

KAUNO TECHNOLOGIJOS UNIVERSITETAS  
INFORMATIKOS FAKULTETAS  
INFORMACINIŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

TOMAS SKROBLAS

UNIVERSITETO STUDIJŲ KOKYBĖS IR STEBĖSENOS  
PROCESŲ TYRIMAS DUOMENŲ GAVYBOS PRIEMONĖMIS

Magistro darbas

Darbo vadovas  
doc. dr. V. Šakys

KAUNAS, 2013

KAUNO TECHNOLOGIJOS UNIVERSITETAS  
INFORMATIKOS FAKULTETAS  
INFORMACINIŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

TOMAS SKROBLAS

UNIVERSITETO STUDIJŲ KOKYBĖS IR STEBĖSENOS  
PROCESŲ TYRIMAS DUOMENŲ GAVYBOS PRIEMONĖMIS

Magistro darbas

Darbo vadovas  
doc. dr. V. Šakys  
2013-05-

Recenzentas  
doc. dr. A. Barila  
2013-05-

Atliko:  
IFM-1/4 gr. Studentas  
Tomas Skroblas  
2013-05-

KAUNAS, 2013

# AUTORIŲ GARANTINIS RAŠTAS

## DĖL PATEIKIAMO KŪRINIO

2013 - 05 - 24d.

Kaunas

**Autoriai,** \_\_\_\_\_ Tomas Skroblas \_\_\_\_\_  
(vardas, pavardė)

patvirtina, kad Kauno technologijos universitetui pateiktas baigiamasis bakalauro (magistro) darbas (toliau vadinama – Kūrinys)

\_\_\_\_\_  
Universiteto studijų kokybės ir stebėsenos procesų tyrimas duomenų gavybos priemonėmis  
(kūrinio pavadinimas)

pagal Lietuvos Respublikos autorių ir gretutinių teisių įstatymą yra originalus ir užtikrina, kad

- 1) jį sukūrė ir parašė Kūrinyje įvardyti autoriai;
- 2) Kūrinys nėra ir nebus įteiktas kitoms institucijoms (universitetams) (tiek lietuvių, tiek užsienio kalba);
- 3) Kūrinyje nėra teiginių, neatitinkančių tikrovės, ar medžiagos, kuri galėtų pažeisti kito fizinio ar juridinio asmens intelektualinės nuosavybės teises, leidėjų bei finansuotojų reikalavimus ir sąlygas;
- 4) visi Kūrinyje naudojami šaltiniai yra cituojami (su nuoroda į pirminį šaltinį ir autorių);
- 5) neprieštarauja dėl Kūrinio platinimo visomis oficialiomis sklaidos priemonėmis;
- 6) atlygins Kauno technologijos universitetui ir tretiesiems asmenims žalą ir nuostolius, atsiradusius dėl pažeidimų, susijusių su aukščiau išvardintų Autorių garantijų nesilaikymu;
- 7) Autoriai už šiame rašte pateiktos informacijos teisingumą atsako Lietuvos Respublikos įstatymų nustatyta tvarka.

**Autoriai**

\_\_\_\_\_  
Tomas Skroblas  
(vardas, pavardė)

\_\_\_\_\_  
(vardas, pavardė)

\_\_\_\_\_  
(vardas, pavardė)

\_\_\_\_\_  
(vardas, pavardė)

\_\_\_\_\_  
(parašas)

\_\_\_\_\_  
(parašas)

\_\_\_\_\_  
(parašas)

\_\_\_\_\_  
(parašas)

## SANTRAUKA

Kauno technologijos universitetui, kaip ir kitoms organizacijoms, svarbu žinoti, kaip pagerinti savo veiklą. Vienas iš svarbiausių veiksnių, gerinančių jo veiklą, yra studijų kokybė. Yra daugybė metodų tai patikrinti. Vienas iš jų – verslo analitikos priemonės bei galintis pagelbėti šių priemonių įrankis – duomenų gavyba.

