

ŠIAULIŲ UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS KATEDRA

Vidas Buivydas
Informatikos (mokytojas) specialybės II kurso neakivaizdinio skyriaus studentas

NUOTOLINIŲ STUDIJŲ DUOMENŲ GAVYBA

DATA MINING IN E-LEARNING

MAGISTRO DARBAS

Darbo vadovas:
prof. habil. dr. L. Sakalauskas

Recenzentas:
prof. habil. dr. G. Kulvietis

Šiauliai, 2011

„Tvirtinu, jog darbe pateikta medžiaga nėra plagijuota ir paruošta naudojant literatūros sąrašę pateiktus informacinius šaltinius bei savo tyrimų duomenis.“

Darbo autorius _____

(vardas, pavardė, parašas)

Darbo tikslai ir uždaviniai

Tikslas

Sudaryti nuotolinių studijų duomenų gavybos metodiką ir pritaikyti ją „Moodle“ sistemos atveju.

Uždaviniai

- Apžvelgti ir palyginti nuotolinių studijų duomenų gavybą.
- Ištirti duomenų gavybos metodų ir algoritmų tinkamumą nuotolinių studijų duomenų analizei.
- Atlikti duomenų gavybos metodų analizę.
- Ištirti WEKA programinės įrangos pritaikymą duomenų gavybai.
- Sudaryti duomenų gavybos procesą.
- Praktiškai realizuoti duomenų gavybą „Moodle“ sistemoje.

Darbo vadovas _____
(vardas, pavardė, parašas)

TURINYS

1. ĮVADAS.....	6
2. TEORINIS PAGRINDIMAS	8
2.1. DUOMENŲ GAVYBA ELEKTRONINĖSE MOKYMOSI SISTEMOSE.....	8
2.2. DUOMENŲ GAVYBOS METODŲ APŽVALGA.....	10
2.3. KRYŽMINIS PATIKRINIMAS	11
2.4. DUOMENŲ GAVYBA „MOODLE“ SISTEMOJE.....	13
2.5. WEKA APŽVALGA	18
3. PROJEKTINĖ DALIS.	19
3.1. „MOODLE“ DUOMENŲ PIRMINIS APDOROJIMAS	19
4. EKSPERIMENTINĖ DALIS	22
5. IŠVADOS.....	35
6. LITERATŪRA.....	36
7. SANTRUMPŲ IR TERMINŲ ŽODYNAS	37
8. PRIEDAI	38
8.1. PRIEDAS 1. DUOMENŲ ANALZIĖ.....	38
8.2. PRIEDAS 2. PROGRAMOS KODAS.....	43

PAVEIKSLAI

Pav. 1 Duomenų gavybos procesas	9
Pav. 2 Tamsesnieji laukeliai žymi duomenis naudojamus mokymui, šviesesni – patvirtinimui.....	11
Pav. 3 Moodle sistemos testo rezultatų statistikos langas	14
Pav. 4 Studentų prisijungimo statistika	15
Pav. 5 Weka klasterizavimo pavyzdys	16
Pav. 6 Keel analizės rezultatai	17
Pav. 7 Moodle naudojimosi žurnalas.....	19
Pav. 10 Moodle mdl_log lentelės struktūra	22
Pav. 11 Apibendrinančios lentelės mdl_dm_action_summary pagrindinių laukų struktūra	22
Pav. 12 Klasterizavimo vizualizacija.....	30

LENTELĖS

Lentelė 1 Patikrinimo metodų palyginimas.....	13
Lentelė 2 Moodle duomenų bazės svarbiausios lentelės	20
Lentelė 3 apibendrinti duomenys.....	21
Lentelė 4 Kurso „13“ pradinis ARFF duomenų failas.....	23
Lentelė 5 Kurso „13“ paruoštas klasterizavimui ARFF duomenų failas.....	25
Lentelė 6 Kurso „13“ klasterizavimo rezultatai.....	27
Lentelė 7 Kurso „13“ studentų priskyrimas klasteriams	30
Lentelė 8 Kurso „13“ klasterizavimo rezultatai (5 klasteriai)	30
Lentelė 9 Kurso „10“ klasterizavimas, ieškant 4 klasterių	33

1. ĮVADAS

Švietimo duomenų gavyba yra nauja disciplina, susijusi su besivystančiais tyrimo metodais skirtais ištirti unikaliems duomenų tipams, glūdintiems švietimo kontekste. Šis darbas skirtas duomenų gavybos tyrimui bei mokymosi valdymo sistemos „Moodle“ [5] apžvalgai. Darbe bus aprašoma elektroninės mokymo sistemos duomenų gavyba, apžvelgiami pagrindiniai duomenų gavybos metodai: statistiniai duomenys, vizualizacija, klasifikavimas ir klasterizavimas.

Greitai plečiantis internetui, augant jo teikiamų paslaugų populiarumui, internetas užima vis didesnę vaidmenį duomenų rinkimo bei dalinimosi srityse. Dauguma organizacijų ir korporacijų tinkle pateikia informaciją bei paslaugas, kaip kad automatizuotą pagalbą klientams, internetinę prekybą bei daugybę kitų išteklių ir pritaikomųjų paslaugų. Internete greit plinta ir palaiptai tampa įprasta: tinklinės paslaugos, elektroninės prekybos aplinkos, nuotoliniai mokymai, realaus laiko bendradarbiavimas, naujienų sklaidimas.

Internetas tampa neatskiriamu kasdieninės veiklos tarpininku, pradedant vaikais, besidalinančiais su draugais muzikiniais failais, baigiant senoliais, gaunančiais nuotraukas bei žinutes iš anūkų. Universitetiniuose tinklapiuose būdinga talpinti kursų bei užsiėmimų informaciją ir resursus, netgi jei ta informacija yra perteikiama tradicinių užsiėmimų metu. Nėra nuostabu, kad internetas naudojamas pažangioms nuotolinio mokymo sistemoms. Nuotoliniam mokymui buvo gana greit pritaikyta tinklu paremta technologija ir naudojama žinių pasiekimui bei dalinimuisi.

Daugiausia paplitusios nuotolinio mokymo aplinkos: Moodle, Virtual-U bei Web-CT, BlackBoard, turi gausybę mokymosi bei informacijos perteikimo modulių – kurso turinio pristatymo įrankiai, sinchroninės ir asinchroninės konferencijų sistemos, apklausų moduliai, virtualūs darbalaukiai, įvertinimų pateikimo sistemos, žurnalai, priskyrimų pateikimo komponentai.

Virtualioje klasėje mokytojai pateikia tekstus, multimedijas, simuliacijas bei diskusijas. Nuotolinio kurso studentai skatinami sekti išteklius bei dalyvauti klasės veikloje. Kita vertus mokytojams yra pakankamai sunku (laiko atžvilgiu) kruopščiai stebėti ir vertinti visus mokinius daugybėje mokymo aplinkų. Mokymo proceso metu yra pakankamai sudėtinga vertinti kurso struktūrą bei efektingumą.

Mokytojui būtų labai pravartu, jei mokymo sistema automatiškai orientuotų mokytojo veiklas bei parinktų veiklas ar resursus kurie pagerintų mokymą. Kol kas tokių įrankių nėra sukurta.

Kita vertus elektroninės komercijos srityje galima išvelgti automatizuotą prekių pasiūlą, kaip kad Amazon.com rekomenduoja pirkti knygas remdamiesi jau įvykdytų pirkimų istorija.

E–mokyme mokymo sesija gali apimti daugybę priėjimo sesijų. Norint sužinoti tikslus užklausų rezultatus reikia apjungti daug priėjimo sesijų pasiskirsčiusių ilgais laiko intervalais. Žinant tai, jog elektroninėje prekyboje tikslai yra maždaug aiškūs (pavyzdžiui – vartojimo skatinimas ar lojalumo didinimas), e–mokymėsi tikslai nėra visiškai aiškūs, sunkiai apibrėžiami.

Tinklinės kursų pristatymo sistemos, kaip ir bet kuris tinklapis ar aplikacija – remiasi serveriais norėdami gauti prieigą prie resursų. Kiekviena užklausa, kurią priima serveris yra įrašoma žurnale, pažymint užklauso laiką bei resursus užklausiai. Tinklo žurnalas saugo neapdorotus duomenis apie besimokančiuosius. Norint apdoroti tokius duomenis ir gauti reikiamą informaciją reikalingas duomenų valymas ir transformavimas-paruošimas duomenų gavybai [3].

2. TEORINIS PAGRINDIMAS

2.1. DUOMENŲ GAVYBA ELEKTRONINĖSE MOKYMOSEI SISTEMOSE.

Tradicinis elektroninės mokymosi sistemos vystymas yra sudėtingas ir daug laiko reikalaujantis darbas, kurio metu kūrėjas (dažniausiai kurso dėstytojas) pats turi parinkti medžiagą, kuri bus rodoma, nustatyti el. mokymosi sistemos struktūrą taip parinkti labiausiai tinkamas sritis bei elementus kiekvienam kursų vartotojų tipui [7,13]. Esant gausybei galimų (e-mokymosi sistemos) struktūros bei architektūros variantų, neįmanoma parinkti vieno šablono visiems atvejams. Tokiu atveju neišvengiamai reiktų sekti studentų aktyvumą bei, atsižvelgiant į tai, koreguoti: kurso turinį, struktūrą bei navigaciją.

Duomenų gavybos metodų pritaikymas žinių paieškai universiteto srityje gali padėti suprasti vartotojų poreikius [4].

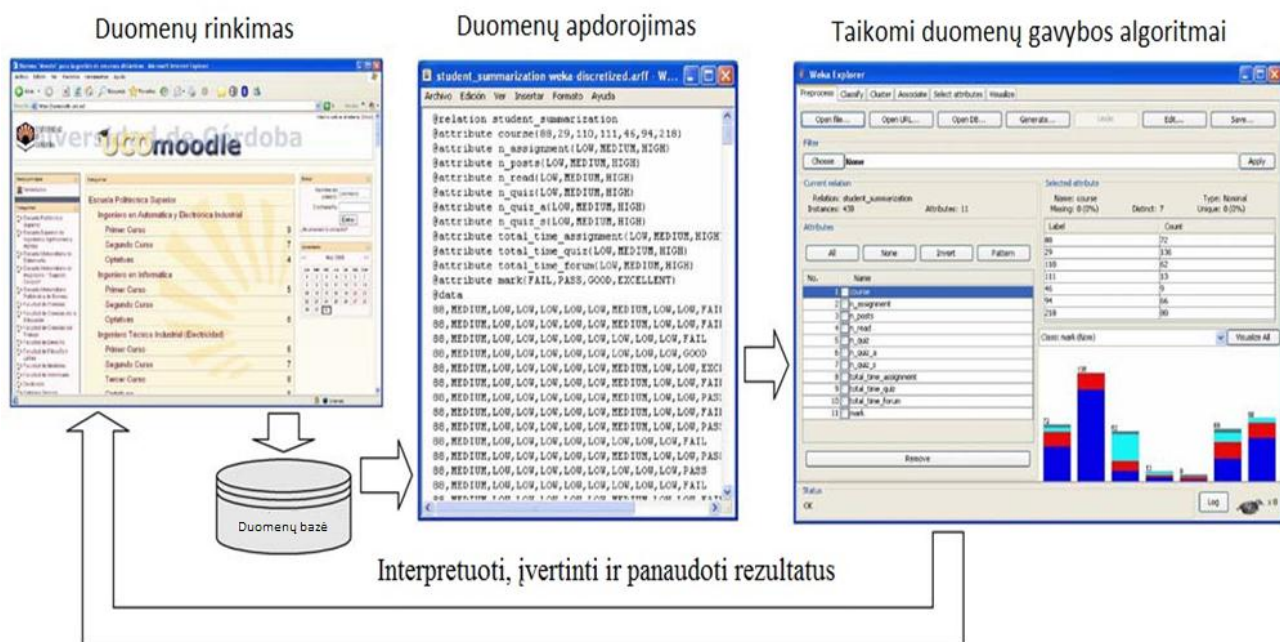
Siekiant palengvinti šį procesą, reiktų naudoti duomenų analizės metodus ir įrankius, sekti moksleivių elgesį ir padėti mokytojams aptikti galimas klaidas, trūkumus ir galimus patobulinimus.

Tradiciskai, e-mokymosi duomenų gavyba pradedama vartotojo užklausa yra lydima duomenų paieškos. Toks metodas pasiteisina vidutiniškos apimties faktorių bei duomenų kiekio atveju, tačiau apimtims didėjant tai nepasiteisina, norint sužinoti daugiau sudėtingų modelių, kurie yra susiję su įvairiais duomenimis, darosi vis sudėtingiau. Alternatyva tradicinei duomenų analizei: duomenų gavyba yra naudojama kaip indukcinis metodas automatiškai atrasti paslėptą informaciją tarp duomenų.

