

E. MOKYMOŠI KURSŲ DUOMENŲ ANALIZĖ, TAIKANT DUOMENŲ GAVYBOS METODUS

Jurgita Lieponienė

Panevėžio kolegija

Anotacija

Sparčiai didėjant apdorojamų ir saugomų duomenų kiekiui, aktuali tampa duomenų analizė, padedanti priimti greitus, efektyvius ir teisingus sprendimus. Duomenų analizei atlikti taikomi duomenų gavybos metodai. Duomenų gavybos metodais iš duomenų išgaunamos naudingos žinios. Straipsnyje apžvelgiamos duomenų gavybos technologijų taikymo e. mokymesi galimybės, atrenkami e. mokymosi kurso duomenys ir parenkami duomenų gavybos metodai e. mokymosi kurso duomenų analizei atlikti. Pristatomi atlikto tyrimo, kuriame duomenų gavybos metodais analizuojami e. mokymosi kurso duomenys, rezultatai. Atlikto tyrimo rezultatų pagrindu apibrėžiamos e. mokymosi kurso tobulinimo kryptys.

Esminiai žodžiai: e. mokymosi sistema, duomenų gavyba, e. mokymosi kursas, klasterizavimas, asociacijų paieška.

Įvadas

E. mokymasis – neatsiejama aukštojo mokslo dalis. Siekiant kokybiškesnių ir efektyvesnių savarankiškų studentų studijų, e. mokymasis integruojamas į tradicines studijas. Užtikrinant e. mokymosi kokybę, aktualus tampa e. mokymąsi realizuojančios aplinkos tobulinimo klausimas. S. Kotzer, Y. Elran (2012) e. mokymosi kursą įvardija kaip vieną iš svarbiausių e. mokymąsi realizuojančios aplinkos komponentų. Todėl e. mokymosi kursų tobulinimas neatsiejama e. mokymosi proceso tobulinimo dalis.

Kuriant e. mokymosi kursus, naudojami įvairūs kursų kūrimo įrankiai arba pasitelkiamos e. mokymosi sistemos integruotos e. mokymosi kursų kūrimo priemonės. Parengti e. mokymosi kursai teikiami e. mokymosi sistemoje, kurios kaupia įvairius studentų mokymosi duomenis: studentų aktyvumo duomenis, vertinimus, duomenis apie studentų mokymosi eigą. Pasak L. M. Anduela (2011), e. mokymosi sistemoje sukauptų duomenų analizė svarbi, tobulinant e. mokymosi procesą.

Analizuoti duomenis galima įvairiais statistikos metodais, tačiau kai duomenų kiekis tampa didelis, tenka taikyti duomenų gavybos metodus (Preidys, Sakalauskas, 2012). Duomenų gavyba – tai procesas, naudojantis įvairius duomenų analizės įrankius, kurie padeda atrasti tokias duomenų struktūras ir ryšius, kurie būtų panaudojami realioms išvadoms ir sistemos rezultatams apibrėžti (Preidys, Sakalauskas, 2012). Duomenų gavybos rezultatai padeda turėti geresnę esamos situacijos išvalgą, suprasti dėsningumus ir numatyti ateitį.

Tyrimo tikslas: naudojant duomenų gavybos metodus, atlikti e. mokymosi kurso duomenų analizę, numatyti e. mokymosi kurso tobulinimo kryptis, siekiant efektyvesnių ir kokybiškesnių studijų.

Tyrimo uždaviniai:

1. Apžvelgti duomenų gavybos taikymo e. mokymesi tyrimus.
2. Atrinkti e. mokymosi kurso duomenis duomenų analizei atlikti.
3. Parinkti tinkamus duomenų gavybos metodus.
4. Pristatyti e. mokymosi kurso duomenų analizės, taikant duomenų gavybos metodus, rezultatus.
5. Numatyti e. mokymosi kursų tobulinimo kryptis.

Tyrimo metodika. Nagrinėjant duomenų gavybos taikymo e. mokymesi tyrimus, parenkant duomenų gavybos metodus, buvo atliekama mokslinių straipsnių analizė. Duomenų analizė atlikta, taikant klasterizavimo ir asociacijų paieškos metodus.

Duomenų gavybos taikymo e. mokymesi tyrimai

Susidomėjimas duomenų gavybos technologijų taikymu e. mokymesi nuolat auga. Pasak L. Shen, M. Wang, R. Shen (2009) e. mokymosi sistemų duomenų bazėje saugoma daug naudingos informacijos, kurią galima efektyviai panaudoti e. mokymosi proceso tobulinimui. Duomenų gavybos metodais analizuojant e. mokymosi sistemos duomenų bazėje sukauptus duomenis, galima nustatyti studentų mokymosi efektyvumą, išsiaiškinti priežastis, kurios įtakoja vienos ar kitos veiklos neigiamus rezultatus (Romero ir kt. 2008). Naudojant e. mokymosi sistemų sukauptus studentų veiklos duomenis bei pritaikius duomenų gavybos metodus, galima nustatyti besimokančiųjų mokymosi stilių, būdą (Liu, Shih 2007), sugrupuoti studentus pagal jų mokymosi tempą, pažangumą (Romero ir kt. 2008).

