

# Studentų akademinės sėkmės nuotolinėse studijose prognozavimo galimybės taikant duomenų tyrybą

Artūras Kukštys

Ny-Krohnborg skole, Rogagaten 9, 5020 Bergen, Norway  
*a.kukstys@gmail.com*

---

**Santrauka.** Aukštosios mokyklos veikia konkurencingoje aplinkoje, kur studentų akademinė sėkmė yra svarbus studijų kokybės rodiklis. Nuotolinės studijos populiarėja, tačiau jos dažniau nutraukiamos nei tradicinės. Šio darbo tikslas – taikant duomenų tyrybą parengti akademinės sėkmės prognozavimo modelį, skirtą mažinti studentų nubyryjimą. Parengtas modelis, paremtas besimokančiųjų mokymosi duomenimis ir atsitiktinio miško algoritmu, leidžia anksti identifikuoti rizikos grupės studentus.

**Raktiniai žodžiai:** akademinė sėkmė, prognozavimas, duomenų tyryba, duomenų tyrybos metodai ir algoritmai, CRISP-DM.

---

## 1 Įvadas

Akademinė sėkmė yra vienas svarbiausių švietimo institucijų tikslų, užtikrinantis visų besimokančiųjų galimybes sėkmingai įgyti išsilavinimą. Pastarojo dešimtmečio tyrimai rodo, kad suaugusieji, nors vis dažniau renkasi nuotolines studijas, jas taip pat dažniau nutraukia nei tradicines [1, 2]. Taigi akademinės sėkmės ir studentų nubyryjimo problema nuotolinėse studijose išlieka aktuali.

Studentų studijų nutraukimas turi neigiamą poveikį individualiu, instituciniu ir nacionaliniu lygmeniu. Jis mažina studentų pasitenkinimą studijomis, silpnina savivertę bei didina socialinės ir profesinės atskirties riziką [3]. Tuo tarpu aukštosioms mokykloms studentų kaitos nevaldymas apsunkina studijų proceso organizavimą, o prasti rezultatai gali neigiamai paveikti institucijos reputaciją. Be to, valstybės skiriami finansiniai ištekliai švietimui tokiu atveju naudojami neoptimaliai [4].

Aukštosios mokyklos veikia konkurencingoje aplinkoje, kurioje studentų akademinė sėkmė tampa svarbiu studijų kokybės ir institucinio efektyvumo rodikliu. Literatūroje išskiriami keli akademinės sėkmės prognozavimo bū-

dai: tiesioginis bendravimas su studentais, apklausos bei duomenų tyryba [1]. Pastaroji technologija vis plačiau taikoma aukštosiose mokyklose studentų nubyrejimui mažinti [1, 2, 5].

Šio darbo tikslas – taikant duomenų tyrybą parengti akademinės sėkmės prognozavimo modelį, skirtą mažinti studentų nubyrejimą.

## 2 Duomenys ir metodai

Prognozuojant besimokančiųjų akademinę sėkmę taikant duomenų tyrybą rekomenduojama remtis CRISP-DM (angl. *Cross Industry Standard Process for Data Mining*) modeliu [6]. Šis modelis sudaro struktūruotą duomenų tyrybos procesą, apimantį šešis etapus [7].

### 2.1 Veiklos supratimas.

Šiame etape siekiant įvertinti duomenų tyrybos taikymo galimybes prognozuoti studentų akademinę sėkmę Kauno technologijos universiteto antrosios pakopos nuotolinėse studijose „Nuotolinio mokymosi informacinės technologijos“, atlikta SSGG analizė, kuri parodė, kad, nepaisant pasireiškiančios akademinės nesėkmės, egzistuoja palankios sąlygos taikyti duomenų tyrybą.

### 2.2 Duomenų supratimas

Prognozuojant studentų akademinę sėkmę analizuojami įvairūs švietimo duomenys, tokie kaip besimokančiųjų veiksmai virtualiojoje mokymosi aplinkoje, akademiniai rezultatai, demografinės charakteristikos ir pan.

