

Vilniaus universiteto priėmimo rodiklių analizė ir prognozavimas (XXI a., I ketvirtis)

Martynas Jokubaitis¹, Igoris Belovas²

¹ Vilniaus universitetas, Šiaulių akademija,
Vytauto g. 84, LT-76352, Šiauliai
martynas.jokubaitis@sa.stud.vu.lt

² Vilniaus universitetas, Matematikos ir Informatikos fakultetas,
Duomenų mokslo ir skaitmeninių technologijų institutas,
Akademijos g. 4, LT-08412, Vilnius
igoris.belovas@mif.vu.lt

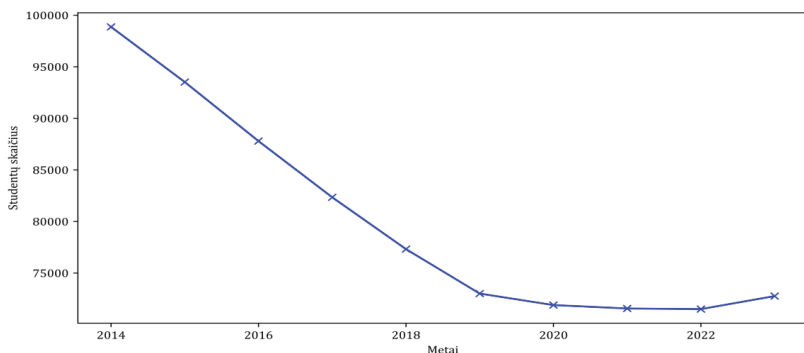
Santrauka. Aukštasis mokslas yra vienas iš svarbesnių veiksnių, teikiančių ne tik asmeninę naudą, bet ir prisidedančių prie geresnės, sveikesnės visuomenės kūrimo. Lietuvoje aukštojo mokslo prieinamumui išaugus, šalis tapo viena labiausiai išsilavinusių visoje Europos Sąjungoje. Ženklus universitetų studentų skaičius mažėjimas 2021–2022 metais stabilizavosi ir pradėjo rodyti teigiamą dinamiką. Vilniaus universitete, kurį pasirenka didžiausia stojančiųjų į universitetus dalis (2023–2024 metais ji sudarė 42–43 %), nuo 2023 metų priimtųjų studijuoti skaičius tik augo ir 2024 metais Vilniaus universitete studijas pradėjo didžiausias per 2019–2024 metų laikotarpį pirmakursių skaičius. Darbe nagrinėjami Vilniaus universiteto XXI amžiaus pirmo ketvirčio priėmimo rodikliai. Galimų ateities tendencijų įvertinimui bei analizei yra taikomas ARIMA modelis. Prognozės rezultatai rodo, kad stojančiųjų rodiklis išliks stabilus. Taip pat darbe nustatoma, jog ARIMA modelis tiksliau įvertina stojančiųjų skaičių už anksčiau naudotą AR modelį.

Raktiniai žodžiai: ARIMA, RMSE, MAE, prognozavimas, studentų skaičius, aukštasis mokslas, priėmimo rodikliai.

1 Įvadas

Smarkiai iki 2021–2022 metų kritęs Lietuvos universitetų studentų skaičius paskutiniaisiais metais ne tik stabilizavosi, bet ir pradėjo rodyti augimo tendencijas [1] (žr. 1 pav.). Tokioje netikėtai besikeičiančioje aplinkoje yra svarbu gebėti prisitaikyti prie naujų aplinkybių. Studentų skaičiaus prognozavimo galimybes nagrinėjančiuose moksliniuose darbuose pabrėžiama, kad pro-

gnozės padeda ne tik numatyti stojimų tendencijas, bet ir priimti duomenimis pagrįstus strateginius sprendimus dėl resursų paskirstymo, gerinti tiek mokymo, tiek mokymosi procesus.