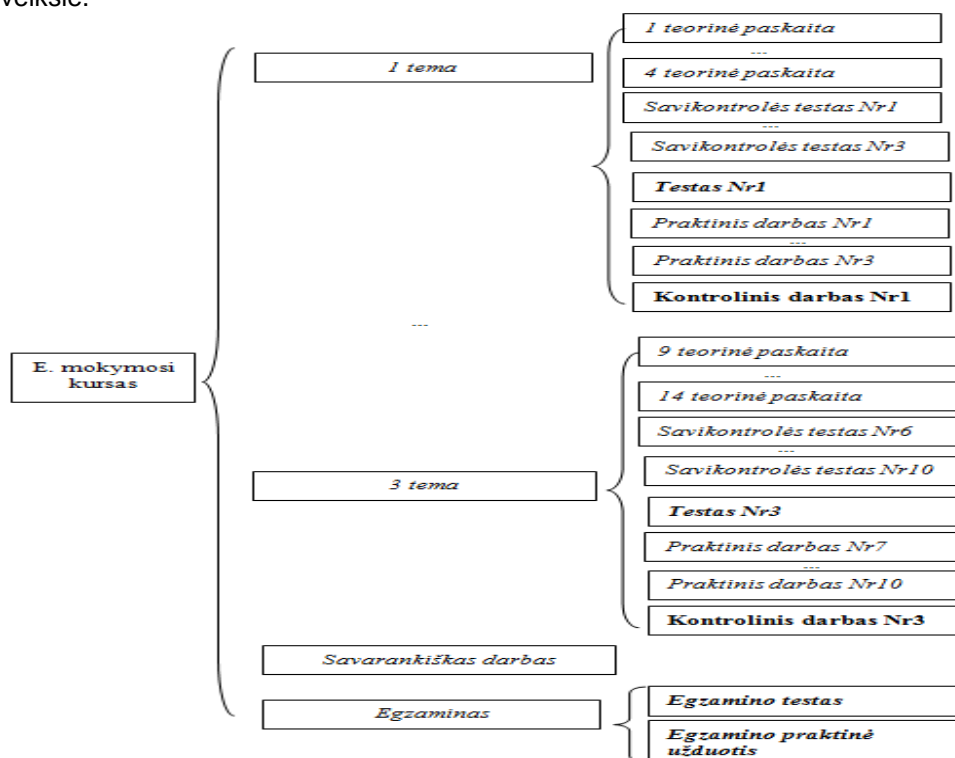
Duomenų gavybos technologijos naudojamos pritaikant e. mokymosi kurso turinį pagal besimokančiojo žinias ir gebėjimus (Klasnja-Milicevic ir kt. 2011), kuriant agentus, siūlančius besimokančiajam reikalingą mokymo medžiagą arba reikalingas nuorodas (Preidys, Sakalauskas 2010). Duomenų gavybos metodais galima diagnozuoti besimokančiųjų problemas, analizuoti besimokančiųjų

veiklą ir nukreipti ją tinkama linkme (Romero ir kt. 2008). Duomenų gavybos technologijos taikomos įrankių, skirtų sumažinti studentų plagiatu ir kitokių nesąžiningų veiklų atsiradimo galimybes, kūrimui (Scherbinin, Butakov 2008). Literatūroje aprašoma eilė mokslinių tyrimų, kuriuose duomenų gavybos metodais analizuojami e. mokymosi kursų vertinamų veiklų duomenys. A. Merceron, K. Yacef (2005) aprašė atlikto tyrimo rezultatus, kuriame duomenų gavybos metodais nagrinėjo studentų veiklos ir galutinių rezultatų priklausomybes. V. Namdeo, A. Singh, D. Singh, R. C. Jain (2010) taikė įvairius klasifikavimo algoritmus tirdami e. mokymosi kurso studijose gautų vertinimų priklausomybes. M. Delgado Calvo-Flores, E. Gibaja Galindo (2006) prognozavo studentų įvertinimus, pasitelkę neuroninius tinklus.

Analizuojant e. mokymosi sistemų duomenų bazėse saugomus duomenis, taikomi įvairūs duomenų gavybos metodai: klasifikavimas, klasterizavimas, asociacijų paieška, prognozavimas. Literatūroje aprašomi įvairūs šių metodų taikymo algoritmai. Duomenų gavybos metodo ir jį realizuojančio algoritmo parinkimas priklauso nuo turimų duomenų, iškeltų tyrimo tikslų ir siektinų rezultatų (Namdeo ir kt. 2011).