Šiame darbe buvo pasirinkti besimokančiųjų mokymosi proceso duomenys, kurie generuojami studijų programos virtualiojoje mokymosi aplinkoje Moodle.

### 2.3 Duomenų parengimas

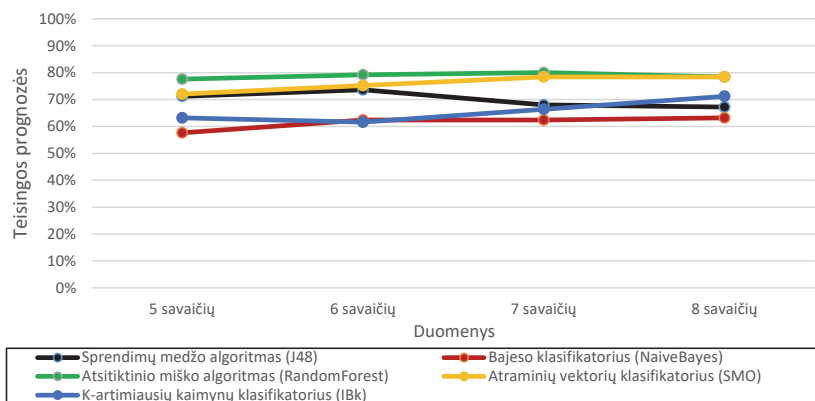
Duomenų paėmimui taikomos SQL užklauskos, kuriomis paimami duomenys iš virtualiosios mokymosi aplinkos Moodle duomenų bazės. Siekiant užtikrinti studentų duomenų apsaugą, jų privatumą ir konfidencialumą, atlikta duomenų pseudonimizacija. Šio etapo metu neapdoroti duomenys buvo filtruojami, pervadinami ir sujungiami į bendrą duomenų rinkinį, kuris vėliau suskirstytas į tris dalis: apmokymo ir dvi testavimo imtis. Duomenų rinkinio struktūra pateikta 1 lentelėje. Kiekviena savaitė modelyje traktuojama kaip atskiras požymis, leidžiantis analizuoti studentų mokymosi aktyvumo dinamiką laiko atžvilgiu.

**1 lentelė.** Akademinės sėkmės prognozavimo modelio apmokymui sudaryto duomenų rinkinio struktūra.

Kintamasis	Aprašymas
TP1_access_week1-week8	Studento prisijungimų prie modulio „Tiriamasis projektas 1“ skaičius kiekvieną analizuojamą savaitę (1–8 savaitės)
TP1_clicks_week1-week8	Studento paspaudimų skaičius modulyje „Tiriamasis projektas 1“ kiekvieną analizuojamą savaitę (1–8 savaitės)
VMP_access_week1-week8	Studento prisijungimų prie modulio „Virtualiojo mokymosi pagrindai“ skaičius kiekvieną analizuojamą savaitę (1–8 savaitės)
VMP_clicks_week1-week8	Studento paspaudimų skaičius modulyje „Virtualiojo mokymosi pagrindai“ kiekvieną analizuojamą savaitę (1–8 savaitės)
key	Raktas, studento identifikacijai pseudonimizuoti (n = 125)
success	Klasės kintamasis, kur N žymi akademinę nesėkmę, o T rodo akademinę sėkmę

## 2.4 Modeliavimas

Modeliavimas vykdomas Weka 3.8 programinėje sistemoje, kurioje buvo parinkti sprendimo medžio algoritmas („J48“), Bajeso klasifikatorius („Naive-Bayes“), atsitiktinio miško algoritmas („RandomForest“), atraminių vektorių klasifikatorius („SMO“) ir K-artimiausių kaimynų klasifikatorius („IBk“). Prognozavimas atliekamas keliais etapais atsižvelgiant į rezultatų pokyčius su 5–8 savaitių duomenimis (1 pav.).



**1 pav.** Pirminiame akademinės sėkmės prognozavimo modelyje taikomų algoritmų rezultatai.

Analizė parodė, kad atsitiktinio miško algoritmas pasiekė aukščiausią prognozavimo tikslumą visais nagrinėtais duomenų atvejais, lyginant su kitais algoritmais. Geriausi rezultatai pasiekti naudojant 7 savaitių duomenis, kai modelio tikslumas siekė 80 %.

