Applications*, (2018) 112:258–273.

Xu B., D. Zhang, S. Zhang, H. Li, H. Lin, “Stock market trend prediction using recurrent convolutional neural networks”, *Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing*, (2018), pp. 166–177.

Zhong X., D. Enke, “Forecasting daily stock market return using dimensionality reduction”, *Expert Syst. Appl.* 67 (2017) 126–139.

Zimek, Arthur; Schubert, Erich, "Outlier Detection", *Encyclopedia of Database Systems*, Springer New York, (2017) pp. 1–5, doi:10.1007/978-1-4899-7993-3_80719-1, ISBN 9781489979933

9. Santrauka

Akcijų kainų prognozavimas remiantis dirbtiniu intelektu pastaruoju laikotarpiu tampa vis labiau plėtojama tyrimų tema. Vyraujant nežinomumui finansų rinkose sparčiai besikeičiančiame pasaulyje, ieškoma naujų sprendimų, gebančių įtraukti kuo daugiau žinomų akcijų kainoms įtaką turinčių veiksnių. Tobulėjant informacinėms technologijoms, akcijų prognozavimą tradiciniais būdais pakeitė mašininis mokymasis, kuris šiai dienai turi didelį potencialą pritaikomumui finansų rinkose, ypač laiko eilučių prognozavime. Mašininis mokymasis, kaip AI atmaina, siūlo platų modelių spektrą laiko eilučių prognozavimui vykdyti, tačiau, kadangi tyrimų nėra daug ir vis dar nėra sukurto vieno tobulo algoritmo, kuris tiksliai prognozuotų kainų pokyčius, ši tema vis dar yra labai aktuali. Populiariausi tiriami metodai yra dirbtiniai neuroniniai tinklai (ANN) bei atraminių vektorių mašinos (SVM) – abu šie metodai įvardijami kaip galingi įrankiai, gebantys apdoroti didelį skirtingų duomenų kiekį. Vienas pagrindinių jų skirtumų yra tas, jog ANN remiasi fundamentalios analizės logika, tuo tarpu SVM – technine analize. Šio tyrimo problema iškelia klausimą, ar prognozuojant S&P 500 indekso kitimą ANN bei SVM metodais galima pasiekti tikslių rezultatų. Tyrimu siekiama išanalizuoti dirbtinio intelekto sampratą ir jo pritaikomumą akcijų kainų prognozavime, tuo remiantis sukurti tyrimo metodologiją bei atlikti empirinį tyrimą praktiškai pritaikant ANN ir SVM indekso kainų kitimui prognozuoti. Duomenų pagrindą sudaro dešimties metų laikotarpio (2010.01.01 – 2020.01.01) S&P 500 indekso uždarymo kainos. ANN modeliui naudoti dvidešimt kintamųjų, suskirstyti į šešias grupes - S&P indekso grąža, finansiniai ir ekonominiai indikatoriai (tokie kaip nafta, auksas), penkių didžiausių indeksui priklausančių įmonių grąža, valiutų kursų pokytis, keturių didžiausių akcijų indeksų grąžos bei prekybos apimtys. SVM modeliui taikyti kintamieji buvo techninės analizės rodikliai, tokie kaip paprastas slenkamasis vidurkis, svertinis slenkamasis vidurkis, santykinio stiprumo indeksas, kaupimo pasiskirstymo oscilatorius ir vidutinis tikrasis diapazonas. Tiek ANN, tiek SVM tyrimų eiga susideda iš duomenų paruošimo bei apdorojimo, modelio mokymosi, modelio testavimo ir modelio tobulinimo. Atlikus ANN modelio tyrimą, pastebėta, jog modelis geba prognozuoti S&P indekso kitimą, tačiau RMSE rodikliui esant toli nuo siekiamos reikšmės, modelis reikalauja papildomo tobulinimo. MAE, rodantis prognozės paklaidą laikotarpio pabaigoje yra gan didelis, todėl kyla abejonių, ar modelis tiksliai gebėtų prognozuoti kainų kitimą jį pritaikius ne tiek indeksui, o konkrečiam akcijų rinkiniui. SVM tyrimas parodė kur kas tikslesnius prognozavimo rezultatus – regresija buvo atlikta testuojant SVM šešiais skirtingais regresijos tipais - linijiniu, kvadratinu, kubiniu, puikiu Gauso, vidutiniu Gauso ir šiurkščiu Gauso. Kintamųjų normalizacijai atlikti buvo naudotas Bajeso optimizavimo algoritmas. Tiksliausias prognozavimo rezultatas pasiektas optimizuotu SVM, kvadratinio regresijos tipo funkcija, vidurkio kvadrato reikšmei pasiekus vieneta, taip pat klaidų statistiniai rodikliai MSE ir MAE rodė žemiausias reikšmes lyginant su kitais testuotais modeliais. Tyrimas įrodė, kad ML grįsti metodai yra tinkami indeksui prognozuoti, tačiau SVM, kuris remiasi techninės analizės rodikliais, geba tiksliau nei ANN prognozuoti laiko eilutes. Tyrimo metu susidurta su kliūtimis, tokiomis kaip techninės galimybės modifikuojant ir mokant pačius modelius bei galimą kintamųjų nepakankamumą. Išsiaiškinus, jog ML algoritmai veikia tiksliau, kai jiems suteikiamas kuo didesnis mokymosi duomenų spektras, padaryta išvada, kad tiek ANN, tiek SVM gali būti dar labiau tobulinami įvedant papildomus kintamuosius, išplėčiant mokymuisi skirtą laikotarpį, taip pat jungiant juos su kitais ML modeliais ir kuriant hibridinius algoritmus. ANN modeliui tobulinti siūloma papildomai prijungti DT algoritmą, išplėsti neuroninį tinklą bet įtraukti daugiau kintamųjų, ypač techninės analizės rodiklių, kurių ANN trūksta. SVM tobulinimui siūloma išplėsti kintamųjų aibę įtraukiant kitus veiksnus, turinčius įtaką kainoms, nes šiuo atveju tyrimas apėmė tik S&P indekso technine analize, todėl tyrimo metu suformuotas optimizuotas SVM nebūtinai tiks kitų turto klasių prognozavimui. Atliktas tyrimas papildė ML mokslinių tyrimų perspektyvą bei pateikė naujas prielaidas tolesnei tyrimų plėtotei.