Duomenų atrinkimas ir apdorojimas

E. mokymosi sistemų duomenų bazėje kaupiami įvairūs besimokančiųjų duomenys: kada besimokantysis prisijungė prie kurso, kokias veiklas atliko, kiek laiko skyrė atskiroms kurso veikloms, kokius vertinimus gavo. Šiuolaikinės e. mokymosi sistemos kurso kuratoriams ir dėstytojams siūlo įvairias ataskaitas, skirtas peržiūrėti studentų aktyvumo, kurso veiklų naudojimo, vertinimų duomenis. E. mokymosi sistemų siūlomi įrankiai įgalina testų analizę. Analizuojant ataskaitų duomenis, kurso kuratorius ir dėstytojas gali įvertinti besimokančiųjų aktyvumą, tobulinti kursų testus, sekti besimokančiųjų pažangą. E. mokymosi sistemų generuojamos ataskaitos teikia paramą kursų kuratoriams ir dėstytojams, tačiau e. mokymosi kursų efektyvumo vertinimui sistemos generuojamų ataskaitų nepakanka. Atliekant tyrimą, pasitelkus duomenų gavybos metodus buvo analizuojami Duomenų bazių projektavimo kurso trejų metų duomenys. Šis e. mokymosi kursas teikiamas e. mokymosi sistemoje Moodle. Analizuojamo e. mokymosi kurso struktūra pateikta 1 paveiksle.



1 pav. E. mokymosi kurso struktūra

Duomenų bazių projektavimo kursui parinktas teminis kurso veiklų išdėstymo šablonas. Analizuojamo e. mokymosi kurso mokomoji medžiaga suskirstyta į 5 temas, kuriose pateikta 14 teorinių paskaitų mokomoji medžiaga, 10 praktinių darbų, teorinių žinių savikontrolėi ir įtvirtinimui numatyta 10 savikontrolės testų, sukurti 3 kontroliniai darbai ir testai, pateikta savarankiško darbo užduotis ir egzamino užduotys. Tyrimo nagrinėti 100 iššestinių ir nuolatinų studijų studentų mokymosi rezultatai.

Tyrimui reikalingi duomenys buvo atrinkti iš e. mokymosi sistemos Moodle duomenų bazės. Atrinkties duomenims apibendrinti buvo sukurtos procedūros. Atrinktą ir apibendrintą duomenų rinkinį sudaro 11 laukų. Atrinktuose įrašuose saugoma: kiekvieno studento atliktų savikontrolės testų skaičius, šiems testams atlikti sugaištas laikas, kiek vidutiniškai kartų studentai atliko savikontrolės testus, savikontrolės testų maksimalių vertinimų vidurkis, testams sugaištas laikas, testų įvertinimų vidurkis, praktinių darbų

įvertinimų vidurkis, galutinis vertinimas ir savarankiško darbo vertinimas. Tyrimui atrinkti duomenys aprašyti 1 lentelėje.

1 lentelė. Duomenys naudojami e. mokymosi kurso duomenų analizėje

| Eil. Nr. | Apibūdinimas |
|----------|---|
| 1 | Atliktų savikontrolės testų skaičius |
| 2 | Savikontrolės testams atlikti sugaištas laikas (min) |
| 3 | Kiek vidutiniškai kartų buvo atlikti savikontrolės testai |
| 4 | Savikontrolės testų maksimalių vertinimų vidurkis |
| 5 | Testams atlikti sugaištas laikas (min) |
| 6 | Testų vidurkis |
| 7 | Praktinių darbų įvertinimų vidurkis |
| 8 | Kontrolinių darbų įvertinimų vidurkis |
| 9 | Skaitytų teorinių paskaitų skaičius |
| 10 | Savarankiško darbo vertinimas |
| 11 | Galutinis vertinimas |

Tyrimui atrinkti ir apibendrinti duomenys buvo konvertuoti į CSV formatą. Duomenų analizei atlikti parinktas laisvai platinamas duomenų gavybos įrankis WEKA. Šis įrankis 1997 metais buvo sukurtas Vaikoto universitete (Naujoji Zelandija). WEKA įrankis realizuotas Java programavimo kalba. Duomenų gavybos įrankis yra funkcionalus, su patogia grafine vartotojo sąsaja (Sharma, Bajpai, Litoriya 2012). Naudojant šį įrankį galima atlikti duomenų pradinį apdorojimą, klasterizavimą, klasifikavimą, asociacijų paiešką, regresiją ir vizualizavimą. WEKA sistema – populiarus duomenų gavybos įrankis. Nors ši sistema nėra pritaikyta didelės apimties duomenų apdorojimui ir analizei, joje įgyvendinti duomenų gavybos metodai pajėgūs susidoroti su nemažomis duomenų aibėmis (Paulauskienė, Kurasova 2013). Įkėlus sutvarkytą duomenų rinkinį į WEKA programą, duomenys buvo papildomai apdorojami, parengiant juos atitinkamo duomenų gavybos metodo taikymui.