Duomenų gavybos taikymas e-mokymosi sistemose yra pasikartojantis ciklas, kuriame gauti duomenys turėtų įeiti į sistemų bei vedlių ciklą, palengvinantį ir stiprinantį mokymąsi kaip visumą, ne tik verčiant duomenis į žinias, bet taip pat išgautų duomenų filtravimą idant būtų priimti atitinkami sprendimai. E-mokymosi duomenų gavybos procesas susideda iš keturių veiksmų (1 pav.):

1. Duomenų rinkimas. LMS sistemą naudoja studentai, naudojimo ir sąveikos informacija yra saugoma duomenų bazėje. (Moodle sistema).
2. Duomenų apdorojimas. Duomenys yra valomi ir transformuojami į atitinkamą formatą. Siekiant apdoroti „Moodle“ duomenis, mes galime naudoti duomenų bazės administratorius įrankius.

3. Taikyti duomenų gavybą. Duomenų gavybos algoritmai taikomi siekiant sukurti ir vykdyti modelį, kuris nustato ir apibendrina palūkanų vartotojo žinias (mokytojų, studentų, administratorių ir pan.). Siekiant tai padaryti, galime naudoti bendruosius ar konkrečius duomenų gavybos įrankius.
4. Interpretuoti, įvertinti ir panaudoti rezultatus. Rezultatai, gauti tyrimo metu, aiškinami ir naudojami tolesniems mokytojo veiksmams. Mokytojas gali naudoti informaciją, aptiktą duomenų gavybos metu, kad galėtų priimti sprendimus „Moodle“ sistemos gerinimui [1,8].



Pav. 1 Duomenų gavybos procesas

Kaip matome, duomenų gavybos taikymas e-mokymesi ne daug skiriasi nuo kitų sričių duomenų gavybos. Tačiau yra keletas svarbių sričių, į kurias reiktų atkreipti dėmesį:

- *Duomenys.* Kitų interneto sistemų naudojami duomenys paprastai yra paprastas serverio prieigos žurnalas, bet e-mokymosi sistemų duomenų bazės saugo daug daugiau informacijos apie mokinių sąveiką. LMS sistemos gali įrašyti bet kurias studentų veiklas, kaip antai skaitymas, rašymas, testų sprendimas, įvairių užduočių sprendimas ir netgi bendravimas tarpusavyje.

- *Tikslas.* Duomenų gavybos tikslas kiekvienoje paraiškoje yra skirtingas ir gali būti daugiau ar mažiau tikslus. E-mokymosi sistemos gerinimas apima: mokymosi sistemos gerinimo bei studentų nukreipimo tinkama linkme parinkimo.
- *Metodai.* Švietimo sistema turi ypatingų savybių, kurios reikalauja skirtingo požiūrio į duomenų gavybą. Kai kurios technikos gali būti taikomos tiesiogiai, kitos negali ir turi būti adaptuotos.

Galima išskirti šiuos internetinių puslapių logų apdorojimo žingsnius:

- Nereikalingų įrašų pašalinimas
- Aktyvių sesijų identifikavimas
- Prisijungimo įrašų susiejimas su mokymosi veiksmis
- Pilna navigacija tarp puslapių
- Integruoti kitą informaciją apie mokinius ir grupes

2.2. DUOMENŲ GAVYBOS METODŲ APŽVALGA

Galime išskirti keturis pagrindinius nuotolinių studijų duomenų gavybos metodus [12,9].

Klasifikavimas - duomenų gavybos metodas, naudojamas prognozuoti duomenų narystės grupėje atvejus. Pavyzdžiui, galbūt norėsite naudoti klasifikaciją prognozuoti, ar tam tikrą dieną bus "saulėta", "lis lietus" ar "snigs sniegas". Populiariausi klasifikavimo metodai apima sprendimų medžius ir neuroninius tinklus.

Asociacijos taisyklių radimas yra gan populiarus ir ištirtas metodas, kurio pagalba randami ryšiai tarp kintamųjų didelėse duomenų bazėse.

Klasterizavimo metu ieškomi elementai, kurie gali būti apjungti grupėje.

K-means algoritmas dar vadinamas evoliucionuojančiu algoritmu, kuris apjungia stebėjimo duomenis į K grupes, kur K pažymi įvesties reikšmę. Tuomet stebėjimo duomenys priskiriami klasteriams pagal jų bendrus požymius artimiausius to klasterio vidurkiui.

Klasterio vidurkis perskaičiuojamas ir procesas kartojamas nuo pradžių. Algoritmo veikimas:

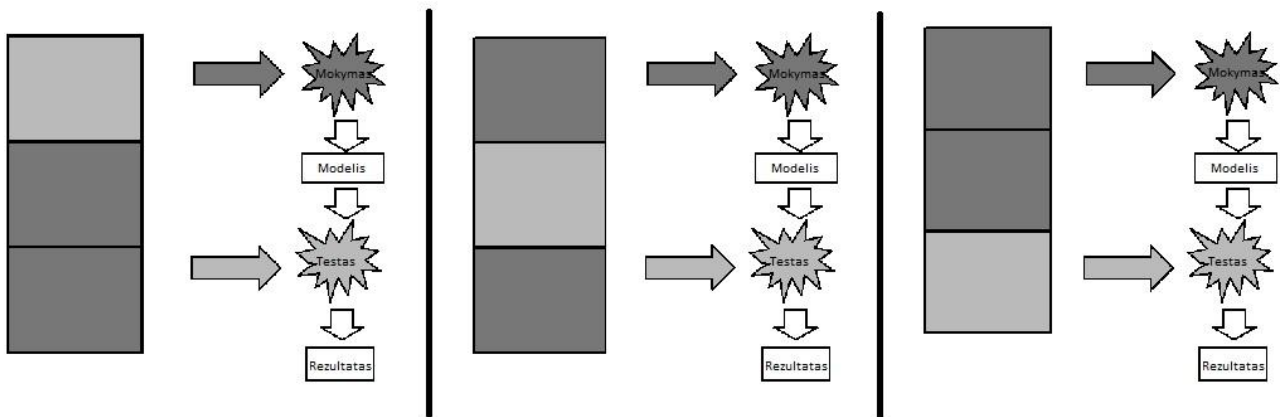
1. Algoritmas atsitiktiniu būdu atrenka K taškų, kurie tampa klasterių centrais.
2. Kiekvienas duomenų taškas priskiriamas artimiausiam klasteriui pagal euklidinį atstumą.
3. Perskaičiuojami kiekvieno klasterio centrai pagal klasterio taškų vidurkį.
4. Kartojami 2 ir 3 žingsniai tol kol nusistovi klasteriai. Konvergencija gali būti apibrėžiama skirtingai, priklausomai nuo įgyvendinimo, tačiau paprastai ji reiškia, kad klasteriai nebesikeičia arba keičiasi nežymiai.

Skaitinio prognozavimo metu rezultatas nėra diskrečioji reikšmė, bet skaitinė reikšmė.

2.3. KRYŽMINIS PATIKRINIMAS

Kryžminis patvirtinimas (Cross-Validation) - tai statistinis metodas, skirtas įvertinti ir palyginti mokymosi algoritmus, dalijant duomenis į du segmentus: „mokyti iš, ar mokyti“ modelį bei modelio patvirtinimo segmentą [6]. Įprastinio kryžminio tikrinimo metu mokymo ir patvirtinimo rinkiniai turi persipinti vienas po kito einančiais etapais, kad kiekvienas duomenų punktas būtų patvirtintas. K-klostė (K-fold) laikoma pagrindine kryžminio tikrinimo forma. Kitos kryžminio tikrinimo formos yra k-klostės kryžminio tikrinimo atvejai arba pasikartojantys k-klostės kryžminio tikrinimo atvejai.

K-klostės kryžminio tikrinimo metu duomenys yra suskirstomi į „k“ vienodų (ar beveik vienodų) dydžių segmentus. Vėliau mokymo ir patvirtinimo „k“ iteracijos yra atliekamos taip, kad per kiekvieną iteraciją iš duomenų imamos skirtingos klostės patikrinimui, kuomet likusios „k-1“ klostės naudojamos mokymui. Paveikslėlyje pavaizduotas k-klostės kryžminis tikrinimas, kuomet $k = 3$.



Pav. 2 Tamsesnieji laukeliai žymi duomenis naudojamus mokymui, šviesesni – patvirtinimui

Duomenų gavyboje labiausiai paplitęs dešimties klosčių ($k = 10$) kryžminis patikrinimas.

Kryžminis patikrinimas naudojamas įvertinti ar palyginti mokymosi algoritmus: kiekvienoje iteracijoje vienas ar daugiau mokymo algoritmų naudoja $k-1$ duomenų klosčių idant išmokti vieną ar daugiau modelių, tada mokymo metu gauti modeliai gauna prognozavimo užklausas apie duomenis patvirtinimo klostėje. Kiekvieno mokymo algoritmo, kiekvienos klostės charakteristika gali būti sekama naudojant tikslinimo metriką (accuracy metric).

Pabaigoje k pavyzdžiai, paimti iš charakteristikos metrikos, tampa prieinami kiekvienam algoritmui. Įvairūs metodai, kaip kad „vidurkio radimo“, gali būti panaudoti imtyje, norint gauti agregatinę dažnį (measure), arba imtys gali būti panaudotos statistinės hipotezės teste, kad parodytų pranašesnę algoritmą.

Yra keli galimi kryžminio tikrinimo tikslai:

- Nustatyti jau išmokto modelio charakteristikai, naudojant esamus duomenis ir vieną algoritmą.
- Palyginti dviejų ar daugiau skirtingų algoritmų charakteristikoms ir rasti geriausią algoritmą esamiems duomenims.

Pakeitimo patikrinimo modelis išmokstamas iš visų esamų duomenų ir tada patikrinamas su tais pačiais duomenimis. Šis patikrinimo procesas naudoja visus prieinamus duomenis, tačiau tai sukelia problemų, nes šis procesas prisitaiko prie konkrečių duomenų. Todėl atsiranda netikslumų, bandant naudoti tą patį algoritmą naujiems nežinomiems duomenims. atsiranda netikslumų.

Tikrinimo atidedant modelio atveju, norint išvengti perpildymo, yra naudojami nepriklausomų testų rinkiniai. Esami duomenys dalinami į du nepersipinančius segmentus: vienas mokymui o kitas tikrinimui. Testavimo dalis neliečiama mokymo metu. Tikrinimo atidedant trūkumas būtų tas, kad procedūra nenaudoja visų duomenų, todėl rezultatas gaunasi stipriai priklausomas nuo mokymo/tikrinimo dalių pasirinkimo. Taipogi informacija gali būti per lengvai arba per sunkiai priskiriama testavimui, todėl galimi netikslūs rezultatai. Reiktų paminėti, jog informacija, parinkta testavimui, irgi gali būti naudinga mokymui. Šie trūkumai gali būti sprendžiami kartojant tikrinimo atidedant algoritmą keletą kartų ir vedant gautų rezultatų vidurkį. Tačiau ir tokiu atveju yra tikimybė, kad informacija naudojama testavimui, niekada nepateks į mokymo dalį. Sprendžiant šią problemą bei siekiant maksimaliai išnaudoti, naudojamas K -klosčių kryžminis patikrinimas.

K -klosčių kryžmininio metodo metu duomenys pirma suskirstomi į k vienodų segmentų. Tada k mokymo iteracijos ir patvirtinimas yra atliekamas tokia tvarka, kuomet kiekvienoje

iteracijoje imama vis kita duomenų juosta tikrinimui, paliekant k-1 juostas mokymui ir duomenys sluoksniuojami į k juostas. Sluoksniavimo metu duomenys pertvarkomi taip, kad kiekviena klostė atspindėtų visą vaizdą.

„Išmetant vieną“ kryžminio patikrinimo metodas sekantis iš K-klosčių kryžminio patikrinimo, kur k žymi kiekvieną atskirą atvejį duomenyse. Kitaip sakant, kiekvienoje iteracijoje imami visi duomenys išskyrus vieną stebimą sritį ir naudojami mokymui, tuomet modelis patikrinamas remiantis ta stebėjimo sritimi.

Siekiant gauti patikimą skaičiavimų charakteristiką, būtina atlikti kuo daugiau skaičiavimų. K-klosčių kryžminio patikrinimo atveju gaunama tik k įvertinimai. Siekiant padidinti skaičiavimų kiekį yra plačiai naudojamas daugkartinis K-klosčių kryžminio patikrinimo metodas. Duomenys yra pertvarkomi bei persluoksniuojami prieš kiekvieną K-klosčių kryžminį patikrinimą.