Raktiniai žodžiai: Mašininis mokymasis, laiko eilučių prognozavimas, dirbtiniai neuroniniai tinklai, atraminių vektorių mašinos.

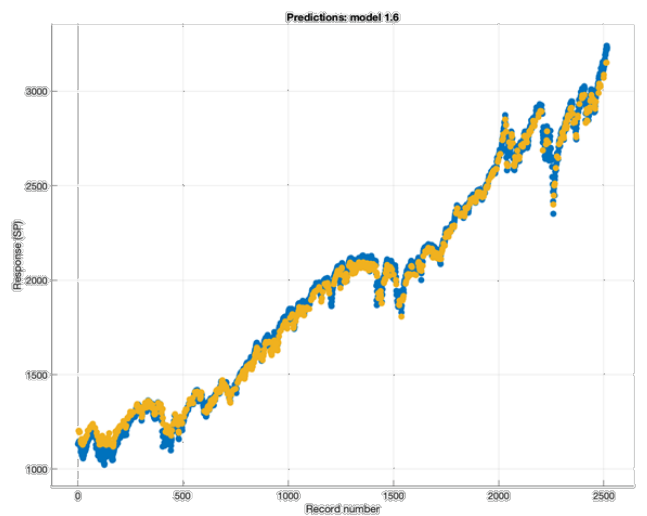
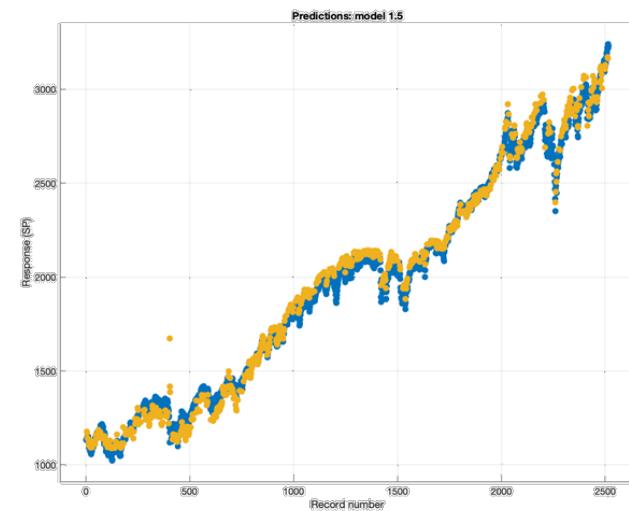