## 2.5 Įvertinimas

Šiame etape įvertinami algoritmų kokybės rezultatai ir tarpusavyje palyginami šie rodikliai: tikslumas, atkūrimas, F-įvertis ir ROC kreivės įvertis (2 lentelė).

**2 lentelė.** Pirminiame akademinės sėkmės prognozavimo modelyje taikomų algoritmų rodikliai.

Algoritmas	Duomenys	Tikslumas	Atkūrimas	F-įvertis	ROC įvertis	Klasė
<b>Sprendimų medis (J48)</b>	5 savaitių	0,5	0,278	0,357	0,531	N
		0,752	0,888	0,814	0,531	T
	6 savaitių	0,556	0,417	0,476	0,676	N
		0,786	0,865	0,824	0,676	T
	7 savaitių	0,441	0,417	0,429	0,592	N
		0,769	0,787	0,778	0,592	T
	8 savaitių	0,429	0,417	0,423	0,598	N
		0,767	0,775	0,771	0,598	T
<b>Bajeso klasifikatorius (Naive-Bayes)</b>	5 savaitių	0,361	0,611	0,454	0,646	N
		0,781	0,562	0,654	0,646	T
	6 savaitių	0,418	0,778	0,544	0,716	N
		0,862	0,562	0,68	0,716	T
	7 savaitių	0,41	0,694	0,515	0,737	N
		0,828	0,596	0,693	0,737	T
	8 savaitių	0,414	0,667	0,511	0,735	N
		0,821	0,618	0,705	0,733	T
<b>Atsitiktinio miško (RandomForest)</b>	5 savaitių	0,9	0,25	0,391	0,648	N
		0,765	0,989	0,863	0,648	T
	6 savaitių	0,857	0,333	0,48	0,734	N
		0,784	0,978	0,87	0,734	T
	7 savaitių	0,824	0,389	0,528	0,772	N
		0,796	0,966	0,873	0,772	T
	8 savaitių	0,765	0,361	0,491	0,716	N
		0,787	0,955	0,863	0,716	T

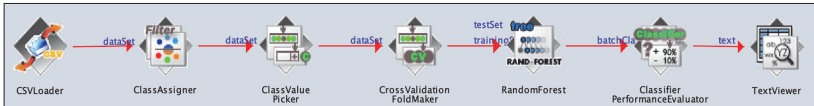
Algoritmas	Duomenys	Tikslumas	Atkūrimas	F-įvertis	ROC įvertis	Klasė
Atraminių vektorių klasifikatorius (SMO)	5 savaičių	0,556	0,139	0,222	0,547	N
		0,733	0,955	0,829	0,547	T
	6 savaičių	0,692	0,25	0,367	0,603	N
		0,759	0,955	0,846	0,603	T
	7 savaičių	0,737	0,389	0,509	0,666	N
		0,792	0,944	0,862	0,666	T
	8 savaičių	0,737	0,389	0,509	0,666	N
		0,792	0,944	0,862	0,666	T
K-artimiausių kaimynų klasifikatorius (IBk)	5 savaičių	0,381	0,444	0,41	0,561	N
		0,759	0,708	0,733	0,561	T
	6 savaičių	0,357	0,417	0,385	0,539	N
		0,747	0,697	0,721	0,539	T
	7 savaičių	0,425	0,472	0,447	0,613	N
		0,776	0,742	0,759	0,613	T
	8 savaičių	0,5	0,5	0,5	0,655	N
		0,798	0,798	0,798	0,655	T