Lentelė 1 Patikrinimo metodų palyginimas

Patikrinimo metodas	Privalumai	Trūkumai
Pakeitimo patikrinimas.	Paprastas.	Per didelis prisitaikymas prie apmokymo duomenų.
Tikrinimas atidedant.	Nepriklausomas mokymas ir tikrinimas.	Mažesni mokymo ir tikrinimo duomenys, smarkiai kintantis.
K-klosčių kryžminis patikrinimas.	Tikslus charakteristikos vertinimas.	Mažas pavyzdžių charakteristikos įvertinimas, galimas mokymo duomenų susilieėjimas, nuvertėjanti veiklos dispersija.
„Išmetant vieną“ kryžminis patikrinimas.	Nešališkas veiklos vertinimas.	Labai didelė dispersija.
Pasikartojantis K-klosčių kryžminis patikrinimas.	Didelis charakteristikos skaičiavimų kiekis.	Prieš kiekvieną ciklą susilieję mokymo ir tikrinimo duomenys, nuvertinta veiklos dispersija.

2.4. DUOMENŲ GAVYBA „MOODLE“ SISTEMOJE

Statistiniai duomenys

„Moodle“ nėra numatytas pagrindinis statistikos modulis, kuriame mokytojai gali gauti specialias ataskaitas apie išsamius statistinius duomenis: kiekvieno mokinio veiklos duomenys (kiek valandų svetainėje studentas praleido prie konkrečios užduoties ir pan.). Būtų labai naudinga gauti, lankomumo ataskaitas (atiduotų studentų registracija į visus kursus, studentų veikla per visą kursą pas konkrečius mokytojus, mokinių istorijos visi kursai kurių buvo imtasi, praleistas laikas prie jų) ir finansines ataskaitas (pajamos už konkretų laikotarpį, visi mokesčiai, surinkti už visus kursus, kurių ėmėsi studentas, bendros pajamos vienam studentui).

„Moodle“ turi statistinių užklausų ataskaitas, kurios parodo papunktinę analizę (3 pav.).

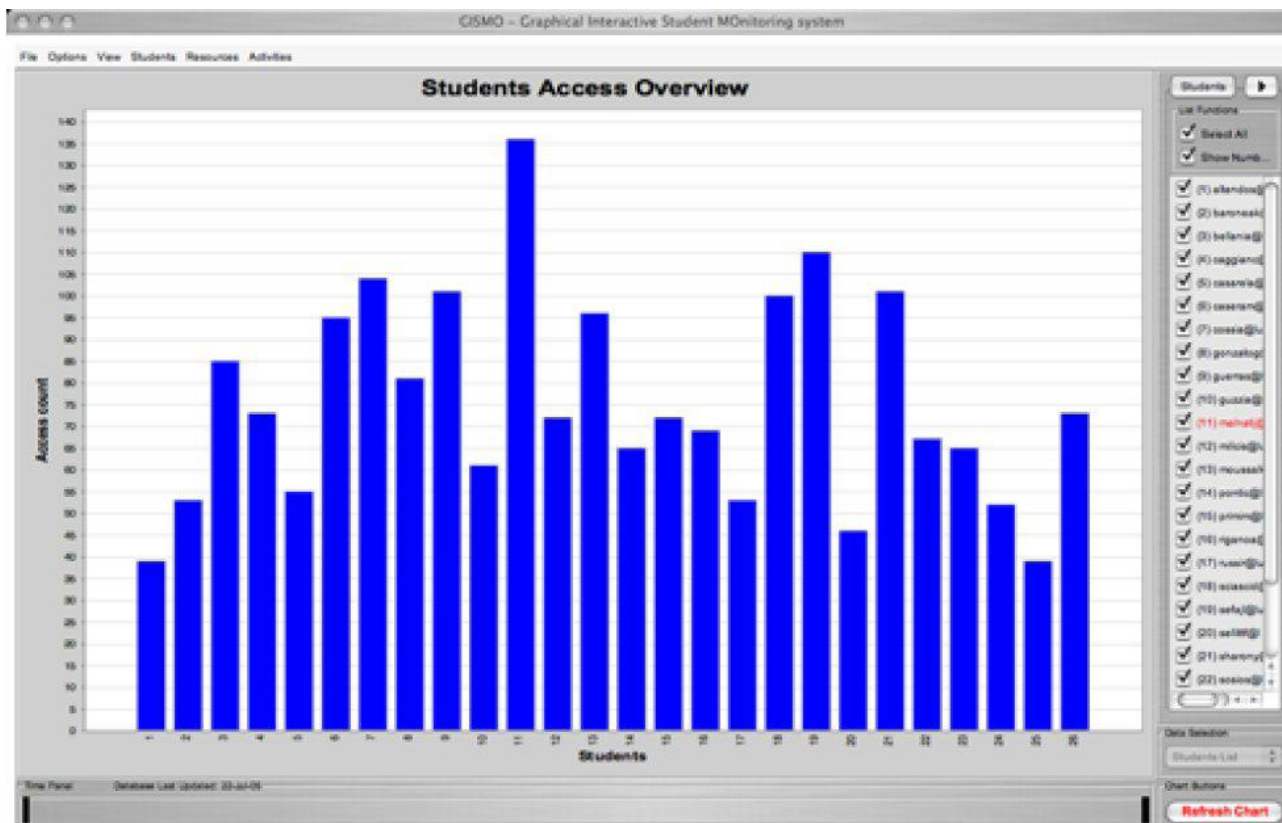
Q#	Question text	Answer's text	partial credit	R. Counts	R.%	% Correct Facility	SD	Disc. Index	Disc. Coeff.
(13)	Pirmas klausimas: paprastas klausimas sudarytas tik iš teksto.	Wrong answer	(0.00)	0/1	(0%)	63%	0.000	0.00	-999.00
		Another wrong answer	(0.00)	0/1	(0%)				
		Correct answer	(1.00)	0/1	(0%)				
		Answer that gives half the credit	(0.50)	1/1	(100%)				
	Sekantis paprastas klausimas	Wrong answer	(0.00)	0/1	(0%)				
		Correct answer	(1.00)	0/1	(0%)				
		Answer that gives half the credit	(0.50)	0/1	(0%)				
		short	(0.00)	1/1	(100%)				
	Pabaigai klausime reikia nurodyti slankaus kablelio skaičių.	65 (64.9..65.1)	(1.00)	1/1	(100%)				
		23.8 (21.8..25.8)	(0.50)	0/1	(0%)				

Pav. 3 Moodle sistemos testo rezultatų statistikos langas

Vizualizacija

„Moodle“ sistemoje nėra numatytų vizualizacijos įrankių, tai nustato tik tekstinė informacija. Galima atsisiųsti ir įdiegti Gismo į „Moodle“ sistemą. Gismo yra grafinis interaktyvių studentų stebėjimo ir kontrolės sistemos įrankis, kuris išgauna stebėjimo duomenis „Moodle“ ir generuoja grafinius atvaizdus, kurie gali būti ištirti kursų dėstytojų (išnagrinėti įvairiais) aspektais studentams).

Gismo numatyta įvairių grafinių atvaizdų ir ataskaitų duomenų vaizdavimas, pavyzdžiui pranešimų grafikai Studentų patekimo į kursą (4 pav.), grafikų ataskaitų apie galimybę atlikti konkrečius studentui žinomus išteklius, grafikus ir t.t.



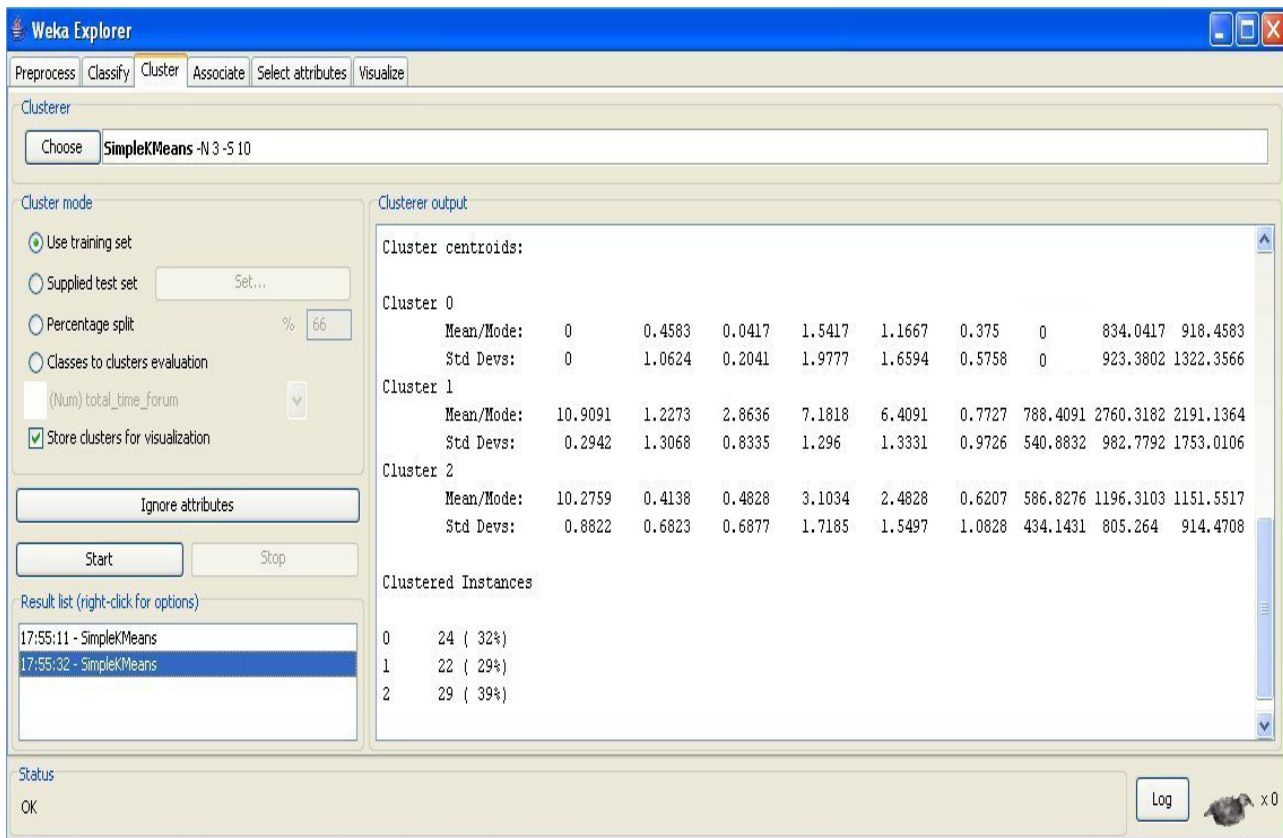
Pav. 4 Studentų prisijungimo statistika

Klasterizavimas

Klasterizavimas yra objektų grupavimo į panašias klases procesas. Straipsnyje pateikiamas klasterizavimo metodo taikymas Šiaulių universiteto Jaunųjų kompiuterininkų mokyklos (ŠU JKM) veiklos analizei. [10]

E-mokymesi klasterizavimas gali būti panaudotas ieškant klasterių su panašiai besimokančiais studentais, ir tada kurtiant panašiai besimokančių studentų grupes ir taikant jiems tas pačias mokymo priemones.

Klasterizavimui galime naudoti kelis „WEKA“ klasterizavimo algoritmus [2,11], pavyzdžiui „KMeans“ (5 pav).



Pav. 5 Weka klasterizavimo pavyzdys

Klasifikavimas

E-mokymesi, klasifikacija gali būti naudojama aptikti galimoms studentų grupėms, kurios turi panašias savybes ir reakcijas į konkrečias pedagogines strategijas, kad prognozuoti mokinių veiklą ir jų galutinis įvertinimus, siekiant nustatyti mokinių piktnaudžiavimo ar žaidimo atvejus.

Klasifikavimui tirti galima panaudoti „Keel“ programinės įrangos 4.5C algoritimą (6 pav.).

```
@decisiontree
if ( n_quiz_a = LOW ) then { mark = "FAIL" }
elseif ( n_quiz_a = MEDIUM ) then {
  if ( total_time_assignment = LOW ) then {
    if ( n_quiz = LOW ) then { mark = "GOOD" }
    elseif ( n_quiz = MEDIUM ) then {
      if ( course = 88 ) then { mark = "GOOD" }
      elseif ( course = 110 ) then {
        if ( n_quiz_s = LOW ) then { mark = "GOOD" }
        elseif ( n_quiz_s = MEDIUM ) then {
          if ( total_time_forum = LOW ) then {
            elseif ( total_time_forum = MEDIUM ) then {
              if ( n_assignment = LOW ) then {
                elseif ( n_assignment = MEDIUM ) then {
                  elseif ( n_assignment = HIGH ) then {
                }
              }
            }
          }
        }
      }
    }
  }
  elseif ( n_quiz = HIGH ) then { mark = "GOOD" }
}
elseif ( total_time_assignment = MEDIUM ) then { mark = "GOOD" }
elseif ( total_time_assignment = HIGH ) then { mark = "GOOD" }
}
elseif ( n_quiz_a = HIGH ) then { mark = "EXCELLENT" }

@TotalNumberOfNodes 7
@NumberOfLeafs 19
```

Pav. 6 Keel analizės rezultatai

2.5. WEKA APŽVALGA

Weka išsiskiria plačiu įrankių rinkiniu, įgalinančiu atlikti duomenų išankstinį apdorojimą, taipogi moderniausiais mokymo algoritmais [12]. Programoje galima greitai išbandyti naujus duomenis su jau sukurtais metodais, atlikti eksperimentinį duomenų gavybos procesą: įvesties duomenų paruošimą, statiškai įvertinti mokymo schemas, vizualizuoti įvesties bei mokymo rezultatų duomenis.

