



**Vilniaus
universitetas**

Ataskaitinė informatikos krypties doktorantų konferencija 2024-10-04

Aivaras Bielskis (VU DMSTI doktorantas, Išmaniųjų technologijų tyrimų grupė)

Darbo tema.

Vertybinių popierių kainų prognozavimas ir investicinio portfelio formavimas naudojant mašininio mokymosi algoritmus.

Stock price forecasting and investment portfolio formation using machine learning algorithms.

Darbo vadovas.

prof. dr. Igoris Belovas.

Doktorantūros studijų laikotarpis.

2022 m. spalio mėn. 1 d. – 2026 m. rugsėjo mėn. 30 d.

Ataskaitinis laikotarpis.

2024 m. kovo mėn. 29 d. – 2024 m. spalio mėn. 3 d.

Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai	
	Planas	Įvykdyta
I (2022/2023)	2	3
II (2023/2024)	1	1
III (2024/2025)	1	
IV (2025/2026)		
Iš viso:	4	4

Visų studijų planas ir jo vykdymo suvestinė

Vilniaus
universitetas

Studijų metai	Dalyvavimas konferencijose				Publikacijos					
	Tarptautinėse		Nacionalinėse		Su citav. rodikliu			Be citav. rodiklio		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė	Planas	Įvykdyta	Būklė
I (2022/2023)									1	Publikuota ¹
II (2023/2024)		2 ^{2,3}								
III (2024/2025)	1				1					
IV (2025/2026)	1				1					
Iš viso	2	2	0	0	2	0	0	0	1	1

¹ Bielskis A., Belovas, I. *Akcijų kainų ARIMA ir LSTM prognozavimo metodų lyginamoji analizė*. Lietuvos matematikos rinkinys. Serija B, 2022, 63, p. 21-27.

DOI: 10.15388/LMR.2022.29755

Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Dalyvavimas konferencijose				Publikacijos					
	Tarptautinėse		Nacionalinėse		Su citav. rodikliu			Be citav. rodiklio		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė	Planas	Įvykdyta	Būklė
I (2022/2023)									1	Publikuota ¹
II (2023/2024)		2 ^{2,3}								
III (2024/2025)	1				1					
IV (2025/2026)	1				1					
Iš viso	2	2	0	0	2	0	0	0	1	1

² Bielskis A. *Advanced Machine Learning Techniques for Investment Forecasting: An Integrated Approach*. ECML PKDD 2024, Vilnius. Poster session.

³ Bielskis A. *Advanced Machine Learning Techniques for Investment Forecasting: An Integrated Approach*. ECML PKDD 2024, Vilnius. MIDAS: The 9th Workshop on Mining Data for financial applications. Paper presentation.

Ataskaitinio pusmečio darbų planas ir jų vykdymo suvestinė

Egzaminai			
Planas	Įvykdyta	Būklė	
Dalyvavimas konferencijose			
Planas	Įvykdyta	Konferencijos tipas	
	Bielskis A. Advanced Machine Learning Techniques for Investment Forecasting: An Integrated Approach. ECML PKDD 2024, Vilnius. Poster session.	Tarptautinė	
	Bielskis A. Advanced Machine Learning Techniques for Investment Forecasting: An Integrated Approach. ECML PKDD 2024, Vilnius. MIDAS: The 9th Workshop on Mining Data for financial applicationS. Paper presentation.	Tarptautinė	
Publikacijos			
Planas	Įvykdyta	Būklė	Publikacijos tipas

