

Studijų proceso duomenų analizės modeliai ir programinė įranga

Vytautas JANILIONIS, Tomas RUZGAS (KTU)

el. paštas: vyjan@fmf.ktu.lt, tomas.ruzgas@fmf.ktu.lt

1. Įvadas

Darbe pateikti duomenų analizės modeliai ir programinė įranga, kurie palengvina sprendimų priėmimą bei jų pagrindimą sprendžiant studijų proceso organizavimo uždavinius.

Nemažai Europos ir JAV švietimo įstaigų bei universitetų kaupia studijų proceso informaciją, kuria duomenų analizės modelius ir analizės rezultatus naudoja priimant sprendimus. Studijų proceso duomenys kaupiami ir analizuojami ne tik atskirose švietimo įstaigose, bet ir valstybiniu ar net tarpvalstybiniu lygiu. Jungtinių Amerikos Valstijų nacionalinis švietimo statistikos centras kaupia ir analizuoja 40 pasaulio valstybių švietimo įstaigų pateiktus duomenis [7]. Didžiojoje Britanijoje panašų vaidmenį atlieka švietimo ir užimtumo departamentas [8]. Londono universitete, 1998 metais sukurtas socialinės aplinkos faktorių įtakos studijų sėkmei tyrimo modelių kompleksas, Vakarų Australijos universitete analizuojami studentų pažangumo duomenys ir tiriama, kokią įtaką studijoms universitete turi vidurinė mokykla, kurią studentas baigė. Panašūs tyrimai buvo atlikti ir Kauno technologijos universitete [5]. Tačiau, sprendimų priėmimui svarbių, pavienių rodiklių analizė, laukiamų rezultatų dažniausiai neduoda. Siekiant pagerinti analizės rezultatų efektyvumą, reikia integruoti atskirtas rodiklių grupes taip, kad studijų proceso organizatoriai ir vadovai susidarytų sisteminių vaizdą, pagrįstą viena su kita susijusiomis studijų proceso dalimis. Šiuo metu pasaulyje daug dėmesio skiriama sprendimų priėmimo paramos sistemų kūrimui. Jų branduolį sudaro duomenų statistinės analizės modeliai, todėl aktualus uždavinys – šių modelių, jų taikymo metodikų ir atitinkamos programinės įrangos kūrimas.

Duomenys apie studijų procesą. Kauno technologijos universitete (KTU) kaupiami jau virš 20 metų. Duomenų bazėje ABITURIENTAS saugomi duomenys apie KTU studentų mokymąsi vidurinėje mokykloje, duomenų bazėje STUDENTAI – informacija apie studentus (anketiniai duomenys, mokslo kryptis, studijų forma, svartinis mokymosi vidurkis, surinktų kreditų skaičius ir t.t.), duomenų bazėje INDIVIDUALŪS STUDIJŲ PLANAI – informacija apie jų mokymąsi universitete (kokius modulius studijavo, kokius įvertinimus gavo, ar laiku atsiskaitė ir t.t.). Šio darbo tikslas yra:

- įvertinus KTU duomenų bazių struktūrą ir šiuolaikinių informacijos išgavimo iš duomenų technologijų bei programinės įrangos galimybes, sukurti studijų proceso duomenų statistinės analizės sistemos struktūrą;

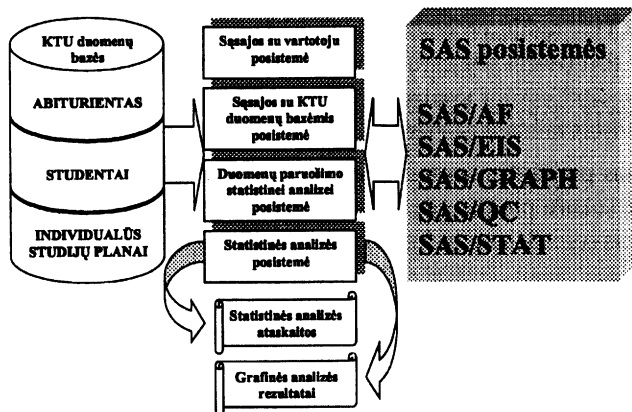
- sukurti KTU studijų proceso duomenų bazių statistinės analizės modelius, kurie leistų tirti įvairių požymių (mokslo kryptis, fakultetas, modulio grupė, studijų forma, kreditai, sesijos svertiniai požymių vidurkiai, atskirų studijų modulių arba jų grupių įvertinimai ir t.t.) tarpusavio ryšius ir priklausomybes;
- panaudojus pasiūlytus duomenų analizės modelius ir statistinės analizės sistemos struktūrą, sukurti programinę įrangą, kuri palengvintų sprendimų priėmimą bei jų pagrindimą sprendžiant studijų proceso organizavimo klausimus.