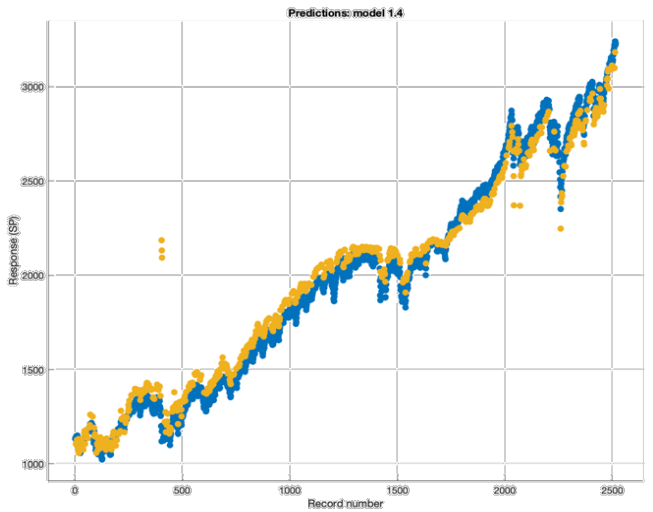
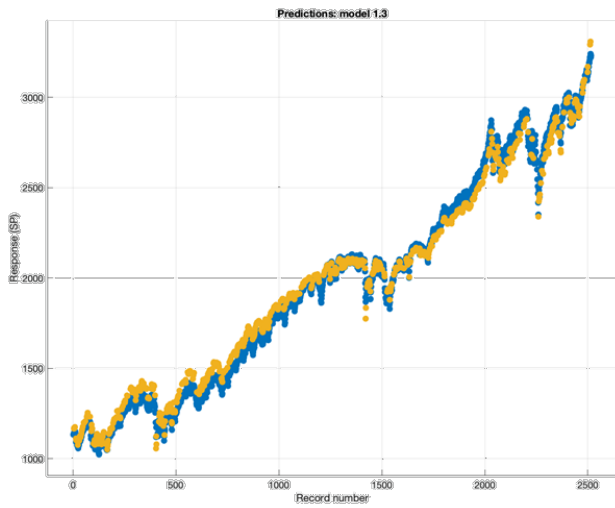
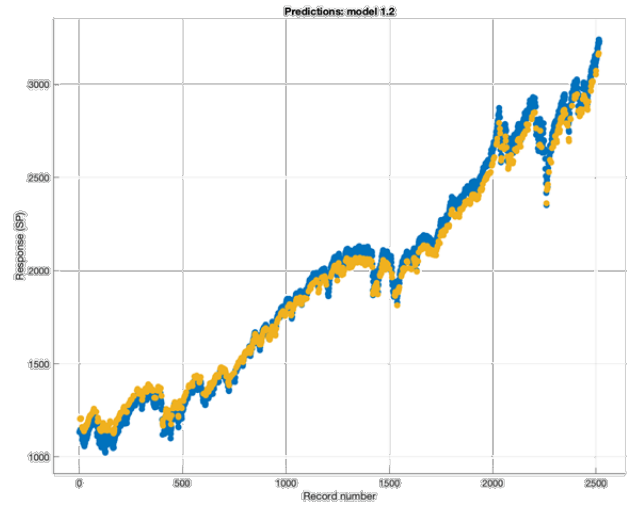
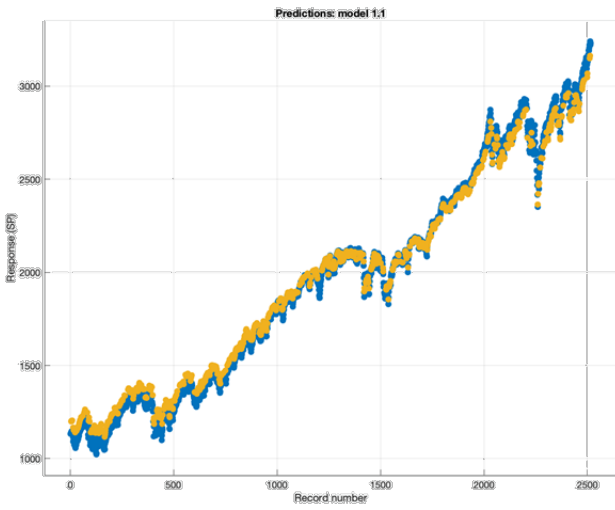
Summary