Weka sukurta Waikato universitete Naujojoje Zelandijoje (Waikato environment for Knowledge Analysis). Programa parašyta Java programavimo kalba ir yra platinama GNU GPL licencijos pagrindu. Suderinama su Linux, Windows bei Macintosh operacinėmis sistemomis.

Weka įgalina mokymo algoritmų diegimą taip, kad vartotojas gali lengvai pritaikyti tiriamiems duomenims. Taipogi yra nemažas įrankių, skirtų duomenų transformavimui, pasirinkimas, kaip kad diskretizacijos algoritmas. Darbastalis įgalina pasirinkti standartinius duomenų gavybos metodus: regresiją, klasifikavimą, klasterizavimą ir asociacijos taisyklių radimą.

Įvestis algoritmavimui pateikiama vienoje ARFF formato lentelėje, kuri gali būti nuskaityta iš failo arba sugeneruota duomenų bazėje.

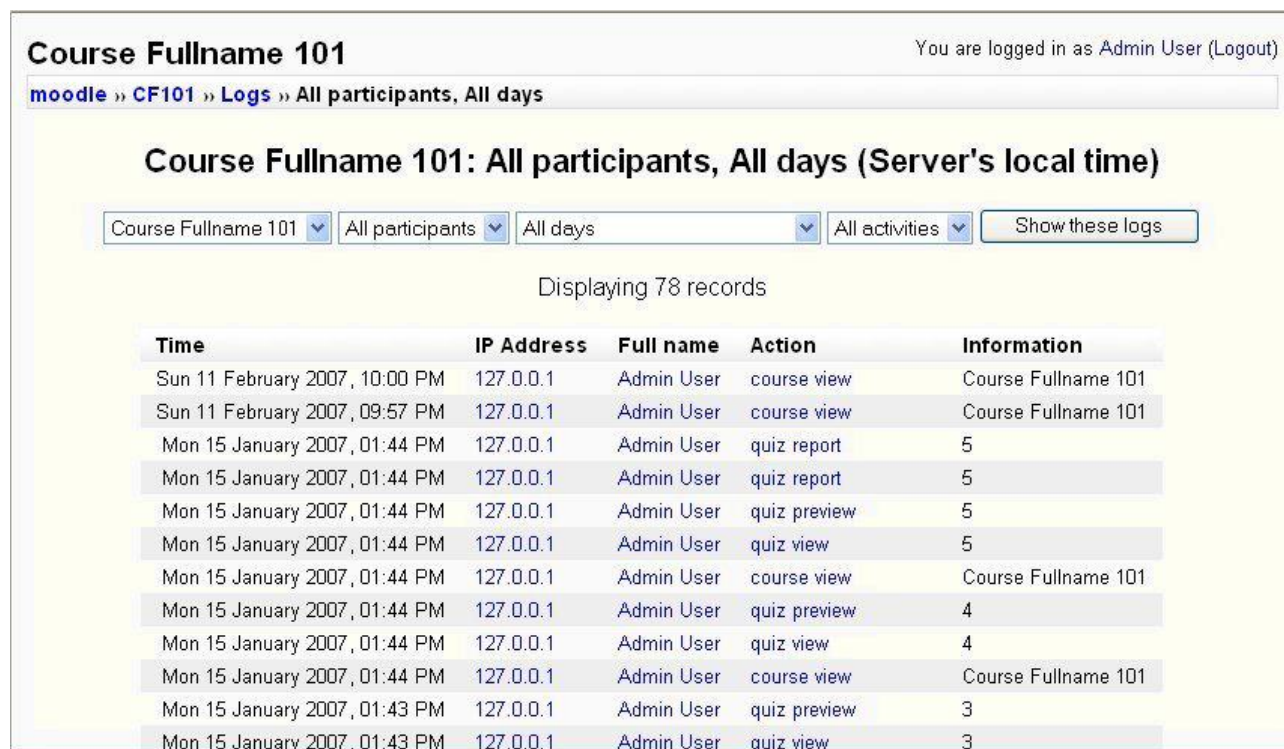
Pirmas Weka naudojimosi būdas: pritaikyti mokymo metodą duomenų grupei ir analizuoti išvestį, siekiant gauti naudingą informaciją. Antras Weka naudojimosi būdas: panaudoti jau išmokus modelius tam, kad sugeneruoti naujų atvejų prognozes. Trečias: pritaikyti keletą skirtingų mokymo būdų ir palyginti charakteristikas.

3. PROJEKTINĖ DALIS.

3.1. „MOODLE“ DUOMENŲ PIRMINIS APDOROJIMAS

„Moodle“ yra atviro kodo mokymosi kursų valdymo sistema, sukurta siekiant padėti mokytojams kurti veiksmingas internetines (on-line) mokymosi bendruomenes. „Moodle“ modulinis dizainas leidžia lengvai kurti naujus kursus, papildyti turinį.

„Moodle“ yra lanksti, įvairius modulius ir išteklius apjungianti sistema, kurioje galima sukurti penkių tipų statinio kurso medžiagas (teksto puslapį, tinklalapį, nuorodą į ką nors internete, į ką nors intranete, katalogų peržiūrą ir lauką, kuris rodo į bet kokį tekstą ar paveikslėlį), šešių tipų interaktyvių kursų materialus (užduotis, pasirinkimus, žurnalus, pamokas, klausimynus, apžvalgas) bei penkis aktyvumo tipus (pokalbius, forumus, žodynus, „wiki“ bei dirbtuves



The screenshot shows the Moodle interface for 'Course Fullname 101'. At the top right, it says 'You are logged in as Admin User (Logout)'. Below the course name, there is a breadcrumb trail: 'moodle » CF101 » Logs » All participants, All days'. The main heading is 'Course Fullname 101: All participants, All days (Server's local time)'. Below this, there are filters: 'Course Fullname 101', 'All participants', 'All days', 'All activities', and a 'Show these logs' button. It indicates 'Displaying 78 records'. The main content is a table with the following columns: Time, IP Address, Full name, Action, and Information.

Time	IP Address	Full name	Action	Information
Sun 11 February 2007, 10:00 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Sun 11 February 2007, 09:57 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz report	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz report	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz preview	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz view	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz preview	4
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz view	4
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 15 January 2007, 01:43 PM	127.0.0.1	Admin User	quiz preview	3
Mon 15 January 2007, 01:43 PM	127.0.0.1	Admin User	quiz view	3

Pav. 7 Moodle naudojimosi žurnalas

„Moodle“ sistema detaliai žurnalizuoja studentų aktyvumą. „Moodle“ žurnaluose matomas kiekvienas paspaudimas, studentų daromas navigacijos tikslais. „Moodle“ turi žurnalizuotų duomenų peržiūrėjimo sistemą (7 pav.).

„Moodle“ neišsaugo aktyvumo žurnaluose kaip tekstinius failus. Vietoj to ji saugo žurnalus Reliacinėje duomenų bazėje. Taigi, duomenys yra saugomi vienoje duomenų bazėje.

„Moodle“ duomenų bazėje yra apie 145 tarpusavyje susijusios lentelės. Antroje lentelėje pateikiama svarbiausių „Moodle“ lentelių aprašymai:

Lentelė 2 Moodle duomenų bazės svarbiausios lentelės

Pavadinimas	Aprašymas
mdl_user	Informacija apie visus vartotojus.
mdl_user_students	Informacija apie visus studentus.
mdl_log	Žurnalizuoja kiekvieno vartotojo veiksmus.
mdl_assignment	Informacija apie kiekvieną priskyrimą.
mdl_assignment_submissions	Informacija apie priskyrimų pateiktis.
mdl_chat	Informacija apie pokalbių aplinkas.
mdl_chat_users	Sekama kurie vartotojai kuriuose pokalbių aplinkose bendrauja.
mdl_choice	Informacija apie visus pasirinkimus.
mdl_glossary	Informacija apie visus žodynus.
mdl_survey	Informacija apie visas apklausas.
mdl_forum	Informacija apie visus forumus.
mdl_forum_posts	Saugomi forumų pasisakymai.
mdl_forum_discussions	Saugomos forumų diskusijos.
mdl_message	Saugomos žinutės.
mdl_message_reads	Saugomos perskaitytos žinutės.ž
mdl_quiz	Informacija apie visas apklausas.
mdl_quiz_attempts	Saugoma apklausų bandymų informacija.
mdl_quiz_grades	Saugomi apklausų rezultatai/pažymiai.

Išankstinis duomenų apdorojimas leidžia visus duomenis paversti į tinkamą formą konkrečiam duomenų gavybos algoritmui. Prieš vykdant duomenų gavybos algoritmus, reikėtų atlikti paruošiamuosius veiksmus kaip kad: duomenų valymas, vartotojo identifikacijos nustatymas, sesijos identifikavimas, pabaigos maršrutaso radimas, sandorio identifikacijos nustatymas, duomenų transformavimas ir sodrinimas, duomenų integravimas, duomenų kiekio skaičiavimas.

Tačiau duomenų pirminis apdorojimas turi keletą konkrečių aspektų:

- „Moodle“ suteikia vartotojų autentifikaciją (slaptažodžių apsauga), kuomet žurnalų failuose eilutės identifikuojamos pagal vartotojus. Kitaip sakant, vartotojas turi savo

specifinį kodą, kuris išlieka ir kitų prisijungimų atveju, todėl galime pašalinti tipinį vartotoją, paliekant tik jo autentifikacijos id.

- „Moodle“ registruoja studentų naudojimosi informaciją ne tik žurnalų failuose, bet ir tiesiogiai į duomenų bazę.

Taigi duomenys, surinkti iš e-mokymosi sistemų, reikalauja mažesnio duomenų valymo bei apdorojimo.

Atrenkant duomenis, būtina pasirinkti kursus, kurie mus domina.

Lentelė 3 apibendrinti duomenys

Pavadinimas	Aprašymas
course	Kurso identifikavimo numeris.
n_assignment	Paskyrimų skaičius.
n_quiz	Atliktų apklausų skaičius.
n_quiz_a	Sėkmingai atliktų užduočių skaičius.
n_quiz_s	Nesėkmingai atliktų užduočių skaičius.
n_messages	Siųstų žinučių pokalbiams skaičius.
n_messages_ap	Siųstu žinučių mokytojui skaičius.
n_posts	Siųstų žinučių į forumą skaičius.
n_read	Forume perskaitytų žinučių skaičius.
total_time_assignment	Visas laikas praleistas paskyrimui.
total_time_quiz	Visas laikas praleistas apklausose.
total_time_forum	Visas laikas praleistas forumuose.
mark	Galutinis studentų rezultatas gautas mokymo kurso metu.

Sukuriame santraukos lenteles. Būtina sukurti naują lentelę (mdl_summarization) „Moodle“ duomenų bazėje, kuri gali būti apibendrinta informacijai reikiamu lygiu (pvz., studentas). Studentas ir sąveikos duomenys yra keliose lentelėse. Sukuriame naują suvestinę lentelę (3 lentelė.), kuri integruoja svarbiausią informaciją apie mūsų tikslą.

4. EKSPERIMENTINĖ DALIS

Iš moodle sistemos mdl_log eksportuojami vartotojų naudojimosi duomenys.

	Field	Type	Collation	Attributes	Null	Default	Extra
<input type="checkbox"/>	id	bigint(10)		UNSIGNED	No	None	AUTO_INCREMENT
<input type="checkbox"/>	time	bigint(10)		UNSIGNED	No	0	
<input type="checkbox"/>	userid	bigint(10)		UNSIGNED	No	0	
<input type="checkbox"/>	ip	varchar(45)	utf8_unicode_ci		No		
<input type="checkbox"/>	course	bigint(10)		UNSIGNED	No	0	
<input type="checkbox"/>	module	varchar(20)	utf8_unicode_ci		No		
<input type="checkbox"/>	cmid	bigint(10)		UNSIGNED	No	0	
<input type="checkbox"/>	action	varchar(40)	utf8_unicode_ci		No		
<input type="checkbox"/>	url	varchar(100)	utf8_unicode_ci		No		
<input type="checkbox"/>	info	varchar(255)	utf8_unicode_ci		No		

Pav. 8 Moodle mdl_log lentelės struktūra

Turimi duomenys suimportuojami į naujai sukurtą sistemą. Naudojimo duomenys apdorojami ir sukuriama apibendrinanti lentelė mld_dm_action_summary. Ją sudaro vartotojo identifikatorius, kurso pavadinimas ir kiekvieno atlikto veiksmo kiekiai.