1 pav. Lietuvos universitetų studentų skaičiaus dinamika

Lavilles ir Arcilla [2], tirdami Mindanao universiteto stojančiųjų rodiklius, siekiant pagerinti universiteto gebėjimą tiksliau prognozuoti priėmimą ir tobulinti kursų pasiūlą, nustatė, kad laiko eilučių metodai, tokie kaip slenkančiųjų vidurkių, viengubo ir dvigubo eksponentinio glodinimo, pasižymėjo didesniu tikslumu nei anksčiau universiteto naudotas *naive* metodas, kai sekančio žingsnio prognozė yra praėjusio žingsnio reikšmė. Bousnguar ir kt. [3], nagrinėdami priėmimo statistiką, taikė ARIMA, eksponentinio glodinimo, neryškių laiko eilučių (*angl.* Fuzzy Time Series, FTS) ir neuroninių tinklų ilgos trumpalaikės atminties (*angl.* Long Short-Term Memory, LSTM) modelius. Savo darbe jiems pavyko nustatyti, kad esant turimam 18 metų duomenų rinkiniui, tiksliausius rezultatus pateikė FTS metodas. Išsamus stojančiųjų skaičiaus vertinimo tyrimas, siekiantis užtikrinti tvarų edukacinės sistemos vystymą, buvo atliktas Kinijoje [4]. Prognozavimui buvo naudota polinominė regresija, dispersinė analizė ir Holto eksponentinis glodinimas, kuris pasiekė tikslesnius rezultatus.

Lietuvoje studentų skaičiaus mažėjimo problemą nagrinėjo Šimkevičienė, kuri savo magistro baigiamajame darbe atliko Lietuvos didžiųjų miestų aukštųjų mokyklų priėmimo rezultatų lyginamąją statistinę analizę 2008–2021 metų laikotarpiu ir nustatė studentų skaičiaus mažėjimo tendencijas. Verta pažymėti, kad autorės tyrimas apsiribojo duomenimis iki 2021 m., o naujausi duomenys rodo bendrą Lietuvos universitetų studentų skaičiaus

augimo tendenciją. Savo tyrimuose autorė studentų skaičiaus laiko eilutės analizuoti taikė autoregresinį vieno parametro $AR(p)$ modelį [5]. Šio darbo tikslas – žengti sekantį žingsnį ir pritaikyti autoregresinį slenkančiųjų vidurkių ARIMA modelį Vilniaus universiteto priėmimo rodikliams.

2 Duomenys ir metodas

Darbe naudojami Vilniaus universiteto priimtų studentų ir numatytų vietų skaičiaus duomenys 2000-2024 metų laikotarpiu. Apžvelgus atliktus tyrimus [2,3,4,5], reikia pastebėti, kad aukščiau minėtų darbų autoriai neatsižvelgė į nustatyto plano dydį ir vertino tik bendrus priimtų studentų skaičius. Šiame darbe studentų skaičiaus įvertinimui yra naudojamas priimtų studentų ir numatytų vietų santykis, taip atsižvelgiant ir į Universiteto išsikeltus metinius priėmimo tikslus.

Tyrimė yra taikomas ARIMA (angl. *Autoregressive Integrated Moving Average*) metodas, žymimas kaip $ARIMA(p,d,c)$, kur p – autoregresinio proceso eilė, d – stacionarių duomenų eilė, c – slankiojo vidurkio proceso eilė [3]. Modelio tikslumui įvertinti naudojami vidutinės absoliučios paklaidos MAE (angl. *Mean Absolute Error*) ir šaknies vidurkio kvadratinės paklaidos RMSE (angl. *Root Mean Square Error*) rodikliai. $ARIMA(p, d, c)$ modelio parametrai yra apskaičiuojami tinklenio paieškos (angl. *grid search*) metodu, kai p , d , c reikšmės kinta nuo 0 iki 2 ir minėtomis paklaidomis remiantis parenkamas geriausias rezultatus rodantis modelis [3].