Šiame darbe sprendimas yra sukurtas, naudojant SQL Server duomenų bazę ir jos paslaugas. Sukurta programinė įranga gali prognozuoti studijų rezultatus, numatyti, kurie studentai iškris iš universiteto. Taip pat analizuoti tekstą studijų kokybės anketose – suskaičiuoti vieno ar kito studijų dalyko kokybę. Tokia analizė yra vykdoma pagal frazių žodyną, kurį užpildo ekspertas. Sistemos vartotojo sąsaja yra realizuota, pasinaudojant ASP.NET karkaso internetinėmis formomis.

Atliekant šiuos tyrimus buvo suskaičiuota, kaip ilgai ekspertas užtrunka įvertindamas frazes žodyne, taip pat kaip tiksliai duomenų gavybos modeliai atlieka prognozes. Galiausiai yra realizuotas teksto spalvinimas, kuris parodo, kuriose vietose studijų kokybės anketoje yra šio žodyno frazės.

## **UNIVERSITY EDUCATION QUALITY AND MONITORING PROCESSES RESEARCH WITH DATA MINING TOOLS**

For Kaunas University of Technology, as well as other organizations, it is important to know how to improve their performance. One of the most important factors to improve the performance is the quality of education. There are many methods to verify this. One of the modern tools is business intelligence tools. It has a few instruments to explore data, one of which is data mining.

In this paper, solution is developed by using the SQL Server database and its services. Powered software can predict the outcomes of university studies and determine which students will drop out of the university. It can also analyze the text in the questionnaires of quality study – evaluate the quality of one or another subject. Such an analysis is carried out using a phrase dictionary which is filled by an expert. The system's user interface is implemented using ASP.NET web-forms.

The study also included the calculation of how long it takes an expert to assess the dictionary phrases, as well as how accurately data mining models make predictions. Finally, text coloring was implemented to indicate the dictionary phrases in a study quality form.