	Field	Type	Collation	Attributes	Null	Default	Extra
<input type="checkbox"/>	userid	int(11)			No	None	
<input type="checkbox"/>	course	varchar(256)	utf8_unicode_ci		No	None	
<input type="checkbox"/>	add_cnt	int(11)			No	0	
<input type="checkbox"/>	add contact_cnt	int(11)			No	0	
<input type="checkbox"/>	add discussion_cnt	int(11)			No	0	
<input type="checkbox"/>	add entry_cnt	int(11)			No	0	
<input type="checkbox"/>	add mod_cnt	int(11)			No	0	
<input type="checkbox"/>	add post_cnt	int(11)			No	0	
<input type="checkbox"/>	assess_cnt	int(11)			No	0	
<input type="checkbox"/>	assign_cnt	int(11)			No	0	
<input type="checkbox"/>	attachments_cnt	int(11)			No	0	

Pav. 9 Apibendrinančios lentelės mld_dm_action_summary pagrindinių laukų struktūra

Ši lentelė naudojama duomenų gavybai. Iš viso turime 87 atributus. Pasirinkus norimą kursą, galima peržiūrėti to kurso veiksmų statistiką ir išsisaugoti ARFF formato failą.

Lentelė 4 pateikiama kurso „13“ ARFF duomenų failas. Šis kursas turi 503 studentus.

Lentelė 4 Kurso „13“ pradinis ARFF duomenų failas

```
% ARFF file for moodle data
% course 13
@relation moodle_course_students_activity

@attribute userid numeric
@attribute finalgrade numeric
@attribute add numeric
@attribute add_category numeric
@attribute add_comment numeric
@attribute add_contact numeric
@attribute add_discussion numeric
@attribute add_entry numeric
@attribute add_mod numeric
@attribute add_post numeric
@attribute assess numeric
@attribute assign numeric
@attribute attempt numeric
@attribute block_contact numeric
@attribute change_password numeric
@attribute choose numeric
@attribute choose_again numeric
@attribute close_attempt numeric
@attribute continue_attemp numeric
@attribute continue_attempt numeric
@attribute delete numeric
@attribute delete_attempt numeric
@attribute delete_comment numeric
@attribute delete_discussion numeric
@attribute delete_entry numeric
@attribute delete_mod numeric
@attribute delete_post numeric
@attribute duplicate numeric
@attribute edit numeric
@attribute editquestions numeric
@attribute editsection numeric
@attribute end numeric
@attribute enrol numeric
@attribute error numeric
@attribute fields_add numeric
@attribute grade numeric
@attribute history numeric
@attribute info numeric
@attribute links numeric
@attribute login numeric
@attribute logout numeric
@attribute mail_blocked numeric
```


@attribute delete_mod numeric
@attribute delete_post numeric
@attribute duplicate numeric
@attribute edit numeric
@attribute editquestions numeric
@attribute editsection numeric
@attribute end numeric
@attribute enrol numeric
@attribute error numeric
@attribute fields_add numeric
@attribute grade numeric
@attribute history numeric
@attribute info numeric
@attribute links numeric
@attribute login numeric
@attribute logout numeric
@attribute mail_blocked numeric
@attribute mail_error numeric
@attribute mailer numeric
@attribute new numeric
@attribute preview numeric
@attribute recent numeric
@attribute remove_contact numeric
@attribute report numeric
@attribute report_live numeric
@attribute report_log numeric
@attribute report_outline numeric
@attribute report_participation numeric
@attribute report_stats numeric
@attribute review numeric
@attribute search numeric
@attribute start numeric
@attribute start_tracking numeric
@attribute stop_tracking numeric
@attribute submit numeric
@attribute subscribe numeric
@attribute subscribeall numeric
@attribute talk numeric
@attribute templates_view numeric
@attribute unassign numeric
@attribute unblock_contact numeric
@attribute unenrol numeric
@attribute unsubscribe numeric
@attribute unsubscribeall numeric
@attribute update numeric
@attribute update_comment numeric
@attribute update_entry numeric
@attribute update_grades numeric
@attribute update_mod numeric
@attribute update_post numeric
@attribute upload numeric
@attribute user_report numeric
@attribute view numeric

choose	0.0497	0.1701	0
choose_again	0.0027	0.0091	0
close_attempt	0.1762	0.3133	0.1196
continue_attemp	0.2622	0.3805	0.2133
continue_attempt	0	0	0
delete	0	0	0
delete_attempt	0.002	0.0068	0
delete_comment	0	0	0
delete_discussion	0	0	0
delete_entry	0	0	0
delete_mod	0.002	0.0068	0
delete_post	0	0	0
duplicate	0	0	0
edit	0	0	0
editquestions	0.002	0.0068	0
editsection	0.002	0.0068	0
end	0	0	0
enrol	0.0139	0.0272	0.0084
error	0	0	0
fields_add	0	0	0
grade	0	0	0
history	0	0	0
info	0	0	0
links	0	0	0
login	0	0	0
logout	0	0	0
mail_blocked	0	0	0
mail_error	0	0	0
mailer	0	0	0
new	0	0	0
preview	0.002	0.0068	0
recent	0.0241	0.0551	0.0112
remove_contact	0	0	0
report	0.002	0.0068	0
report_live	0	0	0
report_log	0.002	0	0.0028
report_outline	0	0	0
report_participation	0	0	0
report_stats	0	0	0
review	0.002	0.0068	0
search	0.0126	0.0272	0.0066
start	0	0	0
start_tracking	0	0	0

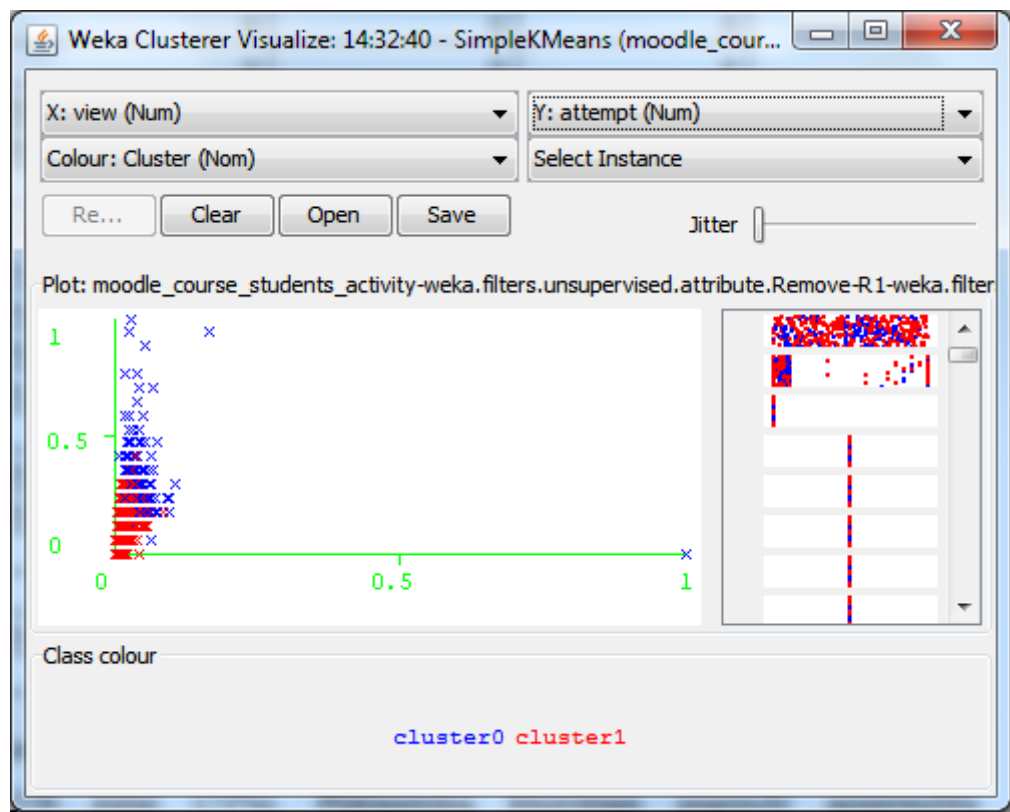
stop_tracking	0	0	0
submit	0	0	0
subscribe	0	0	0
subscribeall	0	0	0
talk	0	0	0
templates_view	0	0	0
unassign	0.002	0.0068	0
unblock_contact	0	0	0
unenrol	0.002	0.0068	0
unsubscribe	0	0	0
unsubscribeall	0	0	0
update	0.0022	0.0071	0.0002
update_comment	0	0	0
update_entry	0	0	0
update_grades	0.002	0.0068	0
update_mod	0.002	0.0068	0
update_post	0	0	0
upload	0.0811	0.1119	0.0684
user_report	0.0095	0.0179	0.0061
view	0.0278	0.049	0.019
view_all	0.0101	0.0226	0.0049
view_discussion	0.1083	0.2255	0.0599
view_entry	0	0	0
view_forum	0.0666	0.1386	0.0369
view_forums	0.0535	0.1218	0.0253
view_grade	0	0	0
view_responses	0	0	0
view_submission	0.002	0.0068	0.0001
write	0	0	0

Klasteriai stabilizavosi po 23 iteracijų. Pirmam klasteriui priskirti 147 įrašai (29%), o antram klasteriui priskirti 356 įrašai (71%). Priklausymą klasteriam daugiausia apsprendžia attempt,change_password,choose,choose_again,close_attempt,continue_attemp,enrol,recent,search,update,upload,user_report,view,view_all,view_discussion,view_forum,view_forums atributai. Pirmam klasteriui priklauso tie vartotojai, kurių atlikti veiksmai viršija vidurkį, o antram - su mažesniais skaičiais. Galima taip išskirti aktyvius ir neaktyvius studentus.

Studentų priskyrimas klasteriams (įrašo numeris, klasterio numeris), pateikiamas Lentelė 7.

Lentelė 7 Kurso „13“ studentų priskyrimas klasteriams

0 1
1 1
2 1
3 0
4 1
5 1
6 1
7 0
8 0
....



Pav. 10 Klasterizavimo vizualizacija

Klasterizavimas, ieškant 5 klasterių:

Lentelė 8 Kurso „13“ klasterizavimo rezultatai (5 klasteriai)

Atributai	Visi duomenys (503)	Klasteris1 (264)	Klasteris2 (58)	Klasteris3 (43)	Klasteris4 (25)	Klasteris5 (113)
finalgrade	0.1478	0.0715	0.0788	0.8952	0.0933	0.0891
add	0.002	0.0038	0	0	0	0
add_category	0	0	0	0	0	0
add_comment	0	0	0	0	0	0

add_contact	0	0	0	0	0	0
add_discussion	0	0	0	0	0	0
add_entry	0	0	0	0	0	0
add_mod	0.002	0.0038	0	0	0	0
add_post	0	0	0	0	0	0
assess	0	0	0	0	0	0
assign	0.002	0.0038	0	0	0	0
attempt	0.1829	0.1081	0.1886	0.1587	0.2494	0.3493
block_contact	0	0	0	0	0	0
change_password	0.0199	0.0114	0.1034	0	0.04	0
choose	0.0497	0	0	0	1	0
choose_again	0.0027	0	0	0	0.0533	0
close_attempt	0.1762	0.1032	0.1795	0.1546	0.2353	0.3404
continue_attemp	0.2622	0.1911	0.276	0.2387	0.2722	0.4277
continue_attempt	0	0	0	0	0	0
delete	0	0	0	0	0	0
delete_attempt	0.002	0.0038	0	0	0	0
delete_comment	0	0	0	0	0	0
delete_discussion	0	0	0	0	0	0
delete_entry	0	0	0	0	0	0
delete_mod	0.002	0.0038	0	0	0	0
delete_post	0	0	0	0	0	0
duplicate	0	0	0	0	0	0
edit	0	0	0	0	0	0
editquestions	0.002	0.0038	0	0	0	0
editsection	0.002	0.0038	0	0	0	0
end	0	0	0	0	0	0
enrol	0.0139	0.0038	0.0172	0	0	0.0442
error	0	0	0	0	0	0
fields_add	0	0	0	0	0	0
grade	0	0	0	0	0	0
history	0	0	0	0	0	0
info	0	0	0	0	0	0
links	0	0	0	0	0	0
login	0	0	0	0	0	0
logout	0	0	0	0	0	0
mail_blocked	0	0	0	0	0	0
mail_error	0	0	0	0	0	0
mailer	0	0	0	0	0	0
new	0	0	0	0	0	0
preview	0.002	0.0038	0	0	0	0
recent	0.0241	0.0114	0.1121	0.0163	0.028	0.0106
remove_contact	0	0	0	0	0	0