Duomenų gavybos metodų parinkimas

Duomenų analizei atlikti parinkti klasterizavimo ir asociacijų paieškos metodai. Klasterizavimas apibrėžiamas kaip objektų suskirstymas į grupes (klasterius), kuriose objektų skirtumai yra kuo mažesni, o tarp grupių skirtumai – kuo didesni. Klasterizavimui pasirinktas vienas iš nehierarchinių klasterinės analizės metodų – k vidurkių metodas. Kadangi nagrinėjama duomenų aibė buvo nuspręsta suskaidyti į du klasterius, šis klasterinės analizės metodas yra tinkamas numatytiems rezultatams pasiekti. Klasterizavimo procedūrą sudaro tokie etapai: atsitiktinai inicijuojami klasterių centrai; kiekvienas analizuojamos duomenų aibės vektorius yra priskiriamas tam klasteriui, iki kurio atstumas nuo centro yra mažiausias; perskaičiuojami kiekvieno klasterio centrai; skaičiuojama kvadratinė paklaida tarp klasterio centro ir klasteriui priskirtų duomenų; algoritmas kartojamas kol analizuojami duomenys nebepasiskirsto kitiems klasteriams (Stefanovič, Kurasova, 2013). K vidurkių metodo kokybei įvertinti skaičiuojama gautų klasterių atstumų nuo klasterio centro iki jam priskirtų duomenų suma (Stefanovič, Kurasova, 2013).

Asociacijų paieška – tai ryšių tarp analizuojamų duomenų suradimas. Atliekant tyrimą, asociacijų paieškai pasirinktas Apriori algoritmas. Taikant Apriori algoritmą, pirmiausia ieškoma dažnų elementų, o paskui iš šių elementų generuojamos kandidatinių sekos (Savulionienė, Sakalauskas, 2011). Norint sutrumpinti susietumo taisyklių paiešką naudojama aprioriškumo savybė, t. y. jeigu rinkinys Z yra nedažnas, tai šio rinkinio papildymas bet koku nauju elementu A šio rinkinio Z nepadaro dažno. Jeigu Z nedažnas, tai Z+A taip pat nedažnas (Savulionienė, Sakalauskas, 2011). Apriori algoritmą sudaro šie žingsniai: dažnų vieno elemento rinkinių radimas; potencialiai dažnų elementų rinkinių generavimas (kandidatų generavimas) ir kandidatų rinkinių dažnumo nustatymas. Pirmame algoritmo žingsnyje peržiūrimas visas duomenų rinkinys ir nustatoma, kiek kartų kiekvienas elementas pasikartoja duomenų rinkinyje. Tęsiant algoritmą apdorojami tik tie elementai, kurie tenkina nustatytą minimalų pasirodymų dažnį. Apriori algoritmas generuoja kandidatinius rinkinius tik iš rastų dažnų rinkinių prieš tai atliktame žingsnyje (Savulionienė, Sakalauskas, 2011).

Duomenų analizės rezultatai

Duomenų analizė pradedama, duomenų rinkinio klasterizavimu. Naudojant WEKA įrankį, analizuojamas duomenų rinkinys suskaidomas į du klasterius. Pirmasis duomenų klasteris apima puikiai, labai gerai ir gerai besimokančiųjų studentų duomenis, antrajam klasteriui priskiriami studentai, kurių analizuojamo dalyko (modulio) žinios ir gebėjimai buvo įvertinti pakankamai, patenkinamai ir silpnai. Pirmasis klasteris sudaro 51,0 proc. visų analizuojamų duomenų, antrajam klasteriui tenka 49,0 proc. nagrinėjamų įrašų (2 lentelė).

2 lentelė. Statistiniai klasterių duomenys

| Atributai | Klasterių vidurinės reikšmės | | |
|---|------------------------------|------------------|------------------|
| | Visi duomenys (100) | 1 klasteris (51) | 2 klasteris (49) |
| Atliktų savikontrolės testų skaičius | 6.33 | 8.6471 | 3.9184 |
| Savikontrolės testams atlikti sugaištas laikas (min) | 181.545 | 296.5588 | 61.8367 |
| Kiek vidutiniškai kartų buvo atlikti savikontrolės testai | 1.334 | 1.9882 | 0.6531 |
| Savikontrolės testų maksimalių vertinimų vidurkis | 7.815 | 9.5098 | 6.051 |
| Testams atlikti skirtas laikas (min) | 85.5 | 86.3922 | 84.5714 |
| Testų vidurkis | 6.775 | 8.3039 | 5.1837 |
| Praktinių darbų įvertinimų vidurkis | 8.105 | 9.6667 | 6.4796 |
| Kontrolinių darbų įvertinimų vidurkis | 7.075 | 8.4608 | 5.6327 |
| Skaitytų teorinės medžiagos temų skaičius | 10.01 | 13.5686 | 6.3061 |
| Savarankiško darbo vertinimas | 8.47 | 9.3137 | 7.5918 |
| Galutinis vertinimas | 6.95 | 8.3529 | 5.4898 |

Analizuojant statistinius klasterių duomenis galime pastebėti, kad skiriasi į klasterius sugrupuotų įrašų atliktų savikontrolės testų skaičius. Pirmąjį klasterį sudaro studentų, kurie įgytų teorinių žinių savikontrolėi ir įvertinimui naudojo savikontrolės testus, duomenys. Pirmojo klasterio analizuojamo atributo vidurinė reikšmė – 8.6471. Kadangi e. mokymosi kurse buvo pateikta 10 savikontrolės testų, tai rodo, kad šio klasterio studentai atliko visus arba beveik visus savikontrolės testus. Antrojo klasterio atliktų savikontrolės testų skaičiaus vidurinė reikšmė – 3.9184, kuri rodo, kad šios grupės studentai savikontrolės testams neskyrė ypatingos reikšmės.