2. Studijų proceso duomenų statistinės analizės sistema

Sprendimų priėmimo paramos sistemos sėkmingas kūrimas ir veikimas priklauso nuo to, kaip švietimo įstaiga organizuoja keturis pagrindinius etapus: duomenų rinkimą, jų pavertimą prasminga informacija, tos informacijos perdavimą bei papildomą pateiktos informacijos detalizavimą pagal vartotojų poreikius. Kalbant informacinių technologijų terminais, šie etapai interpretuojami, kaip:

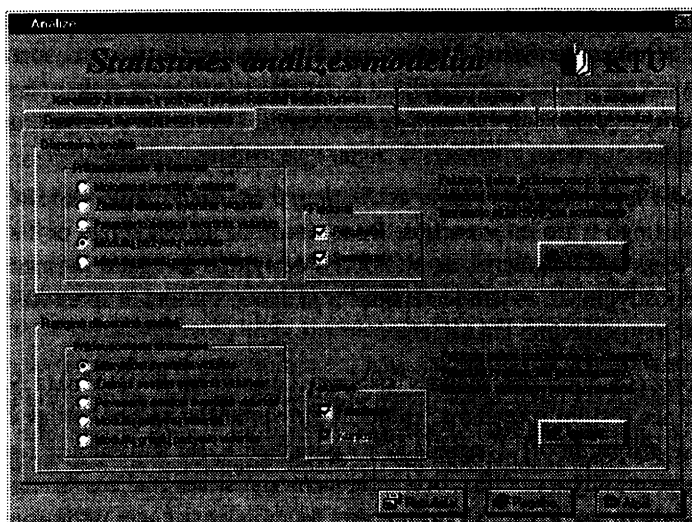
- **Duomenų sandėliavimas (Data Warehousing)** [1]. Sėkmingas šio uždavinio sprendimas padeda standartizuoti duomenų tvarkymą ir tuo pačiu ženkliai palengvina jų analizę. Tik tuo atveju, kai visa informacija yra prieinama ir tinkamai išdėstyta, galima gauti teisingą vaizdą apie studijų procesą.
- **Informacijos iš gavimas iš duomenų (Data Mining)**. Rodiklių parinkimas sprendimų priėmimui ir jų optimalių reikšmių radimas, neturint tinkamų priemonių, gali pareikalauti daug laiko sąnaudų. Čia gali padėti informacijos išgavimo iš duomenų priemonės. Tyrinėjant duomenis, galima sužinoti kokius klausimus švietimo įstaiga turėtų sau užduoti apie vykdomą veiklą ir įvertinti sąryšius tarp skirtingų veiklos rodiklių.
- **OLAP (Online Analytical Processing)**. Ši technologija užtikrinti operatyvią ir interaktyvią sąveiką su informacija, t.y. sistemos vartotojas gauna ne statines, bet dinamines daugiamates hierarchines ataskaitas, kuriose jis pats gali formuoti analizuojamus duomenų pjūvius ir skaičiuoti pasirinktus rodiklius.
- **Stebėtojo technologija (Viewer technology)**. Skirtingo lygio vadovai turi prieiti prie skirtingo lygio informacijos. Be to, informacija reikalinga ne tik vadovams. Kiekvienas administracijos darbuotojas ir dėstytojas turi gauti informaciją tuo detalizacijos lygiu, kuris jam leidžiamas ir ją naudoti priimant sprendimus mokymo proceso organizavimo klausimais.

Studijų proceso duomenų analizės sistemos kūrimui buvo pasirinkta SAS programinė įranga, nes ji siūlo sprendimą, apimančią pirmaujančias technologijas visose aukščiau minėtose srityse [1–4]. Įvertinus KTU duomenų bazių struktūrą ir sistemos SAS galimybes buvo pasiūlyta studijų proceso duomenų statistinės analizės sistemos struktūra (1 pav.). Ją sudaro: sąsajos su KTU duomenų bazėmis posistemė; duomenų paruošimo statistinei analizei posistemė; statistinės analizės posistemė; sąsajos su vartotoju posistemė. Sąsajos su KTU duomenų bazėmis posistemė perkelia reikiamus duomenis iš KTU



1 pav. Studijų proceso duomenų statistinės analizės sistemos struktūra.

duomenų bazių. Duomenų paruošimo statistinei analizei posistemė formuoja duomenų pjūvius ir paruošia duomenų matricą statistinių metodų taikymui. Statistinės analizės posistemę sudaro įvairūs duomenų analizės modeliai. Visas šias tris posistemas apjungia interaktyvios sąsajos su vartotoju posistemė (2 pav.), kurios programos sukurtos panaudojus sistemos SAS modulius SAS/AF ir SAS/EIS [4]. Statistiniai duomenų analizės modeliai sukurti panaudojus SAS modulius SAS/STAT, SAS/GRAPH ir SAS/QC [2–3]. Programinės įrangos naudojimo palengvinimui sukurta pagalbos sistema, kuri integruota į bendrą sistemos aplinką.