report	0.002	0.0038	0	0	0	0
report_live	0	0	0	0	0	0
report_log	0.002	0.0038	0	0	0	0
report_outline	0	0	0	0	0	0
report_participation	0	0	0	0	0	0
report_stats	0	0	0	0	0	0
review	0.002	0.0038	0	0.0001	0	0
search	0.0126	0.0051	0.0115	0.031	0.04	0.0177
start	0	0	0	0	0	0
start_tracking	0	0	0	0	0	0
stop_tracking	0	0	0	0	0	0
submit	0	0	0	0	0	0
subscribe	0	0	0	0	0	0
subscribeall	0	0	0	0	0	0
talk	0	0	0	0	0	0
templates_view	0	0	0	0	0	0
unassign	0.002	0.0038	0	0	0	0
unblock_contact	0	0	0	0	0	0
unenrol	0.002	0.0038	0	0	0	0
unsubscribe	0	0	0	0	0	0
unsubscribeall	0	0	0	0	0	0
update	0.0022	0.004	0.0003	0.0003	0.0004	0.0001
update_comment	0	0	0	0	0	0
update_entry	0	0	0	0	0	0
update_grades	0.002	0.0038	0	0	0	0
update_mod	0.002	0.0038	0	0	0	0
update_post	0	0	0	0	0	0
upload	0.0811	0.066	0.1143	0.0875	0.1158	0.0892
user_report	0.0095	0.0082	0.0192	0.0116	0.0077	0.0073
view	0.0278	0.0207	0.046	0.0233	0.0352	0.035
view_all	0.0101	0.0078	0.021	0.0078	0.0087	0.0109
view_discussion	0.1083	0.0478	0.3912	0.1252	0.0677	0.1069
view_entry	0	0	0	0	0	0
view_forum	0.0666	0.0296	0.2101	0.0654	0.1025	0.0719
view_forums	0.0535	0.0176	0.1661	0.0761	0.0836	0.0644
view_grade	0	0	0	0	0	0
view_responses	0	0	0	0	0	0
view_submission	0.002	0.0039	0	0	0	0
write	0	0	0	0	0	0

Atliekant klasterizavimą, ieškant 5 klasterių, pirmam klasteriui priskirti 262 (52%) įrašai, antram – 58 (12%), trečiam – 43 (9), ketvirtam – 25 (5%), penktam – 113 (22%).

Antram klasteriui priklauso studentai, dažnai keičiantys slaptažodį, peržiūrintys naujienas, diskusijas, forumus, įkeliantys failus.

Trečiam klasteriui priklauso gerai besimokantys studentai.

Penktam klasteriui priklauso studentai, dažnai laikantys testus, dažnai žiūrintys diskusijas.

Ketvirtam klasteriui priklauso studentai, dažniausiai atliekantys paiešką. Jie gan panašūs į penkto klasterio studentus, bet rečiau peržiūri diskusijas.

Pirmam klasteriui priklauso likę studentai, jie atlieka įvairesnius veiksmus nei kituose klasteriuose esantys žmonės.

Lentelė 9 Kurso „10“ klasterizavimas, ieškant 4 klasterių

Atributai	Visi duomenys (354)	Klasteris1 (138)	Klasteris2 (7)	Klasteris3 (162)	Klasteris4 (47)
finalgrade	0.2859	0.7049	0	0.0133	0.038
add	0.0045	0	0.2293	0	0
add_discussion	0.0181	0.0135	0.2609	0.0097	0.0241
add_mod	0.0045	0	0.2293	0	0
add_post	0.0347	0.0267	0.2457	0.0277	0.0511
assign	0.0056	0	0.1905	0.0041	0
attempt	0.0716	0.1294	0.1122	0.0065	0.1201
change_password	0.0028	0	0	0.0062	0
close_attempt	0.0714	0.1297	0.115	0.0056	0.1204
continue_attemp	0.1095	0.1877	0.1801	0.0154	0.1934
delete_discussion	0.0085	0.002	0.2078	0.0056	0.0077
delete_mod	0.0038	0	0.1921	0	0
delete_post	0.0073	0.0048	0.1667	0.0021	0.0089
editquestions	0.0035	0	0.1761	0	0
editsection	0.0042	0	0.2143	0	0
enrol	0.5113	0.5471	0.3571	0.4877	0.5106
preview	0.0028	0	0.1429	0	0
recent	0.0056	0.0015	0.2069	0.0011	0.0037
report	0.0033	0.0003	0.1548	0	0.001
report_log	0.0035	0	0.1429	0.0014	0
review	0.0028	0	0.1429	0	0
search	0.0214	0.0269	0.0204	0.0185	0.0152
subscribe	0.0141	0.0181	0.1429	0.0062	0.0106
talk	0.0048	0.0027	0.1617	0	0.0044
unassign	0.0031	0	0.1429	0.0006	0
unsubscribe	0.0078	0.0036	0.2857	0	0.0053
update	0.0095	0.0043	0.2846	0.0036	0.0041
update_mod	0.0047	0	0.2381	0	0

update_post	0.0195	0.0091	0.2143	0.0062	0.0674
upload	0.0793	0.0445	0.6327	0.0163	0.3161
user_report	0.0444	0.0542	0.2517	0.017	0.079
view	0.1009	0.1024	0.6751	0.0527	0.1769
view_all	0.0466	0.0378	0.5738	0.0165	0.0972
view_discussion	0.1092	0.0932	0.664	0.0787	0.1785
view_forum	0.1024	0.0828	0.7181	0.0703	0.1787
view_forums	0.0375	0.0258	0.2698	0.0254	0.0788
view_submission	0.0028	0	0.1429	0	0

Pirmam klasteriui priklauso 138 (39%) įrašai, antram - 7 (2%), trečiam - 162 (46%), ketvirtam - 47 (13%).

Pirmam klasteriui priklauso vartotojai, gavę gerus įvertinimus, jie dažnai laiko testus, atlieka paiešką.

Antram klasteriui priklauso vartotojai, atliekantys daug administracinių veiksmų (kuriantys įrašus, priskiriantys studentus, tvarkantys diskusijas, forumus), tai tikriausiai mokytojai.

Trečiam ir ketvirtam klasteriams priklauso studentai, gavę prastus įvertinimus.

Trečio klasterio studentai neaktyvus – atlieka mažai veiksmų.

Ketvirto klasterio studentai dažniausiai dalyvauja diskusijose, bando spręsti testus, dažniau nei pirmo klasterio studentai įkrauna failus.

5. IŠVADOS

1. Šiame darbe buvo apžvelgti nuotolinių studijų duomenų gavybos būdai ir pasirinktas klasterizavimo metodas tyrimui.
2. Pažingsniui apžvelgtas mokymo sistemos „Moodle“ duomenų gavybos procesas, apimantis duomenų rinkimą, pirminį apdorojimą, duomenų gavybą bei rezultatų analizę ir panaudojimą. Taipogi buvo apžvelgti duomenų gavybos analizės metodai kaip kad „aprašomoji statistika“, „vizualizacija“.
3. Realizuota sistema, leidžianti atlikti duomenų gavybą, naudojant „Moodle“ duomenis: importuoti duomenys apibendrinami, tuomet duomenys paruošiami duomenų gavybai (pašalinami nereikalingi atributai, normalizuojamos reikšmės), atliekamas klasterizavimas, gautus rezultatus ir paruoštą „ARFF“ duomenų failą galima išsaugoti tolimesnei analizei.
4. Tyrimo metu buvo naudoti Šiaulių Universiteto „Moodle“ sistemos duomenys, atlikus klasterizavimą pagal sistemos naudojimą, vartotojai pasiskirstė į skirtingas grupes (aktyvius, neaktyvius, gerai besimokančius studentus ir mokytojus).
5. „Weka“ duomenų gavybos algoritmus lengva pritaikyti kuriant integruotas sistemas.
6. Taupant laiką ir nenaudojant gausybės skirtingų duomenų gavybos programinių paketų ateityje būtų labai pravartu turėti specialias programas, sukurtas (arba integruotas į e-mokymosi sistemą) būtent e-mokymosi duomenų gavybai.

6. LITERATŪRA

1. Dekker G.; Pechenizkiy M.; Vleeshouwers J. Predicting Students Drop Out: A Case Study. Proceedings of the International Conference on Educational Data Mining. Cordoba, Spain. Educational data mining, 2009.
2. Hall M.; Frank, E.; Holmes G.; Pfahringer B.; Reutemann P.; Witten I. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Tomas 11, Nr. 1, 2009
3. Khabaza T. Nine Laws of Data Mining. 2010. [žiūrėta 2010-11-03]. Prieiga per internetą: <http://www.khabaza.com/>
4. Mamčenko, Jelena [Mamcenko, Jelena]; Belevičiūtė, Inga [Beleviciute, Inga]. Data mining for knowledge management in technology enhanced learning / Jelena Mamcenko, Inga Beleviciute // Proceedings of the 6th WSEAS International Conferences on Applications of Electrical Engineering (AEE'07): Istanbul, Turkey, May 27-29, 2007. Istanbul : WSEAS Press, 2007. ISBN 9789608457713. p. 115-119.
5. Moodle. [žiūrėta 2010-10-15]. Prieiga per internetą: <http://moodle.org/>
6. Refaeilzadeh P.; Tang L.; Liu H. Cross Validation. Encyclopedia of Database Systems, Springer, 2009.
7. Romero C.; Ventura S. Educational data mining: a survey from 1995 to 2005. Expert Systems with Applications. Nr. 33, p. 135-146, 2007.
8. Romero C.; Ventura S., Espejo P. G.; Hervas C. Data Mining Algorithms to Classify Students. Proceedings of the first International Conference on Educational Data Mining, 8-17, 2008.
9. Simoudis E., Reality check for data mining, IEEE expert, Vol. 11 Nr. 5, 1996, pp. 28-30.
10. Turskienė S., Kulvietis G., Burbaitė R. Duomenų gavybos technologijų panaudojimas jaunųjų kompiuterininkų mokyklos veiklos analizei, Lietuvos matematikos rinkinys, 2003, t. 43, spec.nr., p.297-303. ISSN 0132-2818.
11. Weka. [žiūrėta 2010-12-07]. Prieiga per internetą: <http://www.cs.waikato.ac.nz/ml/weka/>
12. Witten I. H.; Frank E. Data Mining Practical Machine Learning Tools And Techniques, Morgan Kaufmann, 2 leidimas 2005. ISBN 0120-884-07-0.
13. Zaiane O. Web Usage Mining for a Better Web-Based Learning Environment. Proceedings of Conference on Advanced Technology for Education (CATE'01), p. 60-64, Banff, Alberta, 2001.

7. SANTRUMPŲ IR TERMINŲ ŽODYNAS

1. **LMS (Learning Management System)** – mokymosi tvarkymo sistema. Programinės įrangos administravimo, dokumentacijos, stebėjimo ir ataskaitų apie mokymo programas, klases ir internetinius renginius sistema.
2. **ARFF (Attribute-Relation File Format)** – atributų santykio failo formatas naudojamas Weka.
3. **GNU GPL (General Public License)** – GNU bendroji viešoji licencija.