Analizuojant išskirtų klasterių atributą – kiek vidutiniškai kartų buvo atlikti savikontrolės testai, paaiškėjo, kad pirmojo klasterio studentai tam tikrus savikontrolės testus kartuoja po kelis kartus, kai tuo tarpu antrojo klasterio studentų šio atributo vidurinė reikšmė siekia tik 0.6531 karto. Ženkliai skiriasi į klasterius sugrupuotų studentų testų, kontrolinių darbų ir praktinių darbų įvertinimų vidurkiai. Pirmojo klasterio studentų testų vidurkio vidurinė reikšmė siekia – 8.3039, kai tuo tarpu antrojo klasterio tik – 5.1837. Pirmojo klasterio kontrolinių darbų įvertinimų vidurkio vidurinė reikšmė – 8.9412, tuo tarpu antrojo klasterio – 5.6327. Gerai, labai gerai arba puikiais pažymiais analizuojamą kursą baigusių besimokančiųjų praktinių darbų įvertinimų vidurkio vidurinė reikšmė – 9.6667, silpnai, patenkinamai arba pakankamai įvertintų besimokančiųjų – 6.4796.

Analizuojant klasterių duomenis paaiškėjo, kad pirmojo klasterio studentai atsakingai studijavo teorinių paskaitų mokomąją medžiagą. Šių studentų skaitytų teorinės medžiagos temų skaičiaus vidurinė reikšmė – 13.5686. Kadangi e. mokymosi kurse buvo pateikta 14 teorinių paskaitų mokomoji medžiaga, klasterių statistiniai duomenys rodo, kad pirmojo klasterio studentai skaitė visas arba beveik visas teorinių paskaitų temas. Antrojo klasterio studentų skaitytų teorinės medžiagos temų skaičiaus – 6.3061 rodo, kad šio klasterio studentai mažai studijavo kurso teorinę mokomąją medžiagą.

Duomenų analizė tęsiama, taikant asociacijų paieškos metodą. Asociacijų paieškos metodu ieškomi sąryšiai tarp studentų skaitytų teorinės medžiagos temų skaičiaus ir galutinio kurso vertinimo. Parengiant duomenis šio metodo taikymui, vykdomas duomenų diskretizavimas. Galutiniai studentų vertinimai suskirstomi į tris kategorijas: jei įvertinimas yra tarp 5 ir 6 balų, priskiriama reikšmė - SILPNAI ARBA PATENKINAMAI, jei galutinis vertinimas yra 7 arba 8 – PAKANKAMAI ARBA GERAI, jei galutinis vertinimas apima 9 arba 10 tuomet priskiriama reikšmė – LABAI GERAI ARBA PUIKIAI. Atributo Skaitytų teorinės medžiagos temų skaičius reikšmės taip pat suskirstomos į tris kategorijas: jei skaitytų temų skaičius yra mažesnis nei 6 temos, priskiriama reikšmė – MAŽIAU NEI PUSĖ; jei skaitytų temų skaičius lygus 6, 7, 8 temoms priskiriamas atitikmuo – PUSĖ; jei skaitytų temų skaičius yra didesnis nei 8 priskiriamas atitikmuo – DAUGIAU NEI PUSĖ. Asociacijų paieška vykdoma, taikant Apriori algoritimą. Nustatyta minimali ieškomų asociatyvių taisyklių patikimumo reikšmė – 0.8. Gautos asociacijos (3 lentelė) rodo, kad studentai, kurie kurso studijas baigė silpnais arba patenkinamais pažymiais studijavo mažiau nei pusę kurso teorinės medžiagos.

3 lentelė. Asociacijos rezultatai, kai ieškomi sąryšiai tarp galutinio kurso vertinimo ir skaitytų teorinės medžiagos temų skaičiaus

| Taisyklė | Patikimumas |
|--|-------------|
| Galutinis vertinimas=LABAI GERAI ARBA PUIKIAI⇒Skaitytų teorinės medžiagos temų skaičius=DAUGIAU NEI PUSĖ | 0.92 |
| Galutinis vertinimas=PAKANKAMAI ARBA GERAI⇒Skaitytų teorinės medžiagos temų skaičius=DAUGIAU NEI PUSĖ | 0.89 |
| Galutinis vertinimas=SILPNAI ARBA PATENKINAMAI⇒Skaitytų teorinės medžiagos temų skaičius=MAŽIAU NEI PUSĖ | 0.91 |

Taikant asociacijų paieškos metodą, ieškomi sąryšiai tarp galutinio kurso vertinimo ir atliktų savikontrolės testų skaičiaus. E. mokymosi kurse numatyta 10 savikontrolės testų. Parengiant duomenis asociacijų paieškos metodo taikymui, atributo – Atliktų savikontrolės testų skaičius reikšmės suskirstomos į tris kategorijas: jei atliktų savikontrolės testų skaičius mažesnis nei 5 priskiriama reikšmė – MAŽIAU NEI PUSĖ; jei savikontrolės testų skaičius yra lygus 5, 6 priskiriamas atitikmuo – PUSĖ; jei atliktų savikontrolės testų skaičius yra didesnis nei 6 priskiriama reikšmė – DAUGIAU NEI PUSĖ. Kaip ir pirmuoju atveju asociacijų paieškai taikomas Apriori algoritmas su nustatyta minimalia asociacijos taisyklių patikimumo reikšme - 0.8. Gautos asociacijos (4 lentelė) rodo, kad studentai, kurie kurso studijas baigė silpnais arba patenkinamais įvertinimais, atliko mažiau nei pusę savikontrolės testų, puikiai arba labai gerai įvertinti studentai žinių savikontrolėi ir įtvirtinimui aktyviai naudojo savikontrolės testus.