Stock price forecasting based on artificial intelligence methods has recently become an increasingly evolving research topic. With the uncertainty prevailing in the financial markets in a rapidly changing world, new solutions are being sought that are able to incorporate as many known factors as possible which influence stock prices. With the development of information technology, stock forecasting in the traditional ways has been replaced by machine learning techniques, which has great potential for applicability in financial markets, especially in time series forecasting. Machine learning, as a significant part of AI, offers a wide range of models for time series prediction, however, as there are not many researches available and no single perfect algorithm exists which is able to accurately predict price changes, this topic still remains very relevant. The most popular studied methods are artificial neural networks (ANN) and support vector machines (SVM), both of which are identified as powerful tools capable of processing large amounts of different data. One of their main differences is that ANN is based on the logic of fundamental analysis, while SVM is based on technical analysis. The problem of this study raises the question of whether accurate results can be obtained by predicting the change of the S&P 500 index using the ANN and SVM methods. The research aims to analyze the concept of artificial intelligence and its applicability in stock price forecasting, on the basis of which to develop a research methodology and conduct empirical research by practically applying the ANN and SVM to forecast index price changes. The data is based on the closing prices of the S&P 500 index for a period of ten years (01.01.2010 - 01.01.2010). The ANN model used twenty variables divided into six groups: S&P index returns, financial and economic indicators (such as oil, gold), returns of the five largest companies in the index, exchange rate changes, returns of the four largest stock indices, and trading volumes. The variables used for the SVM model were technical analysis indicators such as simple moving average, weighted moving average, relative strength index, accumulation distribution oscillator, and average true range. Both ANN and SVM based researches consist of the following steps such as data preparation and processing, model learning, model testing, and model refinement. After researching the ANN model, it was observed that the model is able to predict the change of the S&P index, but as the RMSE indicator is far from the target value, the model requires further improvement. The MAE showing the forecast error at the end of the period is quite large, so there are doubts whether the model would be able to accurately forecast price changes by applying it not only to index forecasting, but to the specific set of stocks. The SVM study showed much more accurate prediction results - regression was performed by testing SVM in six different types of regression - linear, quadratic, cubic, excellent Gaussian, mean Gaussian, and coarse Gaussian. The Bayesian optimization algorithm was used to normalize the variables. The most accurate prediction results were achieved by the optimized SVM, including the quadratic regression type function. The mean square value reached one, and the error statistics MSE and MAE showed the lowest values compared to the other models tested. The study proved that ML based methods are suitable for index prediction, but SVM, which relies on technical analysis indicators, is able to predict time series more accurately than ANN. The study encountered obstacles such as technical possibilities for modifying and teaching the models themselves, and possible failure of the variables. Finding that ML algorithms work more accurately when given the widest possible range of learning data, it is concluded that both ANN and SVM can be further improved by introducing additional variables, extending the learning period, and combining them with other ML models and developing hybrid algorithms. To improve the ANN model, it is proposed to additionally connect the DT algorithm, expand the neural network but include more variables, especially technical analysis indicators, which ANN is missing. To improve the SVM, it is proposed to expand the set of variables to include other factors influencing prices, as in this case the study covered only the technical analysis of the S&P index, therefore the optimized SVM generated during the study may not be suitable for forecasting other asset classes. The study complemented the ML research perspective and provided new assumptions for further research development.