Palyginus algoritmų rezultatus (naudojant 7 savaičių duomenis), nustatyta, kad atsitiktinio miško algoritmas yra pranašesnis prognozuojant akademinę sėkmę (klasę „T“) nei atraminių vektorių klasifikatorius. Vis dėlto, nepaisant aukšto bendro modelio tikslumo, išryškėja modelio ribotumas identifikuojant akademinės nesėkmės atvejus. Atsitiktinio miško algoritmo atkūrimo rodiklis „N“ klasei, naudojant 7 savaičių duomenis, siekia tik 0,389, o tai rodo, kad modelis teisingai atpažįsta mažiau nei pusę faktinių nesėkmės atvejų. Šis ribotumas gali būti siejamas su klasių disbalansu duomenų rinkinyje ir nedideliu imties dydžiu ( $n = 125$ ). Atsižvelgiant į tai, ateities tyrimuose tikslinga taikyti klasių disbalanso sprendimo metodus, tokius kaip SMOTE, siekiant pagerinti nesėkmės atvejų identifikavimo tikslumą.

Taigi po rezultatų įvertinimo sudarytas galutinis akademinės sėkmės prognozavimo modelis, kuriame taikomas atsitiktinio miško algoritmas.

## 2.6 Taikymas

Akademinės sėkmės prognozavimo modelis buvo realizuotas Weka 3.8 programinėje sistemoje, naudojant duomenų įkėlimo, klasės priskyrimo, validavimo ir klasifikavimo komponentus. Modelio struktūra ir duomenų apdorojimo seka pateikta 2 pav.

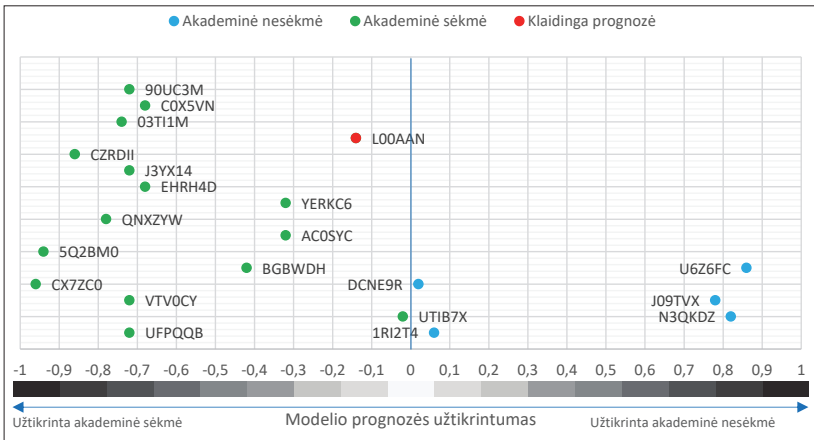


2 pav. Studentų akademinės sėkmės prognozavimo modelis Weka 3.8 programinėje sistemoje.

### 3 Rezultatai

Siekiant įvertinti modelio tinkamumą studijų programoje, atliktas jo testavimas naudojant tris duomenų rinkinius: apmokymo rinkinį, 2021 m. įstojusiujų duomenis (n = 21) ir 2022 m. įstojusiujų duomenis (n = 74).

Pirmiausia analizuojami rezultatai, gauti prognozuojant 2021 m. studentų akademinę sėkmę (žr. 3 pav.). Modelis nustatė, kad penkiems besimokantiesiems gresia studijų nebaigimas (klasė „N“).