4 lentelė. Asociacijos rezultatai, kai ieškomi sąryšiai tarp galutinio kurso vertinimo ir atliktų savikontrolės testų skaičiaus

| Taisyklė | Patikimumas |
|---|-------------|
| Galutinis vertinimas=LABAI GERAI ARBA PUIKIAI⇒Atliktų savikontrolės testų skaičius=DAUGIAU NEI PUSĖ | 1 |
| Galutinis vertinimas=PAKANKAMAI ARBA GERAI⇒Atliktų savikontrolės testų skaičius=DAUGIAU NEI PUSĖ | 0.85 |
| Galutinis vertinimas=SILPNAI ARBA PATENKINAMAI⇒Atliktų savikontrolės testų skaičius=MAŽIAU NEI PUSĖ | 0.84 |

Taikant asociacijų paieškos metodą, ieškomi sąryšiai tarp galutinio kurso vertinimo ir maksimalių savikontrolės testų vertinimų vidurkio. Parengiant duomenis asociacijų paieškos metodo taikymui, atributo – Maksimalių savikontrolės testų vertinimų vidurkis reikšmės suskirstomos į tris kategorijas: ≤ 7 ; > 7 ir < 10 ; $= 10$. Gautos asociacijos pateiktos 5 lentelėje. Puikiai arba labai gerai kurso studijas pabaigę studentai savikontrolės testus kartojo tiek kartų kol gavo maksimalius savikontrolės testų vertinimus. Silpnai arba patenkinamai įvertintiems studentams nepavyko visus savikontrolės testus atlikti maksimaliam vertinimui, šios studentų grupės savikontrolės testų maksimalių vertinimų vidurkis yra mažesnis nei 7.

5 lentelė. Asociacijos rezultatai, kai ieškomi sąryšiai tarp galutinio kurso vertinimo ir maksimalių savikontrolės testų vertinimų vidurkio

| Taisyklė | Patikimumas |
|--|-------------|
| Galutinis vertinimas=LABAI GERAI ARBA PUIKIAI⇒Savikontrolės testų maksimalių vertinimų vidurkis =10 | 0.92 |
| Galutinis vertinimas= SILPNAI ARBA PATENKINAMAI⇒Savikontrolės testų maksimalių vertinimų vidurkis $\Rightarrow \leq 7$ | 0.97 |

Analizuojant sąryšius tarp galutinio kurso vertinimo ir praktinių darbų įvertinimų vidurkių, buvo taikomas asociacijų paieškos metodas. Praktinių darbų įvertinimų vidurkis suskaidytas į tris kategorijas: ≤ 7 ; > 7 ir < 10 ; $= 10$. Gautos asociacijos pateiktos 6 lentelėje. Praktiniai darbai atliekami praktinių užsiėmimų metu. Jei studentas praktinį darbą atliko namuose, šį praktinį darbą jis turi atlikti ir apginti iki nurodyto termino. Praktinio darbo vertinimas už kiekvieną pavėluotą savaitę mažinamas 1 balu. Gautos asociacijos rodo, kad puikiai ir labai gerai kurso studijas pabaigę studentai praktinius darbus atliko atsakingai, kruopščiai ir tiksliai. Silpnai ir patenkinamai įvertinti studentai ne visus praktinius darbus spėjo atsiskaityti laiku, juos atliko nepilnai arba nesugebėjo atliktų praktinių darbų tinkamai apginti.

6 lentelė. Asociacijos rezultatai, kai ieškomi sąryšiai tarp galutinio kurso vertinimo ir praktinių darbų vertinimų vidurkio

| Taisyklė | Patikimumas |
|---|-------------|
| Galutinis vertinimas=LABAI GERAI ARBA PUIKIAI⇒Praktinių darbų įvertinimų vidurkis=10 | 1 |
| Galutinis vertinimas=PAKANKAMAI IR GERAI⇒Praktinių darbų įvertinimų vidurkis =>7 ir <10 | 0.93 |
| Galutinis vertinimas=SILPNAI IR PATENKINAMAI⇒Praktinių darbų įvertinimų vidurkis =><=7 | 0.96 |

Tęsiant tyrimą buvo ieškoma asociacijų tarp galutinio ir savarankiško darbo vertinimo. Atributo – Savarankiško darbo vertinimas reikšmės diskretizuojamos, taikant tas pačias taisykles kaip ir skaidant į kategorijas galutinius kurso vertinimus. Asociacijų paieškos rezultatai pateikti 7 lentelėje. Gautos asociacijos rodo, kad studentai, kurie baigė kurso studijas labai gerais arba puikiais pažymiais tokiais pat pažymiais apsigynė ir savarankišką darbą. Silpnai arba patenkinamai įvertinti studentai savarankišką darbą atliko pakankamais arba gerais pažymiais.