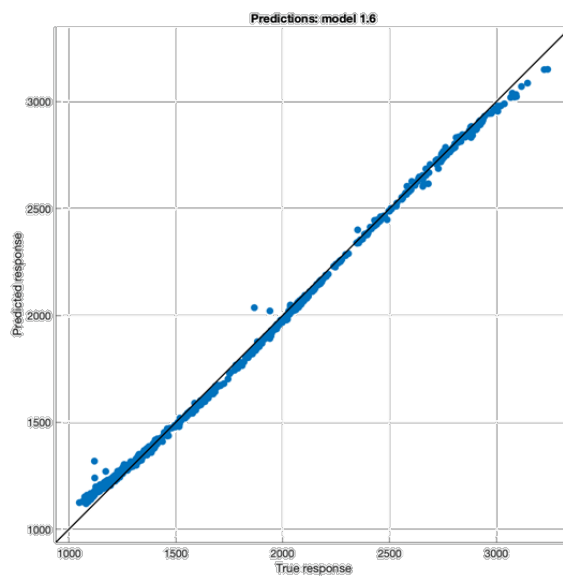
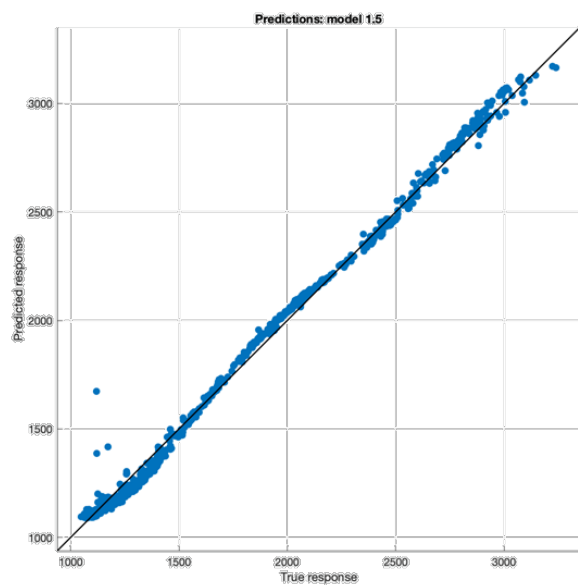
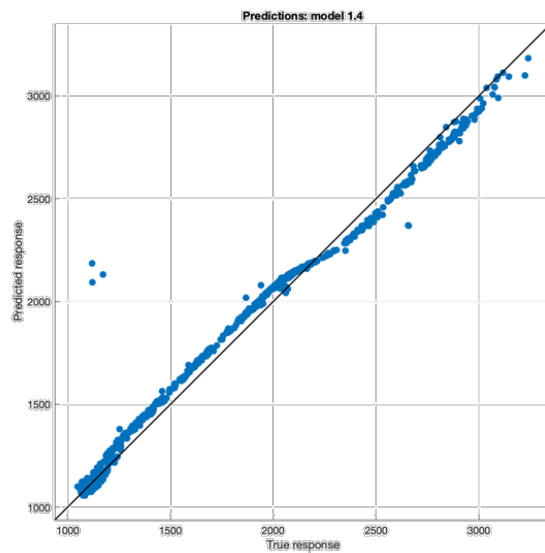
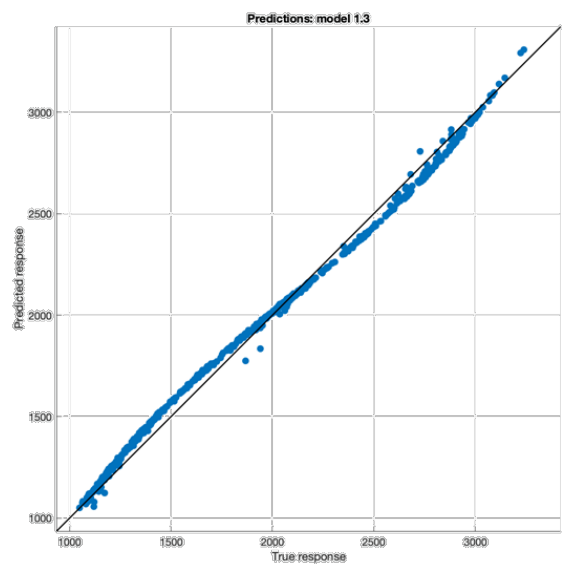
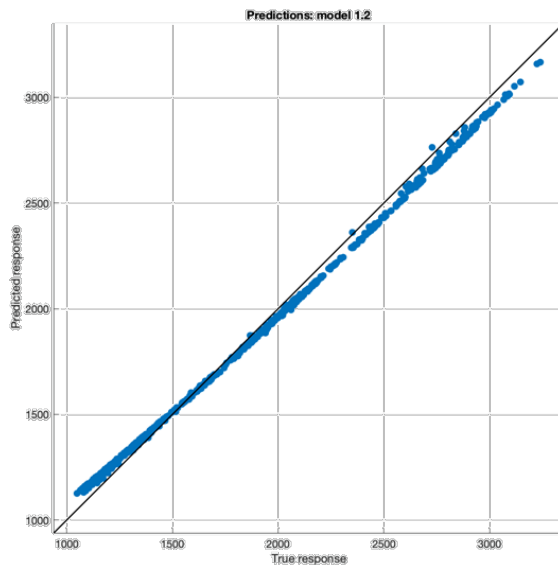
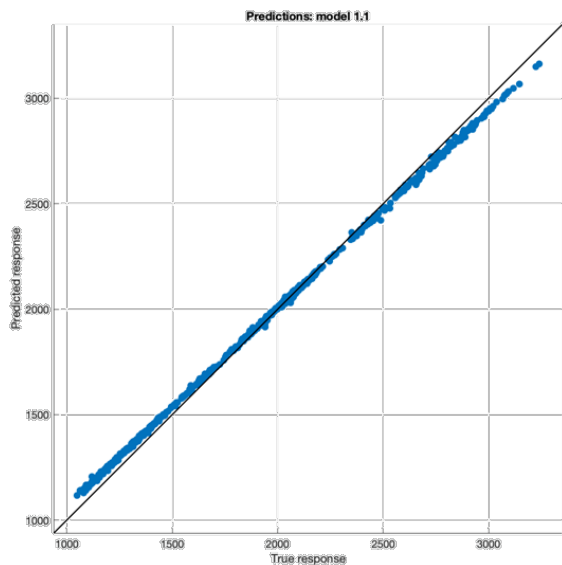
Keywords: Machine Learning, Time-series forecasting, Artificial Neural Networks, Support Vector Machines.