Visų mokslinių tyrimų ir disertacijos rengimo etapai

	Darbo pavadinimas	Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė: 1.1. Analitinės apžvalgos atlikimas. 1.2. Disertacijos tyrimo objekto detalizavimas.	2022 m. spalio mėn. – 2023 m. kovo mėn.	Įvykdyta
	1.3. Analitinės apžvalgos užbaigimas. 1.4. Mokslinių problemų susietų su tyrimo objektu identifikavimas ir tyrimo tikslo suformavimas	2023 m. balandžio mėn. – 2023 m. rugsėjo mėn.	Įvykdyta
2.	Mokslinio tyrimo vykdymas:		
	2.1. Tyrimo metodikos sudarymas: 2.1.1. Uždavinių, skirtų tyrimo tikslui pasiekti, suformulavimas. 2.1.2. Tyrimo metodikos išsikeltiems uždaviniams spręsti parinkimas. 2.1.3. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.	2023 m. spalio mėn. – 2024 m. kovo mėn.	Įvykdyta
	2.2. Teorinis tyrimas: 2.2.1. Mašininio mokymo metodų, planuojamų panaudoti tyrime parinkimas ir analizė. 2.2.2. Parinktų metodų adaptavimas ir modifikavimas, atsižvelgiant į problematikos specifiką.	2024 m. balandžio mėn. – 2024 m. rugsėjo mėn.	Įvykdyta

Visų mokslinių tyrimų ir disertacijos rengimo etapai

	2.3. Empirinis tyrimas: 2.3.1. Parinktų mašininio mokymo metodų įdiegimas, algoritmų konstravimas. 2.3.2. Įdiegtų algoritmų bei jų modifikacijų efektyvumo palyginamasis tyrimas.	2024 m. spalio mėn. – 2025 m. kovo mėn.	
	2.3.3. Geriausių algoritmų adaptavimas ir jų taikymas sprendžiant konkrečius praktinius uždavinius.	2025 m. balandžio mėn. – 2025 m. rugsėjo mėn.	
	2.4. Gautų rezultatų analizė ir apibendrinimas	2025 m. spalio mėn. - 2026 m. kovo mėn.	
1.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų ir kt.) parengimas: 3.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas. 3.2. Analitinės disertacijos dalies parengimas. 3.3. Teorinės disertacijos dalies parengimas. 3.4. Eksperimentinės disertacijos dalies parengimas. 3.5. Bendrųjų išvadų formulavimas.	2026 m. balandžio mėn. – 2026 m. rugsėjo mėn.	
2.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2026 m. birželio mėn.	
3.	Daktaro disertacijos gynimas	2026 m. rugsėjo mėn.	

Tyrimo tema

- Akcijų kainų prognozavimas yra plačiai nagrinėjama tema tiek pasaulyje, tiek Lietuvoje. Egzistuojantys metodai yra nuolat tobulinami, yra ieškoma naujų kelių, vykdomi aktualių įvykių ir procesų tyrimai.
- Doktorantūroje pasirinkta tema „Vertybinių popierių kainų prognozavimas ir investicinio portfelio formavimas naudojant mašininio mokymosi algoritmus“ yra mano bakalauro baigiamojo darbo – „Investicinio portfelio optimizavimas“ ir magistro baigiamojo darbo – „Prognozavimo metodų analizė“ tęsinys.
- Pasirinkta tema yra svarbi, nes darbo rezultatai gali padėti investuotojams ir finansų analitikams, siekiantiems efektyviai valdyti investicinius portfelius ir priimti optimalius sprendimus.
- Mašininio mokymosi algoritmai yra aktualus įrankis prognozuojant vertybinių popierių kainų svyravimus sudarant investicinius portfelius.

Tyrimo objektas

- Tyrimo objektas yra viešai kotiruojamų įmonių akcijos, biržoje prekiaujami fondai, prekės (angl. commodities), obligacijos, valiutos, išvestinės priemonės ir kiti produktai, kuriais galima prekiauti rinkose.
- Tyrimu siekiama iširti, kaip naudojant mašininio mokymosi algoritmus galima prognozuoti vertybinių popierių kainų kitimus ir kurti investicinius portfelius, kurie būtų optimaliai sudaryti atsižvelgiant į kainų kitimo tendencijas.
- Darbo metu planuojama tyrinėti tiek Lietuvos, tiek pasaulio vertybinius popierius.

Tyrimo tikslas

- Tyrimo tikslas yra pasiūlyti naujus (arba modifikuoti esamus) patobulintus algoritmus ir jų pagrindu sukurti programą, formuojančią optimalų investicinį portfelį.
- Programa turėtų naudojant mašininio mokymosi algoritmus efektyviai prognozuoti vertybinių popierių kainų (arba finansinių indeksų) svyravimus ir konstruoti portfelius, siekiant maksimizuoti pelną ir minimalizuoti riziką.