ARIMA modelis apskaičiuotas naudojant 20 metų laikotarpį (nuo 2000 iki 2019 metų). Sekantiems 5 metams (nuo 2020 iki 2024 metų) modelis yra testuojamas ir vertinamos paklaidos; toliau yra pateikiamos prognozės 5 metams į priekį.

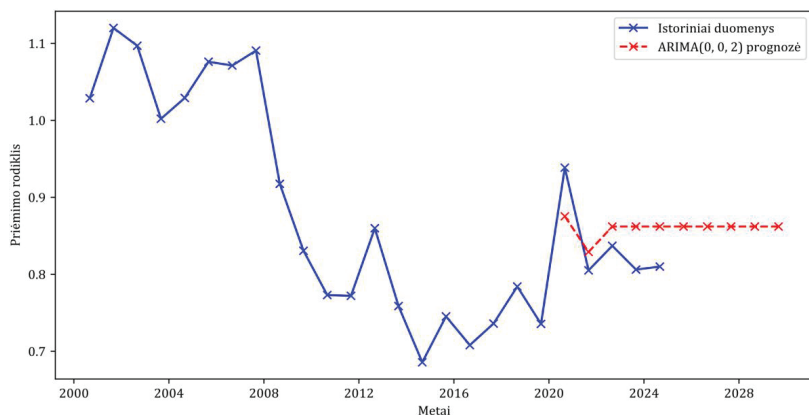
3 Rezultatai

Analizuojant MAE ir RMSE paklaidas, pateiktas 1 lentelėje, matoma, kad ARIMA modelis, kurio autoregresinio proceso p ir stacionarių duomenų d eilės lygios 0 ir slankiojo vidurkio proceso eilė c lygi 2, $ARIMA(0, 0, 2)$, pasiekia geriausias rezultatus iš visų naudotų modelių. Ties šiais parametrais tiek vidutinės absoliučios paklaidos MAE, tiek ir šaknies vidurkio kvadratinės paklaidos RMSE reikšmės yra mažiausios.

1 lentelė. Skirtingų parametų ARIMA modelių paklaidų rezultatai.

ARIMA(p,d,c)	MAE	RMSE
ARIMA(0, 0, 0)	0,066	0,068
ARIMA(0, 0, 1)	0,074	0,081
ARIMA(0, 0, 2)	0,044	0,047
ARIMA(0, 1, 0)	0,104	0,116
ARIMA(0, 1, 1)	0,110	0,121
ARIMA(0, 1, 2)	0,106	0,120
ARIMA(0, 2, 0)	0,249	0,254
ARIMA(0, 2, 1)	0,150	0,155
ARIMA(0, 2, 2)	0,173	0,177
ARIMA(1, 0, 0)	0,102	0,114
ARIMA(1, 0, 1)	0,103	0,115
ARIMA(1, 0, 2)	0,155	0,161
ARIMA(1, 1, 0)	0,107	0,118
ARIMA(1, 1, 1)	0,106	0,117
ARIMA(1, 1, 2)	0,098	0,111
ARIMA(1, 2, 0)	0,174	0,178
ARIMA(1, 2, 1)	0,150	0,155
ARIMA(1, 2, 2)	0,156	0,160
ARIMA(2, 0, 0)	0,101	0,114
ARIMA(2, 0, 1)	0,093	0,109
ARIMA(2, 0, 2)	0,151	0,157
ARIMA(2, 1, 0)	0,110	0,123
ARIMA(2, 1, 1)	0,100	0,115
ARIMA(2, 1, 2)	0,102	0,114
ARIMA(2, 2, 0)	0,111	0,122
ARIMA(2, 2, 1)	0,176	0,181
ARIMA(2, 2, 2)	0,166	0,171

2 pav. pateiktas grafikas, rodantis istorinius analizuoto rodiklio duomenis bei naudoto modelio reikšmes. Iš grafiko matyti, kad modelis gana tiksliai nuspėja 2020, 2021 ir 2022 metų rodiklio pokyčius testuoto laikotarpio metu. Ir pateikia prognozes 5 metams į priekį, kur prognozuojamas rodiklis nemažėja. Tai atitinka anksčiau aptartą studentų skaičiaus mažėjimo tendencijų stabilizavimąsi.