2 pav. Statistinės analizės sistemos vartotojo sąsaja.

3. Statistiniai duomenų analizės modeliai

Atlikus KTU duomenų bazėse saugomos informacijos analizę ir įvertinus kokia informacija reikalinga sprendimų priėmimui, pasiūlyti šie statistinės analizės modeliai:

1. Aprašomosios statistikos daugiamatės ataskaitos ir grafikai:

- studentų duomenų bazės ataskaitos (hierarchija: **Studentai => Atsiskaitymas => Fakultetas => Kursas => Grupė => Lytis**);
- individualių studijų planų duomenų bazės ataskaitos (hierarchija: **Studijos => Studijų forma => Semestras => Mokslo sritis => Mokslo kryptis => Mokslo šaka => Modulis => Kreditų skaičius**);
- studentų ir individualių studijų planų duomenų bazių duomenų grafinis vaizdavimas.

2. Parametrų įverčiai ir hipotezių tikrinimas:

- įvairių rodiklių taškiniai ir intervaliniai įverčiai;
- suderinamumo hipotezių apie rodiklių (mokymosi svertinis vidurkis, modulių grupės įvertinimų vidurkis ir t.t.) skirstinius tikrinimas;
- parametrinių hipotezių apie vidurkių lygybę tikrinimas (rudens ir pavasario semestrų svertinių vidurkių, moterų ir vyrų mokymosi svertinių vidurkių ir t.t.).

3. Klasterinė analizė:

- fakultetų klasifikavimas pagal studijų modulių grupių pažymių vidurkius.

4. Dispersinė analizė:

- kurso ir fakulteto faktorių įtakos mokymosi svertiniams vidurkiams tyrimas;
- fakulteto faktoriaus įtakos studijų modulio pažymių vidurkiui tyrimas;
- fakulteto ir semestro faktorių įtakos studijų modulių grupės pažymių vidurkiui tyrimas.

5. Koreliacinė analizė ir požymių priklausomumo lentelių tyrimas:

- sąryšio stiprumo tarp mokymosi svertinių vidurkių tyrimas;
- sąryšio stiprumo tarp studijų modulio pažymių ir studijų modulių grupių pažymių vidurkių tyrimas, grupuojant duomenis pagal semestrus ir fakultetus;
- sąryšio tarp lyties, fakulteto, studijų formos, vietos (laisvo klausytojo, papildoma, valstybinė), būsenos (išbrauktas, studijuoja, kita) ir kitų požymių tyrimas.

6. Logistinė regresija ir kiti modeliai:

- mokymosi svertinių vidurkių ir studijų modulių grupių pažymių vidurkių prognozė priklausomai nuo studijų formos ir kurso;

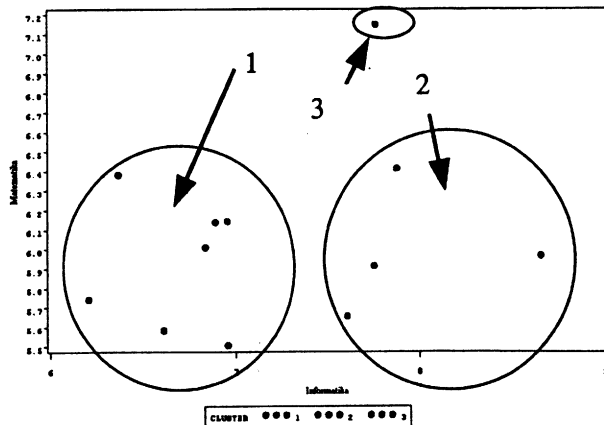
- studentų iškritimo ir neatsiskaitymo už studijų modulius tikimybių įverčių radimas grupuojant duomenis pagal fakultetus ir t.t.

Sukurtų modelių ir visos sistemos testavimas, atliktas naudojant realius KTU studijų proceso duomenis, parodė, kad sistema yra pajėgi spręsti statistinės analizės užduotis esant didelėms imtims. Pateiksime vieno duomenų analizės modelio taikymo pavyzdį. Panaudojus klasterinę analizę suklasifikuoti objektai (fakultetai) į grupes pagal matematikos ir informatikos kryptčių studijų modulių pažymių vidurkius. Buvo analizuoti pirmo kurso, dieninių bakalauro studijų studentų 1999/2000 mokslo metų pažymiai.