Mokslinės problemos

1. Didelių finansinių duomenų surinkimo ir filtravimo (parengimo) proceso automatizavimas
2. Egzistuojančių mašininio mokymosi algoritmų vertybinių popierių kainų prognozavimui bei investicinio portfelio valdymui modifikavimas (tobulinimas):
 - Prognozės tikslumo problema;
 - Rizikos valdymo problema;
 - Strategijos parinkimo problema;
 - Stabilumo problema;
 - Palyginimo ir verifikacijos problemos;
3. Nefinansinių veiksnių (pvz., globalių reiškinių: pandemijų, soc. krizių) poveikis investicinio portfelio valdymo procesui.

Tyrimo uždaviniai

1. Surinkti ir apdoroti finansinius duomenis, būtinus investicinių priemonių kainų prognozavimui ir portfelio formavimui. Išnagrinėti investicinių priemonių rinką, kad būtų galima įvertinti naujas investicines galimybes, kurias galima naudoti portfelio diversifikavimui ir rizikos minimizavimui.
2. Iširti esančius mašininio mokymosi algoritmus, skirtus vertybinių popierių kainų prognozavimui ir investicinio portfelio formavimui. Iširti esančius optimalaus investicinio portfelio formavimo metodus (algoritmus).
3. Palyginti mašininio mokymosi modelių prognozių tikslumą ir investicinių portfelių valdymo metodus, pasitelkiant istorinius bei sintetinius duomenis.

Tyrimo uždaviniai

4. Pasiūlyti ir pritaikyti naujus mašininio mokymosi algoritmus, kurie geriau prognozuotų vertybinių popierių kainų kitimą ir leistų geriau optimizuoti investicinį portfelį. Pasiūlyti naujas strategijas, kaip naudoti mašininio mokymosi algoritmus investicinio portfelio valdyme ir kaip padidinti grąžą bei sumažinti riziką. Kombinuoti tradicinius finansų modelius su mašininio mokymosi algoritmais, kad būtų sukurti nauji investiciniai modeliai. Integruoti naujus finansų rinkos duomenų šaltinius, pvz., susijusius su klimato kaita arba pandemijos poveikiu, į investicinio portfelio prognozavimo ir formavimo modelius. Atsižvelgti į etinius ir socialinius klausimus priimant investicijų sprendimus, pvz., remiantis įmonių socialine atsakomybe arba etiškais investavimo principais. Pasirinktinai pritaikyti socialinių tinklų duomenis investicijų priimti sprendimams ir investicijų portfelio formavimui.
5. Palyginti sukurtų mašininio mokymosi modelių ir investicinių strategijų veiksmingumą su kitų autorių sukurtomis strategijomis ir modeliais.

Tyrimo metodika

- Skaidyti tyrimą į dvi dalis: prognozavimo užduotis ir portfelio formavimo užduotis.
- Prognozavimo uždaviniui atlikti - jungti visus long short term memory (LSTM), Hilbert–Huang Transform (HHT) ir extreme gradient boost (XGBoost) bei deep multilayer perceptron (DMLP) metodus, arba šių metodų mažesnį mišinį ir pridėti stochastic dominance (SD) filtrą.
- Prognozavimo uždaviniui atlikti – patobulinti, arba apjungti su kitais metodais Random Forest (RF) metoda.
- Portfelio formavimo uždaviniui atlikti naudoti ir palyginti Markowitz mean-variance (MV), omega portfolio optimization models (Omega-CVaR), diversified mean-variance with forecasting (DMVF) metodus.
- Prie tyrimo pridėti maximum-drawdown (MDD) metrika ir iš jos išplaukiančią calmar ratio (CR) metrikas vertinant rezultatus.