# TURINYS

<b>LENTELIŲ SĄRAŠAS .....</b>	<b>8</b>
<b>PAVEIKSLŲ SĄRAŠAS.....</b>	<b>9</b>
<b>TERMINŲ IR SANTRUMPŲ ŽODYNAS.....</b>	<b>11</b>
<b>ĮVADAS.....</b>	<b>12</b>
<b>1. UNIVERSITETO STUDIJŲ KOKYBĖS IR STEBĖSENOS PROCESŲ ANALIZĖ .....</b>	<b>15</b>
1.1. TYRIMO OBJEKTAS, SRITIS IR PROBLEMA .....	15
1.2. TYRIMO TIKSLAS IR UŽDAVINIAI.....	15
1.3. STUDIJŲ KOKYBĖS IR STEBĖSENOS PROCESŲ ANALIZĖS TIKSLAS .....	15
1.4. PASIRINKTI STUDIJŲ KOKYBĖS IR STEBĖSENOS PROCESŲ ANALIZĖS METODAI .....	15
1.5. STUDIJŲ KOKYBĖS ANALIZĖ UNIVERSITETE .....	16
1.6. VARTOTOJAI DALYVAUJANTYS STUDIJŲ KOKYBĖS IR STEBĖSENOS PROCESU.....	21
1.6.1. <i>Vartotojų aibė, tipai ir savybės.....</i>	<i>21</i>
1.6.2. <i>Vartotojų tikslai ir problemos.....</i>	<i>22</i>
1.7. ESAMŲ SPRENDIMŲ ANALIZĖ.....	24
1.7.1. <i>Šiuo metu atlikti darbai, naudojant duomenų gavybą.....</i>	<i>24</i>
1.7.2. <i>Šiuo metu siūlomi duomenų gavybos sprendimai .....</i>	<i>26</i>
1.7.3. <i>Sprendimai, naudojantys lietuvių kalbos atpažinimą.....</i>	<i>27</i>
1.8. DUOMENYS, PAGAL KURIŲS REIKIA KURTI SPRENDIMĄ .....	28
1.9. SIEKIAMAS SPRENDIMAS, GERINANTIS STUDIJŲ KOKYBĘ.....	29
1.10. RIZIKA, KURI KYLA, KURIANT DUOMENŲ GAVYBOS SPRENDIMUS .....	30
1.11. REZULTATO KOKYBĖS KRITERIJAI .....	30
1.12. ANALIZĖS IŠVADOS .....	31
<b>2. KURIAMOS SISTEMOS, GERINANČIOS KOKYBĖS IR STEBĖSENOS PROCESUS, REIKALAVIMAI 32</b>	
2.1. KURIAMO SPRENDIMO REIKALAVIMŲ SPECIFIKACIJA .....	32
2.1.1. <i>Kuriamo duomenų gavybos sprendimo funkcionalumo aprašymas .....</i>	<i>32</i>
2.1.2. <i>Kuriamo sprendimo studijų analizei nefunkciniai reikalavimai .....</i>	<i>42</i>
2.2. ANALIZUOJAMOS DALYKINĖS SRITIES MODELIS .....	43
2.3. DUOMENŲ GAVYBOS PRAPLĖTIMO KALBOS NAUDOJIMAS .....	44
2.4. REIKALAVIMŲ ANALIZĖS APIBENDRINIMAS.....	47
<b>3. UNIVERSITETO STUDIJŲ KOKYBĖS IR STEBĖSENOS PROCESUS GERINANČIOS SISTEMOS PROJEKTAS .....</b>	<b>47</b>
3.1. STUDIJŲ PROCESO TYRIMO SPRENDIMO REIKALINGUMAS .....	47
3.2. PANAUDOTI SPRENDIMAI KURIANT SISTEMĄ .....	48
3.2.1. <i>Sistemos loginė architektūra.....</i>	<i>48</i>
3.2.2. <i>Vartotojui atvaizduojama internetinė sąsaja .....</i>	<i>48</i>
3.2.3. <i>Duomenų gavybos struktūros bei modelio suformavimas, pasinaudojant ASP.NET karkasu bei DMX užklausomis .....</i>	<i>50</i>
3.2.4. <i>Duomenų gavybos rezultatų ištraukimas, pasinaudojant DMX užklausų kalba .....</i>	<i>54</i>
3.2.5. <i>Panaudotos duomenų bazės paslaugos.....</i>	<i>56</i>
3.2.6. <i>Teksto gavybos atlikimas, pasinaudojant integravimo paslaugomis .....</i>	<i>57</i>
3.2.7. <i>Teksto gavyba tiesiogiai ieškant frazių tekste .....</i>	<i>61</i>
3.3. FRAZIŲ ĮVERTINIMO REKOMENDACIJOS .....	62
3.4. TYRIMŲ ATLIKIMAS, PASINAUDOJANT SUKURTA SISTEMA.....	63
3.4.1. <i>Duomenų gavybos modelių ir struktūrų kūrimas, naudojantis sprendimu.....</i>	<i>63</i>
3.4.2. <i>Duomenų gavybos gautų modelių rezultatų peržiūra naudojantis sprendimu .....</i>	<i>66</i>
3.5. TYRIMAMS NAUDOJAMOS DUOMENŲ BAZĖS SCHEMA.....	68
<b>4. SUKURTO DUOMENŲ GAVYBOS SPRENDIMO ĮVERTINIMAS.....</b>	<b>75</b>
4.1. EKSPERIMENTO PLANAS .....	75
4.2. DUOMENŲ GAVYBOS ALGORITMŲ KŪRIMO IR TIKSLUMO PATIKRINIMAS.....	75
4.2.1. <i>Duomenų gavybos sukurto modelio atitikimas nurodytam formoje .....</i>	<i>75</i>
4.2.2. <i>Duomenų gavybos modelio tikslumo įvertinimas .....</i>	<i>77</i>
4.3. TEKSTO GAVYBOS ĮVERTINIMAS.....	79
4.3.1. <i>Naudojama metodika vertinti teksto gavybą.....</i>	<i>79</i>
4.3.2. <i>Eksperto darbo su frazių žodynu įvertinimas.....</i>	<i>80</i>

4.3.3.	<i>Integravimo paslaugų projekto pritaikymo tinkamumo įvertinimas .....</i>	<i>81</i>
4.3.4.	<i>Teksto gavybos taikymo, tiesiogiai ieškant frazių tekste, įvertinimas .....</i>	<i>82</i>
4.3.5.	<i>Teksto gavybos taikymo, kai žodynas yra papildytas eksperto įvertinimas.....</i>	<i>83</i>
4.3.6.	<i>Morfologiškai anotuoto teksto panaudojimo galimybė.....</i>	<i>84</i>
<b>5.</b>	<b>IŠVADOS.....</b>	<b>85</b>
	<b>LITERATŪRA.....</b>	<b>87</b>
	<b>PRIEDAI .....</b>	<b>89</b>