Gautos trys fakultetų grupės (3 pav.):

- ▷ į pirmą grupę pateko fakultetai, kuriuose studentų matematikos kryptties studijų modulių pažymių vidurkiai yra žemi, o informatikos kryptties – vidutiniai (Elektrotechnikos ir automatikos, Mechanikos, Dizaino ir technologijų, Informatikos, Statybos ir architektūros, Chemijos technologijų, Fundamentalųjų mokslų fakultetai);
- ▷ į antrą grupę pateko fakultetai, kuriuose studentų matematikos kryptties modulių pažymių vidurkiai žemi, o informatikos kryptties – aukšti (Ekonomikos ir vadybos, Telekomunikacijų ir elektronikos ir Socialinių mokslų fakultetai bei Tarptautinių studijų centras);
- ▷ į trečią grupę pateko fakultetas, kuriame studentų matematikos ir informatikos kryptties studijų modulių pažymių vidurkiai yra aukšti (Panevėžio institutas).

Gavus šiuos rezultatus, studijų proceso organizatorius ir vadovus, turėtų dominti šie klausimai: ar Panevėžio institute mokosi geresni studentai, ar ten geriau dėsto matematiką ir informatiką, ar ten mažiau reikalaujama iš studentų atsiskaitant už modulius? Kodėl toks žemas matematikos pažymių vidurkis pirmajame semestre likusiuose KTU fakultetuose? Ar informatikos pažymių vidurkis pirmame kurse objektyviai atspindi jų pasirošimą informatikos srityje? Ar tikrai vadybos ir socialinių mokslų fakultetų studentai geriau žino informatikos pagrindus už tikslųjų mokslų fakultetų studentus?



3 pav. Klasterinės analizės rezultatai.

4. Išvados

Įvertinus KTU studijų proceso duomenų bazių struktūrą, šiuolaikinių statistinės analizės sistemų ir informacijos išgavimo iš duomenų technologijų galimybes, pasiūlyta studijų proceso duomenų statistinės analizės sistemos struktūra, kuri realizuota panaudojus SAS programinę įrangą.

Panaudojus aprašomosios statistikos, koreliacinės analizės, požymių priklausomumo lentelių, klasterinės analizės ir kitus statistikos metodus, duomenų sandėliavimo, daugiamatčių hierarchinių ataskaitų ir grafinio duomenų vaizdavimo technologijas, sukurti KTU studijų proceso duomenų statistinės analizės modeliai, kurie leidžia tirti įvairių faktorių tarpusavio ryšius ir priklausomybes.

Sukurta sistema gali būti taikoma atliekant studijų proceso duomenų statistinę analizę, tiriant atskirų studijų modulių arba jų grupių mokymosi rezultatų įtaką studijų sėkmei, analizuojant studentų iškritimo priežastis ir sprendžiant kitus uždavinius. Sistemą nesunkiai galima pritaikyti kitų statistinės analizės uždavinių sprendimui, vienus statistinius modelius pakeičiant kitais, arba naujus prijungiant prie jau esančių.

Literatūra

- [1] R. Kimball, *Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*, SAS Institute Inc., Cary (2001).
- [2] R.P. Cody, J.K. Smith, *Applied Statistics and the SAS Programming Language*, Prentice Hall, New Jersey (1997).
- [3] M.E. Stokes, C.S. Davis, G.G. Koch, *Categorical Data Analysis Using the SAS System*, SAS Institute Inc., Cary (1995).
- [4] *SAS/AF[®] it Software: Usage and Reference, Version 6*, SAS Institute Inc., Cary (1995).
- [5] V. Janilionis, Matematikos ir fizikos mokymosi rezultatų mokykloje ir KTU statistinė analizė, *Matematika ir matematikos dėstymas, Konferencijos pranešimų medžiaga*, Technologija, Kaunas, 137–142 (1999).
- [6] V. Janilionis, T. Ruzgas, Studijų proceso duomenų statistinės analizės modeliai, *Matematika ir matematinis modeliavimas, Konferencijos pranešimų medžiaga*, Technologija, Kaunas, 69–74 (2001).
- [7] *National Center for Education Statistics*, <http://www.nces.ed.gov>.
- [8] *Department for Education and Employment*, <http://www.dfee.gov.uk>.

Models and software of study process data analysis

V. Janilionis, T. Ruzgas

The models of study process data analysis and statistical software for decision-making support are presented in this paper. The models and software were developed using applied statistics methods (descriptive statistics, analysis of variance, nonparametric tests, cluster analysis, correlation and logistic regression analysis) as well as data warehousing, multidimensional reporting, visualisation, object programming technologies and software SAS. The developed system is used for the analysis of study process data of Kaunas University of Technology. It enables to analyse the relationships and dependencies between various factors (e.g. study areas, faculties, study modules groups, form of studies, credits, weighted averages of the exams grades, grades of study modules or their groups etc.) and use the results for the effective decision-making.