Teorinio ir empirinio tyrimo planas

1. Prognozavimo uždaviniui naudojamų metodų (LSTM, HHT, XGBoost, DMLP ir SD filtras) įgyvendinimas ir šių metodų jungimas.
2. Prognozių atlikimas naudojant įgyvendintus metodus. Atliktų prognozių rezultatų lyginimas.
3. Trūkumų šalinimas ir esant poreikiui metodų modifikavimas.
4. Gautų rezultatų analizė.
5. Portfelio formavimo uždaviniui naudojamų metodų (MV, Omega-CVaR, CMVF) įgyvendinimas.
6. Prognozavimo ir portfelio formavimo sistemos sudarymas.
7. Sistemos testavimas, trūkumų šalinimas ir esant poreikiui metodų modifikavimas.
8. Galutinių rezultatų analizė.

1. Mašininio mokymo metodų, planuojamų panaudoti tyrime parinkimas ir analizė. Parinktų metodų adaptavimas ir modifikavimas, atsižvelgiant į problematikos specifiką.
2. Parinktų metodų adaptavimas ir modifikavimas buvo atliktas remiantis penkių plačiai naudojamų investicinių indeksų duomenimis, gautais iš Yahoo Finance. Įdiegta paprasta ilgos ir trumpos pozicijos strategija, o tikslumas buvo vertinamas naudojant tokius rodiklius kaip MSE, MAE ir R-squared.
3. Mokyta trijų modelių (LSTM, DMLP, XGBoost) su penkių indeksų istorinių kainų duomenimis.
4. HHT taikymas nepadidino prognozių tikslumo, o LSTM ir DMLP kombinacija davė geriausius rezultatus, ypač DAX indekse.
5. XGBoost susidūrė su sunkumais prognozuojant už treniravimo diapazono ribų.

Gautos adaptacijos išvados:

1. HHT efektyvumas: Hilberto-Huango transformacijos taikymas reikšmingai nepadidino prognozių tikslumo.
2. Indeksų pasirinkimas: Modelių veikimas skyrėsi priklausomai nuo indeksų, parodant, kad jų efektyvumas priklauso nuo konkretaus indekso.
3. XGBoost apribojimai: XGBoost turėjo sunkumų prognozuodamas už treniravimo duomenų ribų, pabrėžiant ekstrapoliacijos ribotumus.
4. LSTM pranašumas: LSTM šiek tiek lenkė kitus modelius.
5. Rodikliai ir finansiniai rezultatai: Geriausi MAE, MSE arba R kvadratu rodikliai ne visada lemia optimalius finansinius rezultatus.

Bendruosius gebėjimus stiprinančios veiklos

Vilniaus
universitetas

Dalyko pavadinimas	Data	Programos trukmė val.	Programos apimtis kreditais ECTS	Trumpas turinio aprašas
Lietuvos akademinė elektroninė biblioteka (eLABa): autoriaus sąsaja	2022-11-08	1,5	0,15	Paskaitos tikslas – supažindinti su Lietuvos akademinės elektroninės bibliotekos sistemos funkcijomis, struktūra, galimybėmis, eLABa autoriaus sąsaja, dokumentų registravimu ir viso teksto (failų) įkėlimu, įrašų importu, licencinių sutarčių pasirašymu ir eLABa ataskaitų formavimu.
LATEX	2022-11-08 2022-11-15	8	1,25	Šiame dalyke studentai supažindinti su teksto redagavimo programa LaTeX. 1. Instaliavimas ir paruošimas darbui; 2. Preambulė; 3. Teksto redagavimas; 4. Formulės ir specialūs simboliai; 5. Paveikslėliai ir lentelės; 6. Nuorodos ir bibliografija; 7. Aplinkos; 8. Pateiktys.
Mokslinių tyrimų duomenų valdymo įvadas	2023-11-08	1,5	0,15	Mokymų tikslas – supažindinti su mokslinių tyrimų duomenų ir jų valdymo sąvokomis bei mokslinių tyrimų duomenų gyvavimo ciklu. Juose bus aptariami svarbiausi sudėtiniai mokslinių tyrimų duomenų valdymo aspektai ir pateikiama praktiškų darbinių duomenų priežiūros bei organizavimo patarimų.
Mokslinių tyrimų duomenų valdymo plano sudarymas	2023-11-09	1	0,1	Paskaitos tikslas - supažindinti su duomenų valdymo plano kaip kartu su projekto paraiška teikiamo dokumento specifika. Paskaitoje bus aptariama, į ką vertėtų atkreipti dėmesį rengiant duomenų valdymo planą ir bus pristatyti duomenų valdymo plano rengimą palengvinantys įrankiai (DVP Vedlys, DMP Online ir kt.).
FAIR duomenų principai	2023-11-09	1	0,1	Mokymų tikslas – supažindinti su mokslinių tyrimų duomenų pateikimo rekomendacijomis, plačiau žinomomis kaip „FAIR principai“. Šių principų laikymasis padidina mokslinio tyrimo metu sukurtų duomenų pridėtinę vertę, be to šių principų laikytis reikalauja Europos Sąjungos mokslo finansavimo programos ir kai kurie akademiniai leidėjai. Mokymuose sužinosite, kokias duomenų savybes įvardija kiekviena akronimo FAIR raidė ir ką daryti, kad šias savybes turėtų Jūsų sukurti duomenys. Formatas: paskaita (1 val.) ir savarankiško darbo užduotis