2 pav. Istoriniai duomenys ir ARIMA(0, 0, 2) prognozė

2 lentelėje pateikti nuo 2020 iki 2024 metų testavimo periodo istoriniai duomenys bei modelio prognozė, kartu su prognozuojamomis reikšmėmis sekantiems 5 metams. Apibendrinant, galime teigti, kad yra nustatyta, jog Vilniaus universiteto studentų priėmimo rodiklį tiksliausiai nusako ARIMA(0, 0, 2) parametų modelis.

2 lentelė. ARIMA(0, 0, 2) modelio prognozė

Data	Istoriniai duomenys	ARIMA(0, 0 2) prognozė
9/1/2020	0,939	0,875
9/1/2021	0,805	0,829
9/1/2022	0,837	0,862
9/1/2023	0,806	0,862
9/1/2024	0,810	0,862
9/1/2025		0,862
9/1/2026		0,862
9/1/2027		0,862
9/1/2028		0,862
9/1/2029		0,862

4 Išvados ir aptarimas

Šimkevičienė savo tyrime naudojo vieno parametro $AR(p)$ modelius bei nustatė, kad studentų skaičių geriausiai aprašo $AR(3)$ modelis. Šiame darbe pritaikius $AR(p)$ modelius, kai p parametras kinta nuo 1 iki 3, analizuojamiems duomenims, jų prognozių paklaidos pateiktos 3 lentelėje. Kaip matyti iš lentelės, visi šie modeliai stipriai atsilieka nuo šiame darbe nustatyto $ARIMA(0, 0, 2)$ modelio. Jo prognozių paklaidos yra daugiau nei du kartus mažesnės nei bet kurio pritaikyto $AR(p)$ modelio, todėl jį galima laikyti geresniu pasirinkimu.

3 lentelė. $AR(p)$ modelių paklaidų rezultatai

Modelis	MAE	RMSE
AR(1)	0,102	0,114
AR(2)	0,101	0,114
AR(3)	0,121	0,132
ARIMA(0, 0, 2)	0,044	0,047

Nors darbe yra akcentuojamas $ARIMA$ modelis (kaip sekantis žingsnis po anksčiau naudoto AR modelio), tęsiant tyrimus ateityje, planuojama įtraukti į palyginimą kitus užsienio autorių naudotus metodus: eksponentinio glodinimo, neryškių laiko eilučių (angl. *Fuzzy Time Series*, *FTS*) ir neuroninių tinklų ilgos trumpalaikės atminties (angl. *Long Short-Term Memory*, *LSTM*) modelius.

Literatūra

[1] Oficialiosios statistikos portalas. URL: <https://osp.stat.gov.lt/statistiniu-rodikliu-analize?hash=4deccb8e-1383-46e9-a86f-ecee7485f558#/>

[2] Lavilles, R. Q., Arcilla M. J. B. (2012). Enrollment Forecasting for School Management System. *International Journal of Modeling and Optimization*, 2(5), 563-566. <https://doi.org/10.7763/IJMO.2012.V2.183>

[3] Bousnguar, H., Najdi, L., Battou, A. (2022). Forecasting approaches in a higher education setting. *Education and Information Technologies*, 27, 1993-2011. <https://doi.org/10.1007/s10639-021-10684-z>

[4] Li, X. (2023). Sequence Model and Prediction for Sustainable Enrollments in Chinese Universities. *Sustainability*, 15, 214. <https://doi.org/10.3390/su15010214>

[5] Šimkevičienė, I., (2023). Studentų priėmimo į Lietuvos didžiųjų miestų aukštąsias mokyklas lyginamoji statistinė analizė. *Magistro baigiamasis darbas*. URL: <https://talpykla.elaba.lt/elaba-fedora/objects/elaba:192828115/datastreams/MAIN/content>