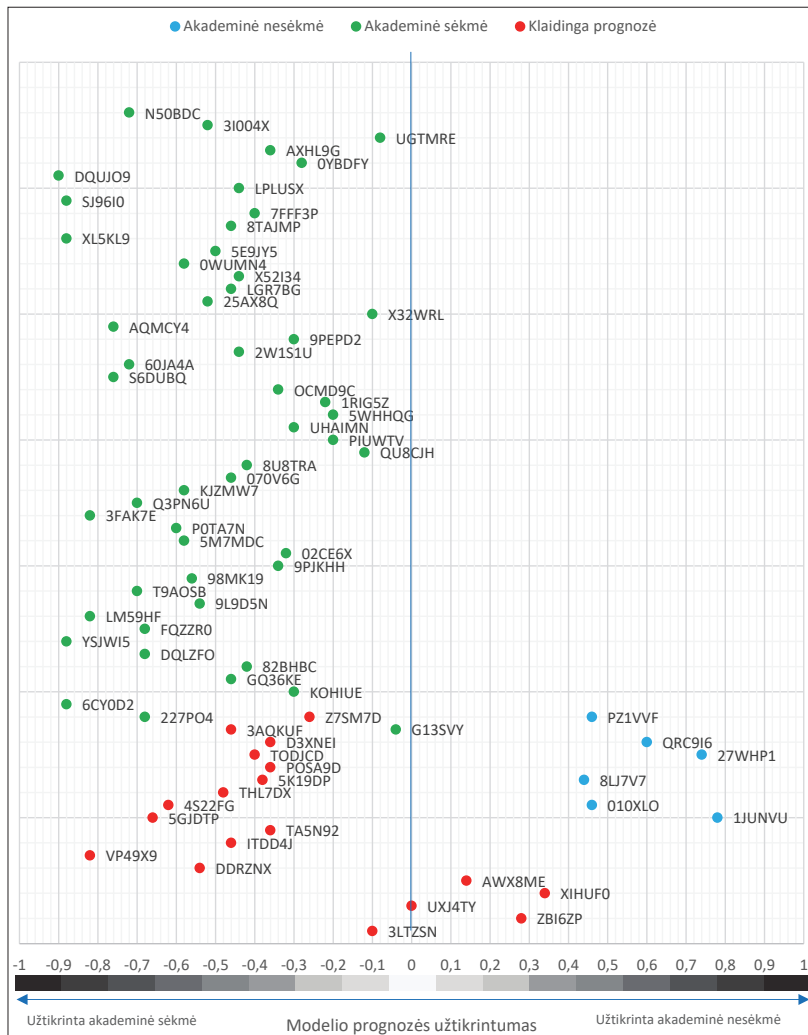


3 pav. Akademinės sėkmės prognozės 2021 m. įstojusiems studentams.

Remiantis užtikrintumo įverčiais, dviejų studentų prognozės dėl nesėkmės laikytinos mažiau patikimomis, nes jų reikšmės yra mažesnės nei 0,1, o dar dviejų studentų prognozės dėl sėkmės – taip pat mažai tikėtinos, nes užtikrintumo rodikliai yra mažesni nei -0,2.

Toliau analizuojami 2022 m. studentų akademinės sėkmės prognozavimo rezultatai (žr. 4 pav.). Modelis nustatė, kad 10 iš 74 studentų gresia studijų nutraukimas (klasė „N“).

Analuojant užtikrintumo įverčius, vieno studento priskyrimas klasei „N“ buvo silpnas, o dar vieno studento atveju modelis abiem klasėms priskyrė vienodą tikimybę (užtikrintumo įvertis lygus 0). Taip pat nustatyta, kad penkių studentų priskyrimas klasei „T“ pasižymi mažesniu užtikrintumu (žemiau  $-0,2$ ).



4 pav. Akademinės sėkmės prognozės 2022 m. įstojusiems studentams.

Akademinės sėkmės prognozavimo rezultatai buvo pateikti studijų programos administracijai, siekiant palyginti juos su realia studentų situacija. Remiantis gauta informacija modelio klaidos vizualizuotos 3 ir 4 paveiksluose.

Palyginus prognozes su realia situacija, nustatyta, kad modelis 11 iš 15 atvejų teisingai priskyrė „N“ reikšmę. Neteisingai klasifikuoti arba mažo užtikrintumo atvejai (8 studentai) dažniausiai pasižymėjo akademinės nesėkmės požymiais, tokiais kaip akademinės skolos, žemi pasiekimai ar paskaitų nelankymas.

Taip pat nustatyta, kad dalis studentų, kuriems buvo prognozuotas studijų tęsimas („T“), dėl įvairių priežasčių nutraukė studijas arba išėjo akademinį atostogų. Tai rodo, kad modelio tikslumas gali būti toliau tobulinamas siekiant sumažinti klaidų tikimybę.

## 4 Išvados

Šiame darbe parodyta, kad duomenų tyrybos technologijos taikymas leidžia prognozuoti besimokančiųjų akademinę sėkmę ir identifikuoti studentus, kuriems gresia nuotolinių studijų nutraukimas. Prognozavimas buvo vykdomas remiantis CRISP-DM modeliu, užtikrinančiu nuoseklų ir kokybišką tyrybos procesą.