## LENTELIŲ SĄRAŠAS

1.1 lentelė. Vartotojų problemų aprašymas .....	23
1.2 lentelė. Duomenų gavybos sprendimų palyginimas .....	27
1.3 lentelė. Atsakymų į anketos klausimus skaičiai 2011 m. gruodžio mėn – 2012 m. sausio mėn laikotarpiu .....	29
2.1 lentelė. Panaudojimo atvejo „Tvarkyti teksto gavybos žodyną“ aprašymas .....	34
2.2 lentelė. Panaudojimo atvejo „Sukelti duomenų failą“ aprašymas .....	34
2.3 lentelė. Panaudojimo atvejo „Sukurti duomenų gavybos modelį“ aprašymas.....	35
2.4 lentelė. Panaudojimo atvejo „Apmokytį modelį“ aprašymas .....	36
2.5 lentelė. Panaudojimo atvejo „Suformuoti ataskaitą pagal duomenų gavybos rezultatus“ aprašymas.....	37
2.6 lentelė. Panaudojimo atvejo „Paimti tiriamus duomenis“ aprašymas.....	39
2.7 lentelė. Panaudojimo atvejo „Atlikti teksto analizę“ aprašymas .....	39
2.8 lentelė. Panaudojimo atvejo „Parodyti analizės rezultatą“ aprašymas .....	40
2.9 lentelė. Panaudojimo atvejo „Rodyti frazių buvimo vietas tekste“ aprašymas.....	40
2.10 lentelė. Panaudojimo atvejo „Peržiūrėti grafiškai duomenų gavybos modelį“ aprašymas .....	41
2.11 lentelė. Panaudojimo atvejo „Peržiūrėti ataskaitą“ aprašymas .....	41
2.12 lentelė. Panaudojimo atvejo „Išsaugoti ataskaitą“ aprašymas .....	42
2.13 lentelė. Duomenų gavybos užklausų kalbos panaudojamumo palyginimas su vartotojo sąsaja .	44
3.1 lentelė. Galimi frazių įvertinimai .....	62
3.2 lentelė. Apklausa lentelė .....	68
3.3 lentelė. Atestato_dalykai lentelė .....	69
3.4 lentelė. Stojimo_balai lentelė.....	69
3.5 lentelė. Dalykas lentelė .....	69
3.6 lentelė. Destytojas lentelė .....	69
3.7 lentelė. Individualus_planas lentelė .....	70
3.8 lentelė. Studentas lentelė.....	70
3.9 lentelė. Duomenų bazės klasifikatorių sąrašas.....	71
4.1 lentelė. Būsenų pasiskirstymas, kai prognozuojantys duomenys yra universiteto modulių pažymiai .....	78
4.2 lentelė. Būsenų pasiskirstymas, kai prognozuojantys duomenys yra universiteto dėstytojų parašyti įvertinimai .....	78
4.3 lentelė. Būsenų pasiskirstymas, kai prognozuojantys duomenys yra mokykloje gautų dalykų įvertinimai .....	79
4.4 lentelė. Būsenų pasiskirstymas, kai naudojami apjungti duomenys .....	79
4.5 lentelė. Frazių žodyno įvertinimo trukmės .....	80