8. PRIEDAI

8.1. PRIEDAS 1. DUOMENŲ ANALZIĖ

Veiksmų kiekiai:

```
SELECT ACTION , COUNT( * ) FROM EVENTS GROUP BY 1
```

<u>action</u>	<u>count(*)</u>
add	306
add contact	76
add discussion	555
add entry	21
add mod	297
add post	852
assess	135
assign	18
attachments	4
attempt	18368
block contact	30
change password	366
close attempt	17160
continue attemp	115217
choose	106
choose again	29
delete attempt	2004
delete discussi	381
delete mod	173
delete post	91
edit	4
editquestions	1622
editsection	57
enrol	2676
error	14611
grade	1

<u>action</u>	<u>count(*)</u>
guest	6
history	258
info	4
links	10
login	47936
logout	24097
mail blocked	339
mail error	988
mailer	1066
mark read	1
new	1
newattachment	13
preview	196
prune post	1
recent	389
remove contact	29
removeattachmen	2
report	3446
report live	27
report log	33
report outline	3
report particip	6
review	24643
search	393
submit	93
subscribe	75
subscribeall	36
talk	3125
unassign	64
unblock contact	14
unenrol	52
unsubscribe	87

<u>action</u>	<u>count(*)</u>
unsubscribeall	9
update	4979
update entry	3
update feedback	5
update grades	649
update mod	1945
update post	408
upload	3099
user report	2775
view	315805
view all	27924
view discussion	21513
view entry	411
view forum	23230
view forums	6098
view responses	11
view submission	2425
write	1691

Viso veiksmų:

```
SELECT COUNT( * ) FROM EVENTS
```

count(*)

695573

Vartotojų:

```
SELECT COUNT( DISTINCT userid ) FROM EVENTS
```

count(distinct userid)

3055

Kursų:

```
SELECT COUNT( DISTINCT course ) FROM EVENTS
```

count(distinct course)

23

Arff kursui 4:

```
% ARFF file for moodle data
@relation moodle_course_students_activity
@attribute userid numeric
@attribute add numeric
@attribute add_contact numeric
@attribute add_discussion numeric
@attribute add_entry numeric
@attribute add_mod numeric
@attribute add_post numeric
@attribute assess numeric
@attribute assign numeric
@attribute attachments numeric
@attribute attempt numeric
@attribute block_contact numeric
@attribute change_password numeric
@attribute close_attempt numeric
@attribute continue_attemp numeric
@attribute choose numeric
@attribute choose_again numeric
@attribute delete_attempt numeric
@attribute delete_discussi numeric
@attribute delete_mod numeric
@attribute delete_post numeric
@attribute edit numeric
@attribute editquestions numeric
@attribute editsection numeric
@attribute enrol numeric
@attribute error numeric
@attribute grade numeric
@attribute guest numeric
@attribute history numeric
@attribute info numeric
@attribute links numeric
@attribute login numeric
@attribute logout numeric
@attribute mail_blocked numeric
@attribute mail_error numeric
@attribute mailer numeric
@attribute mark_read numeric
@attribute new numeric
@attribute newattachment numeric
@attribute preview numeric
@attribute prune_post numeric
@attribute recent numeric
@attribute remove_contact numeric
@attribute removeattachmen numeric
@attribute report numeric
```


8.2. PRIEDAS 2. PROGRAMOS KODAS

```
<?php session_start(); ?>
<html>
<head>
<title>Logs</title>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">

<style type="text/css">
.row_odd {
color:gray;
}
.row_even {
background: #ecf6fc;

}
.tableHeaderLabel {
    color:black;
    font-weight: bold;
}
.tableCustom {
    border-collapse: collapse;
    border: 1px solid #95bce2;

}
.tableCustom td{
    border-bottom: 1px solid #95bce2;
}
.tableCustom th{
    background: #3e83c9;
    color: #fff;
    font-weight: bold;
    text-align: center;
    border: 1px solid #3e83c9;
}
.tableHeader {

}

</style>

<link href="./uploadify/uploadify.css" type="text/css" rel="stylesheet" />
<script type="text/javascript" src="./uploadify/jquery-1.4.2.min.js"></script>
<script type="text/javascript" src="./uploadify/swfobject.js"></script>
<script type="text/javascript"
src="./uploadify/jquery.uploadify.v2.1.4.js"></script>
<script type="text/javascript">
function reload() {
window.location='index.php';
}
$(document).ready(function() {
    $('#file_upload').uploadify({
        'uploader' : './uploadify/uploadify.swf',
        //'script'   : './uploadify/uploadify.php',
```

```

        'script'      : './index.php',
        'cancelImg'  : './uploadify/cancel.png',
        'folder'    : './uploads',
        'auto'       : true,
        'buttonText' : 'Pasirinkite CSV',
        'fileExt'    : '*.csv',
        'onAllComplete' : reload

    });
});

</script>

</head>

<body>

<?php

function db_query($sql) {

    $DB_IP = "127.0.0.1";
    $DB_USER = "logs";
    $DB_PASS = "logs";
    $DB_DATABASE = "logs";
    $link = mysql_connect($DB_IP, $DB_USER, $DB_PASS);
    mysql_query('SET NAMES utf8');
    mysql_query('SET CHARACTER SET utf8');
    mysql_query('SET character_set_client="utf8"');
    mysql_query('SET character_set_connection="utf8"');
    mysql_query('SET character_set_results="utf8"');

    if (!$link) {
        die('Not connected : ' . mysql_error());
    }

    $db_selected = mysql_select_db($DB_DATABASE, $link);

    if (!$db_selected) {
        die ('Can\'t use foo : ' . mysql_error());
    }
    $result = mysql_query($sql, $link);

    $rez = array();

    if($result == false)
    {
        die("Query failed: " . mysql_error() );
    }
    elseif(mysql_num_rows($result) == 0)
    {
    }
    else
    {
        while($row = mysql_fetch_assoc($result))
        {
            $rez[]=$row;
        }
    }
}

```

```

}

if (!$link) {
mysql_close($link);
}
return $rez;
}

function db_query2($sql) {

$DB_IP = "127.0.0.1";
$DB_USER = "logs";
$DB_PASS = "logs";
$DB_DATABASE = "logs";
$link = mysql_connect($DB_IP, $DB_USER, $DB_PASS);
mysql_query('SET NAMES utf8');
mysql_query('SET CHARACTER SET utf8');
mysql_query('SET character_set_client="utf8"');
mysql_query('SET character_set_connection="utf8"');
mysql_query('SET character_set_results="utf8"');

if (!$link) {
    die('Not connected : ' . mysql_error());
}

$db_selected = mysql_select_db($DB_DATABASE, $link);

if (!$db_selected) {
    die ('Can\'t use foo : ' . mysql_error());
}
$result = mysql_query($sql, $link);

if (!$link) {
mysql_close($link);
}
return $result;
}

function printSelect($rez, $name) {
    print "<select name=\"". $name . "\" >";
    foreach ($rez as $row) {
        print "<option value=\"". $row[$name]. "\">". $row[$name]. "</option>";
    }
    print "</select>";
}

function printEventsTable($rez, $tableDesc) {
    print "<br/>";
    $tableDesc = $tableDesc." ". count($rez);

    print "<div class=\"tableHeaderLabel\">". $tableDesc. "</div>";

    print "<table class=\"tableCustom\">";
    print "<tr class=\"tableHeader\">";

```

```

print '<th>'. 'Laikas'. '</th>';
print '<th>'. 'Vartotojas'. '</th>';
print '<th>'. 'IP'. '</th>';
print '<th>'. 'Kursas'. '</th>';
print '<th>'. 'Modulis'. '</th>';
print '<th>'. 'Komanda'. '</th>';
print '<th>'. 'Veiksmas'. '</th>';
print '<th>'. 'URL'. '</th>';
print '<th>'. 'Info'. '</th>';
print '</tr>';

$cnt = 0;
foreach ($rez as $row) {
    $style = ($cnt % 2 == 0) ? "row_odd" : "row_even";
    print "<tr class=\"\". $style. \"\">";

    $t = str_replace(" ", "&nbsp;" , date('Y-m-d H:m:s', $row['time']));
    print '<td>'. $t. '</td>';
    print '<td>'. $row['userid']. '</td>';
    print '<td>'. $row['ip']. '</td>';
    print '<td style="text-align:center;">'. $row['course']. '</td>';
    print '<td>'. $row['module']. '</td>';
    print '<td>'. $row['cmid']. '</td>';
    print '<td>'. $row['action']. '</td>';
    print '<td>'. $row['url']. '</td>';
    print '<td>'. $row['info']. '</td>';

    print '</tr>';

    $cnt++;
}
print "</table>";
}

function printBigTable($rez, $tableDesc) {
    print "<br/>";
    $tableDesc = $tableDesc;

    print "<div class=\"tableHeaderLabel\">". $tableDesc. "</div>";

    print "<table class=\"tableCustom\">";
    print "<tr class=\"tableHeader\">";

    $header=$rez[0];
    foreach ($header as $c) {
        print '<td>'. $c. '</td>';
    }
    print '</tr>';

    $cnt = 0;
    foreach ($rez as $row) {
        if ($cnt > 0) {
            $style = ($cnt % 2 == 1) ? "row_odd" : "row_even";
            print "<tr class=\"\". $style. \"\">";
            foreach ($row as $c) {
                print '<td>'. $c. '</td>';
            }
            print '</tr>';
        }
    }
}

```

```

        }

        $cnt++;
    }
    print "</table>";
}

function printCountsTable($rez, $field, $cntField, $tableDesc, $fieldDesc,
$cntFieldDesc) {
    print "<br/>";
    $tableDesc = $tableDesc;
    //." ". count($rez);

    print "<div class=\"tableHeaderLabel\">".$tableDesc."</div>";

    print "<table class=\"tableCustom\">";
    print "<tr class=\"tableHeader\">";
    print "<th>".$fieldDesc."</th>";
    print "<th>".$cntFieldDesc."</th>";
    print '</tr>';
    $cnt = 0;
    foreach ($rez as $row) {
        $style = ($cnt % 2 == 0) ? "row_odd" : "row_even";
        print "<tr class=\"".$style.\">";

        print '<td>'.$row[$field].'\</td>';
        print '<td>'.$row[$cntField].'\</td>';

        print '</tr>';

        $cnt++;
    }
    print "</table>";
}

function printGeneralTable($rez, $header_name, $header_desc, $tableDesc) {
    print "<br/>";
    print "<div class=\"tableHeaderLabel\">".$tableDesc."</div>";
    print "<table class=\"tableCustom\">";
    print "<tr class=\"tableHeader\">";
    foreach ($header_desc as $fieldDesc) {
        print "<th>".$fieldDesc."</th>";
    }
    print '</tr>';
    $cnt = 0;
    foreach ($rez as $row) {
        $style = ($cnt % 2 == 0) ? "row_odd" : "row_even";
        print "<tr class=\"".$style.\">";

        foreach ($header_name as $field) {
            print '<td>'.$row[$field].'\</td>';
        }

        print '</tr>';

        $cnt++;
    }
    print "</table>";
}

```



```

}

function clearCache() {
    unset($_SESSION['data_users']);
    unset($_SESSION['data_courses']);
    unset($_SESSION['data_actions']);
    unset($_SESSION['data_useractionstable']);
}

function getUsers() {
    if(isset($_SESSION['data_users'])) {
        $users = $_SESSION['data_users'];
    } else {
        $users = db_query("select distinct userid from events order by 1");
        $_SESSION['data_users'] = $users;
    }
    return $users;
}

function getCourses() {
    if(isset($_SESSION['data_courses'])) {
        $courses = $_SESSION['data_courses'];
    } else {
        $courses = db_query("SELECT course, COUNT( * ) FROM `events` GROUP
BY 1 ORDER BY 2 DESC");
        // $courses = db_query("select distinct course from events order by
1");
        $_SESSION['data_courses'] = $courses;
    }
    return $courses;
}

function getActions() {
    if(isset($_SESSION['data_actions'])) {
        $actions = $_SESSION['data_actions'];
    } else {
        $actions = db_query("select distinct action from events order by
1");
        $_SESSION['data_actions'] = $actions;
    }
    return $actions;
}

function getUserActionsTable() {
    if(isset($_SESSION['data_useractionstable'])) {
        $rez = $_SESSION['data_useractionstable'];
    } else {
        $rez = makeUserActionsTable();
        $_SESSION['data_useractionstable'] = $rez;
    }
    return $rez;
}

function makeUserActionsTable() {
    $users = getUsers();
    $actions = getActions();
    $stable = array();

    $header = array();
    $header[]="userid";
    foreach ($actions as $arow) {

```

```

        $action = $arow["action"];
        $header[]=$action;
    }

    $table[]=$header;

    foreach ($users as $row) {
        $userid = $row["userid"];

        $rez = array();

        $rez["userid"] = $userid;
        foreach ($actions as $arow) {
            $action = $arow["action"];
            $rez[$action] = 0;
        }
        $sql = "select action,count(*) as cnt from events where
userid=\"". $userid."\" group by 1";
        $events = db_query($sql);

        foreach ($events as $arow) {
            $action = $arow["action"];
            $actionCnt = $arow["cnt"];

            $rez[$action] = $actionCnt;
        }

        $table[]=$rez;
    }
    return $table;
}

function showEvents() {
    if (isset($_GET["userid"])) {
        $userid = $_GET["userid"];

        print "<a href=\"#ipTable\">Prisijungimai</a> ";
        print "<a href=\"#userActionsTable\">Veiksmi</a> ";
        print "<a href=\"#timeTable\">Veiksmų trukmės</a> ";
        print "<a href=\"#eventsTable\">Įvykiai</a> ";
        print "<br/>";

        showIPCounts($userid);
        showActionsCounts($userid);

        showUserEvents($userid);
    }

    if (isset($_GET["action"])) {
        if ($_GET["action"] == "view_students_actions") {
            $table = getUserActionsTable();

            print "<a name=\"userActions\"></a>";
            printBigTable($table, "Vartotojų veiksmi");
        }
    }
}