10. Priedai

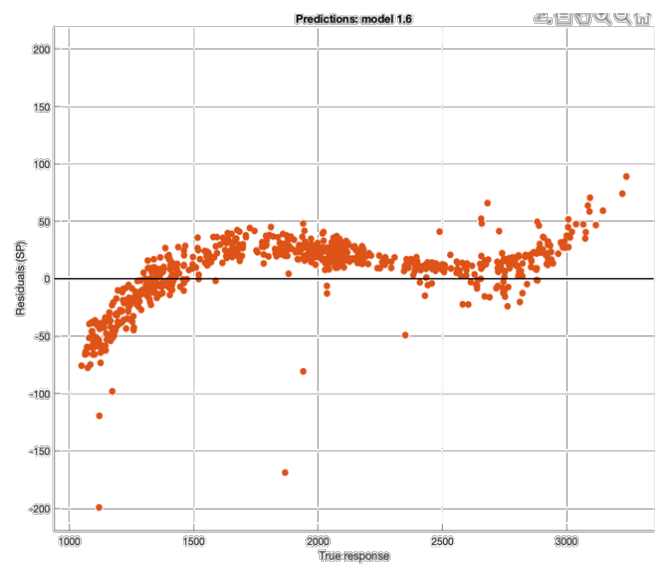
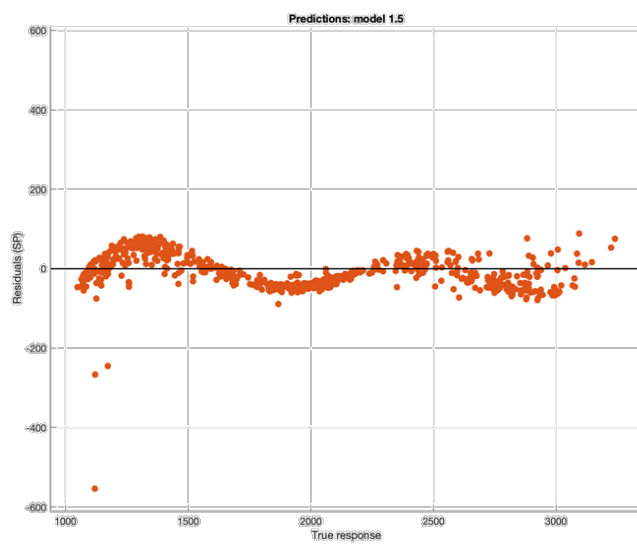
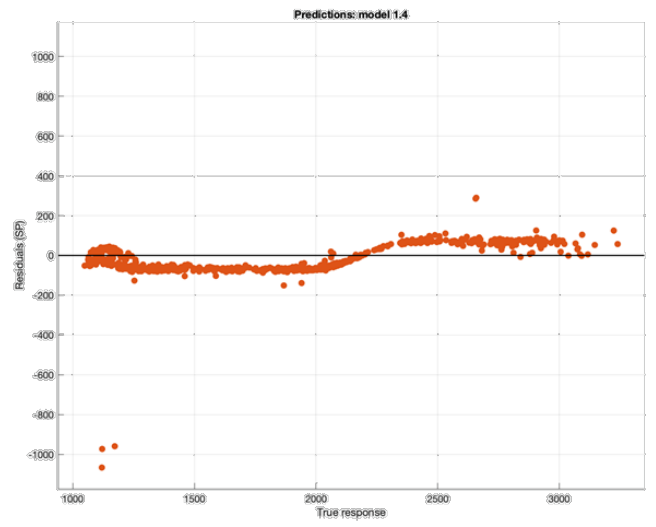