Bendruosius gebėjimus stiprinančios veiklos

Duomenų talpyklos ir duomenų žurnalai	2023-11-10	1	0,1	Paskaitos tikslas - supažindinti su mokslinių tyrimų duomenų saugojimui ir publikavimui skirtomis platformomis - duomenų talpyklomis. Bus aptariami pagrindiniai duomenų talpyklų tipai, pristatomos populiariausios bendro pobūdžio talpyklos, patariama, kaip pasirinkti tinkamą talpyklą savo duomenims. Taip kalbėsime, kas yra duomenų žurnalai ir kuo jie skiriasi nuo įprastų mokslinių žurnalų.
MIDAS naudojimas	2023-11-10	1	0,1	Paskaitoje bus supažindinama su Nacionaliniu mokslinių tyrimų duomenų archyvu (MIDAS) bei jo funkcijomis: publikuotų duomenų peržiūra, duomenų įkėlimu ir publikavimu, bendradarbiavimą palengvinančiomis funkcijomis, naudotojo paskyros tvarkymu.
Informacijos saugojimas ir citavimas naudojantis programa Zotero	2023-11-14	1,5	0,15	Trumpas programos Zotero pristatymas. Praktinis darbas su programa Zotero: įdiegimas ir pasiruošimas darbui; Mokslinių informacijos šaltinių išsaugojimo galimybės; Citavimo nuorodų ir literatūros sąrašo sudarymas; Papildomos programos funkcijos.
Pedagoginė veikla			1	2022-2023 mokslo metų pirmojo pusmečio metu dėčiau „Informacinės saugos pagrindai“ pratybas informacinių sistemų inžinerijos ketvirtakursiams. 2023-2024 mokslo metų pirmojo pusmečio metu dėčiau „Finansų inžinerija ir modeliavimas“ pratybas informacinių sistemų inžinerijos trečiakursiams. 2022-2023 mokslo metų antrojo pusmečio metu Recenzavau du Programų sistemų programos bakalauro baigiamuosius darbus „Giliųjų neuroninių tinklų taikymas kriptovaliutų kainų ir tendencijų prognozavimui“ ir „Nordpool elektros kainų prognozavimas naudojant dirbtinio intelekto algoritmus“; 2023-2024 mokslo metų pirmojo pusmečio metu recenzavau Informacinių technologijų programos bakalauro baigiamąjį darbą „Decentralizuota vertybinių popierių valdymo sistema paremta blokų grandinių technologija“.

Kito pusmečio darbo planas

1. Parinktų mašininio mokymo metodų įdiegimas, algoritmų konstravimas.
2. Įdiegtų algoritmų bei jų modifikacijų efektyvumo palyginamasis tyrimas.
3. Sudalyvauti bendruosius gebėjimus stiprinančiose veiklose (pagal galimybę).



**Vilniaus
universitetas**

Ačiū už dėmesį

Aivaras Bielskis

VU DMSTI doktorantas

aivaras.bielskis@mif.stud.vu.lt