7 lentelė. Asociacijos rezultatai, kai ieškomi sąryšiai tarp galutinio kurso vertinimo ir savarankiško darbo vertinimo

| Taisyklė | Patikimumas |
|---|-------------|
| Galutinis vertinimas=LABAI GERAI ARBA PUIKIAI⇒Savarankiško darbo vertinimas= LABAI GERAI ARBA PUIKIAI | 1 |
| Galutinis vertinimas=SILPNAI ARBA PATENKINAMAI⇒Savarankiško darbo vertinimas=PAKANKAMAI ARBA GERAI | 0.83 |

Apibendrinant tyrimo rezultatus galima teigti, kad studentų mokymosi rezultatai tiesiogiai priklauso nuo to kaip aktyviai besimokantieji studijavo teorinę medžiagą, įgytų teorinių žinių savikontrolei ir įtvirtinimui naudojo savikontrolės testus, kaip atsakingai ir kruopščiai atliko praktinius darbus. Tobulinant analizuojamą e. mokymosi kursą būtina e. mokymosi kursui nustatyti mokymosi eigos valdymą, kuris kontroliuotų, kad baigiamasis kurso egzaminas būtų prieinamas tik tiems studentams, kurie analizavo visas teorinės medžiagos temas, atliko visus savikontrolės testus bei praktinius darbus. Mokymosi eigos kontrolė užtikrintų aukštesnius studentų mokymosi rezultatus.

Išvados

Atlikus duomenų gavybos tyrimų e. mokymosi srityje analizę, nustatyta, kad duomenų gavybos metodai taikomi studentų mokymosi efektyvumo vertinimui, besimokančiųjų problemų diagnozavimui, besimokančiųjų mokymosi stiliaus ir būdo nustatymui. Duomenų gavybos metodais paremta duomenų analizė svarbi, pritaikant e. mokymosi turinį pagal besimokančiojo žinias ir gebėjimus, nustatant mokymosi eigą, kuriant įrankius, skirtus sumažinti studentų plagiatą ir kitokių nesąžiningų veiklų atsiradimo galimybes.

E. mokymosi kurso duomenų analizei atlikti atrinkti e. mokymosi sistemos duomenų bazėje saugomi e. mokymosi kurso duomenys. Atliekant tyrimą, analizuoti e. mokymosi kurso savikontrolės testų, kontrolinių testų, praktinių darbų, kontrolinių darbų, savarankiško darbo duomenys, tirta kaip besimokančiųjų vertinimus įtakoja teorinės mokomosios medžiagos nagrinėjimas. E. mokymosi duomenų analizei atlikti pasirinkti klasterizavimo ir asociacijų paieškos metodai. Duomenų klasterizavimas atliktas, naudojant k vidurkių metodą, asociacijų paieškai taikytas Apriori algoritmas.

Klasterizacijos metu gauti statistiniai duomenys parodė, kad studentai, kurių kurso žinios ir gebėjimai buvo įvertinti puikiais, labai gerais arba gerais pažymiais, aktyviau studijavo teorinę mokomąją medžiagą, teorinių žinių savikontrolei ir įtvirtinimui naudojo savikontrolės testus, geriau atliko kontrolinius testus, kontrolinius darbus, egzaminą užduotis, lyginant su pakankamai, patenkinamai arba silpnai įvertintais studentais.

Taikant asociacijų paieškos metodą nustatyta, kad studentai, kurių kurso žinios ir gebėjimai buvo įvertinti patenkinamai arba silpnai, skaitė tik mažiau nei pusės teorinių paskaitų medžiagą, atliko mažiau nei pusę savikontrolės testų, atliktų savikontrolės testų maksimalių įvertinimų ir praktinių darbų įvertinimų vidurkis buvo mažesnis nei septyni.

Atlikus e. mokymosi kurso duomenų analizę, nuspręsta pakeisti e. mokymosi kursą, nustatant mokymosi eigos kontrolę, kuri valdytų besimokančiųjų mokymosi eigą, kontroliuotų kad e. mokymosi kurso baigiamasis egzaminas būtų pasiekiamas tik tiems besimokantiejiems, kurie atliko visus savikontrolės testus ir gavo šių savikontrolės testų maksimalius vertinimus, studijavo visas e. mokymosi kurso teorinės medžiagos temas.