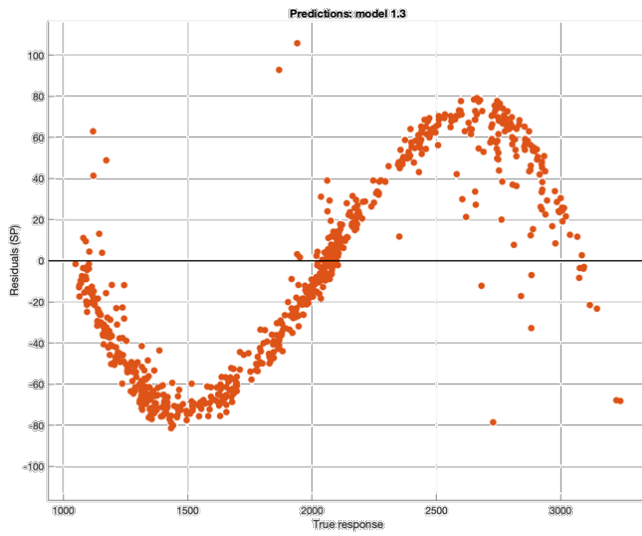
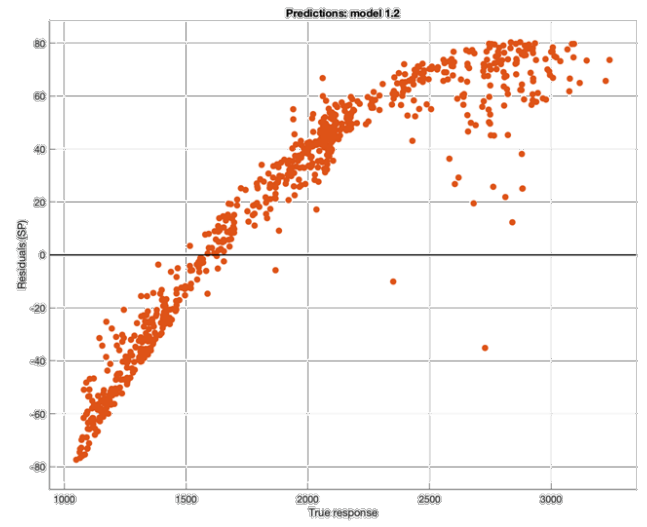
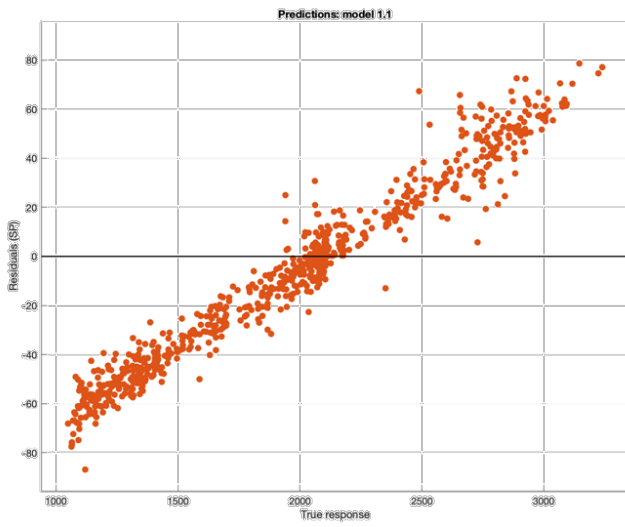
1. SVM Modelių Atsako reikšmių grafikai