```

```

}

if (isset($_GET["action"])) {
    if ($_GET["action"] == "view_course_actions_counts") {
        $course = $_GET["course"];

        showCourseActionsCounts($course);
    }
}

/*
$sql = "select * from events";
$cnt = 0;
if (isset($_GET["userid"])) {
    $userid = $_GET["userid"];

    $cnt++;
    if ($cnt == 1) {
        $sql = $sql." where ";
    } else {
        $sql = $sql." and ";
    }
    $sql = $sql." userid=\"". $userid."\" ";
}

if (isset($_GET["course"])) {
    $course = $_GET["course"];
    $cnt++;
    if ($cnt == 1) {
        $sql = $sql." where ";
    } else {
        $sql = $sql." and ";
    }
    $sql = $sql." course=\"". $course."\" ";
}

$sql = $sql." limit 100 ";

$events = db_query($sql);
printEventsTable($events, "Įvykiai:");
*/

}

function caclTimes($events) {
    // make time
    $actions = getActions();
    $action_info = array();
    foreach ($actions as $arow) {
        $action = $arow["action"];
        $action_info[$action]["name"] = $action;
        $action_info[$action]["sum"] = 0;
        $action_info[$action]["cnt"] = 0;
    }
    $last = NULL;
}

```

```

foreach ($events as $erow) {
    if (isset($last)) {
        if ($last["ip"] == $erow["ip"]) {
            $timeDiff = $erow["time"]-$last["time"];
            $action = $last["action"];
            // perhaps check time diff

            $action_info[$action]["sum"] += $timeDiff;
            $action_info[$action]["cnt"] += 1;

        }
    }
    $last = $erow;
}

$user_actions_info = array();
foreach ($actions as $arow) {

    $action = $arow["action"];
    if ($action_info[$action]["cnt"] > 0) {
        $action_info[$action]["avg_time"] =
$action_info[$action]["sum"] / $action_info[$action]["cnt"];

        $action_info[$action]["avg_time"] =
number_format($action_info[$action]["avg_time"], 2);

        $user_actions_info[$action]=$action_info[$action];
    } else {
        $action_info[$action]["avg_time"] = 0;
    }
}

$header_name = array();
$header_name[]="name";
//$header_name[]="sum";
$header_name[]="cnt";
$header_name[]="avg_time";

$header_desc = array();
$header_desc[]="Veiksmas";
//$header_desc[]="Suma";
$header_desc[]="Kiekis";
$header_desc[]="Vid. trukmė, s";

print "<a name=\"timeTable\"></a>";
printGeneralTable($user_actions_info, $header_name, $header_desc, "Veiksmų
vidutinės trukmės");
}

function showCourseActionsCounts($course) {
    $sql = "select * from mdl_dm_action_summary where course=\"". $course."\"
order by userid";
    $rez = db_query($sql);

    $tableDesc = "Kurso ".$course." vartotojų veiksmų kiekiai";
}

```

```

$actions = getActionNames();

$header_name = array();
$header_desc = array();

$header_name[]="userid";
$header_desc[] = "Vartotojas";

$header_name[]="finalgrade";
$header_desc[] = "Balas";

foreach ($actions as $arow) {
    $action = $arow['action'];
    $header_name[]=$action."_cnt";
    $header_desc[]=$action;
}

printGeneralTable($rez, $header_name, $header_desc, $tableDesc);
//printBigTable($rez, $tableDesc);

/*
$k = array();
$k[]=array("k" => 2);
$k[]=array("k" => 3);
$k[]=array("k" => 4);
$k[]=array("k" => 5);
$k[]=array("k" => 6);

print "Klasterių: ";
printSelect($k, "k");
print "<br/>";
*/

$arff = "";
$arff .= "% ARFF file for moodle data\n";
$arff .= "% course ". $course . "\n";
$arff .= "@relation moodle_course_students_activity\n";
$arff .= "\n";
$arff .= "@attribute ". "userid"." numeric\n";
$arff .= "@attribute ". "finalgrade"." numeric\n";
foreach ($actions as $arow) {
    $action = $arow['action'];

    $arff .= "@attribute ". str_replace(" ", "_", $action)." numeric\n";
}

$arff .= "\n";
$arff .= "@data\n";
$arff .= "% ". sizeof($rez)." \n";
$arff .= "\n";
foreach ($rez as $line) {
    $dl = array();
    $dl[]= $line['userid'];
    $grade = $line['finalgrade'];
    if (empty($grade)) {
        $grade = -1;
    }
    //$grade = 1;

```

```

    $dl[] = $grade;
    foreach ($actions as $arow) {
        $action = $arow['action'];

        $dl[] = $line[$action."_cnt"];
    }

    $data = join(", ", $dl);

    $arff .= $data."\n";
}

//
$myFile = "rez/mdl_course_". $course. ".arff";
$fh = fopen($myFile, 'w') or die("can't open file");
fwrite($fh, $arff);
fclose($fh);

$k = 3;
//
$myRezFile = "rez/mdl_course_". $course. "_cluster_k". $k. ".txt";

$myFile00 = "rez/mdl_course_". $course. "_cluster_data.arff";
$myFile0 = "rez/mdl_course_". $course. "_cluster_data_norm.arff";

$myRezFile1 = "rez/mdl_course_". $course. "_cluster_assigments_k". $k. ".txt";

// $command0 = 'java -classpath weka.jar
weka.filters.unsupervised.attribute.Remove -R 1 < ' . $myFile . ' > ' . $myFile00 . '';
// exec($command0);

$command00 = 'java -classpath weka.jar weka.filters.MultiFilter -F
"weka.filters.unsupervised.attribute.Remove -R 1" -F
"weka.filters.unsupervised.attribute.Normalize -S 1.0 -T 0.0" -F
"weka.filters.unsupervised.attribute.RemoveUseless -M 99.0" < ' . $myFile . ' >
' . $myFile0 . '';
exec($command00);

$command = 'java -classpath weka.jar weka.clusterers.SimpleKMeans -N
' . $k . ' -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10 -t
' . $myFile0 . ' > ' . $myRezFile . '';
exec($command);

$command1 = 'java -classpath weka.jar weka.clusterers.SimpleKMeans -p 0 -N
' . $k . ' -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10 -t
' . $myFile0 . ' > ' . $myRezFile1 . '';
exec($command1);

// $arff = str_replace("\n", "<br/>", $arff);
print "ARFF duomenų failas: ";
print "<a href=\"". $myFile . "\">Parsisiūsti</a><br/>";

print "<textarea readonly rows=\"10\"";

```

```

cols="\80\>".$arff."</textarea><br/>";

$arff_rez = file_get_contents($myRezFile);
$arff_rez1 = file_get_contents($myRezFile1);

print "Klasteriavimo rezultatai: ";
print "<a href=\"\". $myRezFile.\"\">Parsisiųsti</a><br/>";

print "<textarea readonly rows=\"10\"
cols=\"80\>".$arff_rez."</textarea><br/>";

print "Priskyrimai klasteriams: ";
print "<a href=\"\". $myRezFile1.\"\">Parsisiųsti</a><br/>";

print "<textarea readonly rows=\"10\"
cols=\"80\>".$arff_rez1."</textarea><br/>";

}

function showUserEvents($userid) {
    $sql = "select * from events where userid=\"\". $userid.\"\" order by time";
    $events = db_query($sql);

    caclTimes($events);

    print "<a name=\"eventsTable\"></a>";
    printEventsTable($events, "Vartotojo ".$userid." įvykiai");
}

function showIPCounts($userid) {
    $sql = "select IP,count(*) as cnt from events where userid=\"\".
$userid.\"\" group by 1 order by 2";
    $events = db_query($sql);
    print "<a name=\"ipTable\"></a>";
    printCountsTable($events, "IP", "cnt", "Vartotojo ".$userid."
prisijungimai", "IP", "Kiekis");
}

function showActionsCounts($userid) {
    $sql = "select action,count(*) as cnt from events where userid=\"\".
$userid.\"\" group by 1 order by 2";
    $events = db_query($sql);
    print "<a name=\"userActionsTable\"></a>";
    printCountsTable($events, "action", "cnt", "Vartotojo ".$userid."
veiksmiai", "Veiksmas", "Kiekis");
}

function getActionNames() {
    $sql = "SELECT distinct ACTION as action FROM EVENTS";
    $rez = db_query($sql);
    return $rez;
}

function makeSummary() {
    $rez = getActionNames();

    $u_sql = "delete from mdl_dm_action_summary";
    //print $u_sql."<br/>";
}

```

```

mysql_query($u_sql);

    $u_sql = "insert into mdl_dm_action_summary (userid,course) (SELECT
distinct userid,course FROM events);";
    //print $u_sql."<br/>";
    mysql_query($u_sql);

    foreach ($rez as $row) {
        $u_sql = "ALTER TABLE `mdl_dm_action_summary` drop `".
$row['action'] . "`_cnt`";
        //print $u_sql."<br/>";
        mysql_query($u_sql);
    }

    foreach ($rez as $row) {
        $u_sql = "ALTER TABLE `mdl_dm_action_summary` ADD `".
$row['action'] . "`_cnt` INT NOT NULL DEFAULT '0'";
        //print $u_sql."<br/>";
        mysql_query($u_sql);
    }

    $sql = "SELECT userid, course, action, COUNT( * ) as cnt FROM `events`
GROUP BY 1 , 2, 3";
    $rez = db_query($sql);
    foreach ($rez as $row) {
        $action = $row['action']."_cnt";

        $u_sql = "UPDATE `mdl_dm_action_summary` SET `".
$row['cnt'] . "` where userid=".$row['userid'] . " and course=".$row['course'] . ";";
        //print $u_sql."<br/>";

        mysql_query($u_sql);
    }

    print "Duomenų apdorojimas baigas.<br/>";
}

function loadEvents() {

    print '<form action="index.php" method="post" enctype="multipart/form-
data">
    <input id="file_upload" name="file_upload" type="file" />
    Galima importuoti moodle mdl_log lentelės duomenis, CSV formatu.
    Importuojami tik unikalūs duomenys (id).
    <a href="./examples/mdl_log.csv">Pavyzdys</a>
    </form>';

    print "<form action=\"index.php\" method=\"GET\">";
    print "<input type=\"hidden\" name=\"action\" value=\"update_summary\"
/>";
}

```



```

print "<input type=\"submit\" value=\"Apdoroti duomenis\" />";
print "</form>";

if (isset($_GET["action"])) {
    if ($_GET["action"] == "update_summary") {
        makeSummary();
    }
}

print "<br/>";

$users = getUsers();
$actions = getActions();
$courses = getCourses();

$events_cnt_r = db_query("select count(*) as cnt from events;");
$events_cnt = $events_cnt_r[0]['cnt'];
print "Įrašų: ".$events_cnt."<br/>";
print "Vartotojų: ".sizeof($users)."<br/>";
print "Kursų: ".sizeof($courses)."<br/>";
print "Skirtingų veiksmų: ".sizeof($actions)."<br/>";
print "<br/>";

print "<form action=\"index.php\" method=\"GET\">";
print "<input type=\"hidden\" name=\"action\"
value=\"view_students_actions\" />";
print "<input type=\"submit\" value=\"Visų vartotojų veiksmai\" />";
print "</form>";

print "<form action=\"index.php\" method=\"GET\">";
print "<br/>";
print "Vartotojas: ";
printSelect($users, "userid");
print "<input type=\"submit\" value=\"Peržiūrėti vartotojo informacija\"
/>";
print "</form>";

print "<form action=\"index.php\" method=\"GET\">";
print "<br/>";
print "Kursas: ";
printSelect($courses, "course");
print "<input type=\"hidden\" name=\"action\"
value=\"view_course_actions_counts\" />";
print "<input type=\"submit\" value=\"Peržiūrėti kurso informacija\" />";
print "</form>";

}

```

```

loadEvents();
showEvents();

if (!empty($_FILES)) {
    $tempFile = $_FILES['Filedata']['tmp_name'];
    print $tempFile;
    $targetPath = $_SERVER['DOCUMENT_ROOT'] . $_REQUEST['folder'] . '/';
    $targetFile = str_replace('///','/', $targetPath) .
$_FILES['Filedata']['name'];

    $fileTypes = str_replace('*', '', $_REQUEST['fileext']);
    $fileTypes = str_replace(';','|', $fileTypes);
    $typesArray = split('|', $fileTypes);
    $fileParts = pathinfo($_FILES['Filedata']['name']);

    if (in_array($fileParts['extension'], $typesArray)) {
        // Uncomment the following line if you want to make the directory if
it doesn't exist
        // mkdir(str_replace('///','/', $targetPath), 0755, true);

        db_query2("LOAD DATA INFILE '".str_replace('\\', '/', $tempFile)."'
INTO TABLE events FIELDS TERMINATED BY ';' ENCLOSED BY '\"' LINES TERMINATED BY
'\r\n';");
        clearCache();
        move_uploaded_file($tempFile, $targetFile);
        echo str_replace($_SERVER['DOCUMENT_ROOT'], '', $targetFile);
        //
    } else {
        echo 'Blogas failo formatas.';
    }
}

?>

</body>

```