Literatūros sąrašas

1. Anduela, L. M (2011). Analyzing E-Learning Systems Using Educational Data Mining Techniques. *Journal of Social Sciences*, 2(3), p. 403-420.
2. Delgado Calvo-Flores, M., Gibaja Galindo, E. (2006). Predicting students' marks from Moodle logs using neural network models. *Current developments in Technology-Assisted Education*, p. 586-590.
3. Klasnja-Milicevic, A., Vesin, B., Ivanovic, M., Budimac, Z. (2011). Elearning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, 56, p. 84-89.
4. Kotzer, S., Elran, Y.(2012). Learning and teaching with Moodle-based E-learning environments, combining learning skills and content in the fields of Math and Science & Technology. *Moodle Research Conference*, p. 122-131.
5. Liu, F. J., Shih, B. J. (2007). Learning Activity based E-learning Material Recommendation System. *Ninth IEEE International Symposium on Multimedia*, p. 343-348.
6. Merceron, A., Yacef, K. (2005). Educational Data Mining: a Case Study. *Proceedings of the 12th International Conference on Artificial Intelligence in Education AIED*, p.20-26.
7. Namdeo, V., Singh, A., Singh, D., Jain, R.C. (2010). Result Analysis Using Classification Techniques. *International Journal of Computer Applications*, 1(22), p. 50-55.
8. Paulauskienė, K.; Kurasova, O. (2013) Duomenų tyrybos sistemų galimybių tyrimas įvairių apimčių duomenims analizuoti. *Informacijos mokslai*, 65, p. 85-95.
9. Preidys, S., Sakalauskas, L. (2010). Analysis of Students' Study Activities in Virtual Learning Environments Using Data Mining Methods. *Technological and Economic Development of Economy*, 16(1) p. 94-108.
10. Preidys, S., Sakalauskas, L. (2012). Išmaniųjų modulių integravimo į VMA Moodle galimybės: nuo teorijos prie praktikos. *Mokslų taikomųjų tyrimų įtaka šiuolaikinių studijų kokybei*.
11. Preidys, S., Sakalauskas, L. (2011). Nuotolinio mokymosi stilių personalizavimas. *Informacijos mokslai*, 56, p. 42-49.
12. Preidys, S., Žilinskienė, I. (2012). Nuotolinio mokymosi kurso personalizavimo modelis mokymosi veiklų atžvilgiu. *Elektroninis mokymasis, informacija ir komunikacija: teorija ir praktika*, p. 29-37.
13. Romero, C., Ventura, S., Garcia, E. (2008). Data mining in course management systems Moodle case study and tutorial. *Computers & Education*, 51, p. 368-384.
14. Savulionienė, L., Sakalauskas, L.(2011). Statistinis dažnų posekių paieškos algoritmas. *Informacijos mokslai*, p. 126-143.
15. Sharma, N., Bajpai, A., Litoriya, R. (2012). Comparison the various clustering algorithms of weka tools. *International Journal of Emerging Technology and Advanced Engineering*, 2(5), p.73-80.
16. Scherbinin, V., Butakov, S. (2008). Plagiarism detection: the tool and the case study. *International Conference e-Learning*, p. 304-310.
17. Shen, L., Wang, M., Shen, R. (2009). Affective e-Learning: using "emotional" data to improve learning in pervasive learning environment. *Educational Technology & Society*, 12(2), p. 176-189.
18. Stefanovič, P.; Kurasava, O. (2013). Tekstinių dokumentų panašumų paieška naudojant saviorganizuojančius neuroninius tinklus ir k vidurkių metodą. *Informacijos mokslai*, 65, p. 24-33.

Summary

ANALYSIS OF E-LEARNING COURSES DATA USING DATA MINING TECHNIQUES

Data mining means searching for certain patterns within large sets of data, which creates a lot of possibilities for decision makers. By analysing those patterns, better decisions can be made in order to improve e-learning process. The research interest in using data mining in e-learning is constantly increasing. According to L. Shen, M. Wang, R. Shen the database of e-learning system includes much useful information which can be effectively used for the improvement of e-learning process. Authors emphasise that due to the vast quantities of data these systems can generate daily, it is very difficult to analyse this data manually and a very promising approach towards this analysis objective is the use of data mining techniques.

The purpose of this research is to analyze the course data using data mining techniques in order to provide course development trends to more efficient and higher quality studies.

Research methods: the analysis of scientific literature, data analysis using data mining techniques.

In this research, the data of Database design course were analysed. All data necessary for this research were transferred from Moodle database. There were created the procedures to perform calculations. Selected data set consists of 11 fields. The sources of selected data are analysable course records: the number of performed self-test, time spent for self-test, average of test grades, average of control works grades, final mark and so on.

WEKA tool was used to perform data analysis using data mining techniques. For data analysis was chosen descriptive data mining model's techniques: clustering and association rules. For the clustering in this work was used k means algorithm, for the finding association rules was used Apriori algorithm. After clustering technique were obtained statistical data which showed that the students with excellent, very good or good final grade actively studied theoretical training materials, performed self-tests, better performed control tests, control works, exam tasks. Using the association technique was obtained that the students who get minimal (the worst) grades did the course out of continuity; they gave little attention to the study of theoretic materials, self-tests; they passed the final exam without gaining necessary theoretic knowledge. The students who got good marks successfully completed all sections of theoretic materials passed self-tests and passed the final exam only on completing all course activities. Data analysis showed the disadvantages

of course structure. It was resolved that Database design course should be modified introducing the learning process control.

Keywords: e-learning system, data mining, e-learning course, clustering, association technique.