2. SVM Modelių Numatytų/Faktinių reikšmių grafikai



3. SVM Modelių Atsako Reikšmių grafikai



4. Formuliu sąrašas

Formulė 1 ANN aktyvinimo funkcija	37
Formulė 2 ANN aktyvacijos išvesties funkcija.....	37
Formulė 3 ANN sklidimo funkcija.....	37
Formulė 4 Santykinio Stiprumo Indekso (RSI) skaičiavimas	37
Formulė 5 ANN Prekybos signalo skaičiavimas (SOT)	38
Formulė 6 MACD formulė.....	38
Formulė 7 ROC formulė	38
Formulė 8 SMA.....	39
Formulė 9 Slenkamojo vidurkio (SMA) multiplikatorius.....	39
Formulė 10 Ekspontinis slenkamasis vidurkis	39
Formulė 11 SVM Hyper plokštumos lygtis	41
Formulė 12 SVM atributų lygtis, tolygios hyper plokštumai.....	41
Formulė 13 SVM atributų lygtis, virš hyper plokštumos.....	42
Formulė 14 SVM atributų lygtis, žemiau hyper plokštumos	42
Formulė 15 SVM atributų suderinimo funkcija	42
Formulė 16 SVM didžiausia kraštinė hyper plokštuma	42
Formulė 17 ANN Prognozavimo funkcija	51
Formulė 18 SVM: MAPE ir RSME apskaičiavimas.....	54

Forma patvirtinta Mykolo Romerio universiteto
Senato 2012 m. lapkričio 20 d. nutarimu Nr.1SN-10

PATVIRTINIMAS APIE ATLIKTO DARBO SAVARANKIŠKUMĄ

2020 – 12 – 07
Vilnius

Aš, Mykolo Romerio universiteto (toliau – Universitetas), Viešojo valdymo ir verslo fakulteto, Verslo ir ekonomikos instituto (*fakulteto / instituto, programos pavadinimas*)

Studentas (-ė) Valerija Miribian, (*vardas, pavardė*) patvirtinu, kad šis rašto darbas / bakalauro / magistro baigiamasis darbas „Dirbtiniu intelektu grįstų metodų taikymo akcijų kainų prognozei tyrimas“:

1. Yra atliktas savarankiškai ir sąžiningai;
2. Nebuvo pristatytas ir gintas kitoje mokslo įstaigoje Lietuvoje ar užsienyje;
3. Yra parašytas remiantis akademinio rašymo principais ir susipažinus su rašto darbų metodiniais nurodymais.

Man žinoma, kad už sąžiningos konkurencijos principo pažeidimą – plagijavimą studentas gali būti šalinamas iš Universiteto kaip už akademinės etikos pažeidimą.