Atlikta klasifikavimo algoritmų analizė parodė, kad tiksliausių rezultatus pateikė atsitiktinio miško algoritmas, kuris pasižymėjo aukščiausiais kokybės rodikliais.

Gauti rezultatai patvirtino, kad virtualiojoje mokymosi aplinkoje generuojami besimokančiųjų duomenys leidžia anksti identifikuoti akademinės rizikos studentus, sudarant prielaidas kurti ankstyvojo įspėjimo sistemas ir taikyti tikslias intervencijas.

Nepaisant to, modelio tikslumas ribojamas duomenų kokybės ir požymių pasirinkimo, todėl ateityje tikslinga plėsti duomenų rinkinį, įtraukiant papildomus akademinis ir psichologinius kintamuosius bei tobulinti modelio struktūrą.

## Literatūra

- [1] M. Xavier and J. Meneses. Dropout in Online Higher Education: A scoping review from 2014 to 2018. Barcelona: eLearn Center, Universitat Oberta De Catalunya. 2020, 1-61. Prieiga per: <https://femrecerca.cat/meneses/publication/dropout-online-higher-education-scoping-review/dropoutonline-higher-education-scoping-review.pdf>

- [2] A. R. Kowalski et al. Student Dropout in Distance Education in Higher Education: literature review in Portuguese. *EaD em Foco*. 2020, 10(2), 1-24. Prieiga per: [https://www.academia.edu/64271821/Student\\_Dropout\\_in\\_Distance\\_Education\\_in\\_Higher\\_Education\\_literature\\_review\\_in\\_Portuguese](https://www.academia.edu/64271821/Student_Dropout_in_Distance_Education_in_Higher_Education_literature_review_in_Portuguese)
- [3] E. Sosu and P. Pheunpha. Trajectory of University Dropout: Investigating the Cumulative Effect of Academic Vulnerability and Proximity to Family Support. *Front. Educ.* 2019, 4(6), 1-10. Prieiga per: [https://www.researchgate.net/publication/331040381\\_Trajectory\\_of\\_University\\_Dropout\\_Investigating\\_the\\_Cumulative\\_Effect\\_of\\_Academic\\_Vulnerability\\_and\\_Proximity\\_to\\_Family\\_Support](https://www.researchgate.net/publication/331040381_Trajectory_of_University_Dropout_Investigating_the_Cumulative_Effect_of_Academic_Vulnerability_and_Proximity_to_Family_Support)
- [4] A. Behr et al. Dropping out of university: a literature review. *Review of Education*. 2020, 8(2), 614-652. Prieiga per: <https://bera-journals.onlinelibrary.wiley.com/doi/epdf/10.1002/rev3.3202>
- [5] N. G. Rotem, G. Yair and E. Shustak. Dropping out of master's degrees: objective predictors and subjective reasons. *Higher education research & development*, 2021, 40(5), 1070-1084. Prieiga per: <https://www.tandfonline.com/doi/epub/10.1080/07294360.2020.1799951?needAccess=true>
- [6] M. Yildiz and C. Borecki. Predicting Academic Achievement with Machine Learning Algorithms. *Journal of Educational Technology & Online Learning*, 2020, 3(3), 372-392. Prieiga per: <https://dergipark.org.tr/en/download/article-file/1214052>
- [7] O. Niakšu. Duomenų tyryba medicinoje: taikymas, problemos ir galimybės. *Visuomenės sveikata*, 2014, 4(67), 9-19. Prieiga per: [https://visuomenessveikata.hi.lt/uploads/pdf/visuomenes%20sveikata/2014.4\(67\)/VS%202014%204\(67\)%20LIT%20A%20Duomenu%20tyryba.pdf](https://visuomenessveikata.hi.lt/uploads/pdf/visuomenes%20sveikata/2014.4(67)/VS%202014%204(67)%20LIT%20A%20Duomenu%20tyryba.pdf)