## PAVEIKSLŲ SĄRAŠAS

1.1 pav. Universiteto studijų kokybės tikslai pagal Vidinį studijų kokybės užtikrinimo sistemos aprašą .....	16
1.2 pav. Procesų žemėlapis pagal vidinį studijų kokybės užtikrinimo sistemos aprašą .....	17
1.3 pav. Universiteto veiklos sąveikų modelis su išorinėmis organizacijomis ir visuomene .....	18
1.4 pav. Studijų procesų analizės ir tobulinimo procesas .....	20
1.5 pav. Studijų kokybės ir stebėsenos veiklos proceso veiklos objektai .....	21
1.6 pav. Strateginio planavimo ir kokybės departamento pareigybės.....	22
1.7 pav. Universiteto padaliniai susiję su studijų kokybės užtikrinimu.....	22
1.8 pav. Katedros studijų kokybės gerinimo panaudojimo atvejai .....	24
1.9 pav. Paprastam vartotojui matoma duomenų gavybos modelio struktūra .....	25
2.1 pav. Funkcijos, kurias turi turėti kuriamas sprendimas .....	33
2.2 pav. Duomenų gavybos modelio sukūrimo sekų diagrama .....	36
2.3 pav. Ataskaitos suformavimo ir išsaugojimo sekų diagrama.....	38
2.4 pav. Dalykinės srities modelis .....	43
3.1 pav. Panaudota sistemoje architektūra .....	48
3.2 pav. Administratoriaus navigavimas tarp sistemos puslapių .....	49
3.3 pav. Koordinatoriaus navigavimas tarp sistemos puslapių .....	49
3.4 pav. Senato studijų komisijos nario navigavimas tarp sistemos puslapių.....	50
3.5 pav. Analizės paslaugų objektai, skirti atlikti duomenų gavybą.....	51
3.6 pav. Duomenų gavybos struktūros ar modelio sukūrimas .....	52
3.7 pav. Duomenų gavybos struktūros sukūrimo DMX užklausa .....	52
3.8 pav. Duomenų gavybos modelio sukūrimo DMX užklausa .....	52
3.9 pav. Duomenų gavybos objektų sukūrimo sekos diagrama.....	53
3.10 pav. Apsikeitimas duomenimis, vykdant DMX užklausą.....	54
3.11 pav. Duomenų nuskaitymo iš duomenų gavybos modelio DMX užklausa .....	55
3.12 pav. Duomenų gavybos rezultatų gavimo sekų diagrama.....	56
3.13 pav. Analizės paslaugų panaudojimas .....	57
3.14 pav. Duomenų srautai paketuose, kurie skirti vykdyti teksto gavybą.....	57
3.15 pav. Dalis įvertintų žodžių .....	58
3.16 pav. Teksto gavybos integravimo projekto kontrolės srautas .....	59
3.17 pav. Žodyno sukūrimo duomenų srautas .....	59
3.18 pav. Suradimas, kurie žodžiai egzistuoja iš eksperto įvertinto frazių sąrašo .....	60
3.19 pav. Vieno klausimo įvertinimų radimo moduliams duomenų srautas.....	60
3.20 pav. Galutinio įvertinimų sąrašo sudarymo duomenų srautas .....	60
3.21 pav. Frazių spalvinimo atsakymų tekste algoritmas .....	61
3.22 pav. Analizės tipo bei modelio ir struktūros pavadinimų nurodymas.....	63
3.23 pav. Sistemos vaizdas pridėjus laukų į struktūrą .....	64
3.24 pav. Pasirinkimas algoritmo, kuriuo kuriamas modelis.....	65
3.25 pav. Laukai, kurie buvo kuriami anksčiau formose .....	65
3.26 pav. XML bylos pasirinkimo formos vaizdas naršyklėje .....	66
3.27 pav. Duomenų gavybos modelio pasirinkimo formos vaizdas naršyklėje.....	67
3.28 pav. Atributų pasirinkimo formos vaizdas internetinėje naršyklėje .....	67
3.29 pav. Analizės rezultatų lentelė pateikta interneto naršyklės pagalba.....	68
3.30 pav. Duomenų bazės schema be klasifikatorių .....	72
3.31 pav. Duomenų bazės lentelė Studentas su savo klasifikatoriais .....	73
3.32 pav. Visos lentelės su klasifikatoriais išskyrus Studentas.....	74
4.1 pav. Pasirinkti laukai, kurie įtraukiami į duomenų gavybos modelį, matomi formoje.....	76
4.2 pav. Duomenų gavybos modelis analizės duomenų bazėje .....	77
4.3 pav. Frazių spalvinimo forma internetinėje sąsajoje.....	80
4.4 pav. Profesinės kalbos kultūros dalyko teigiamų aspektų įverčiai gauti integravimo paslaugų projektu .....	81

4.5 pav. Profesinės kalbos kultūros teigiamų aspektų įverčiai, tiesiogiai ieškant frazių žodyne.....	82
4.6 pav. Matematika 1 modulio neigiamų aspektų tekstai kai naudojamas ne eksperto sudarytas žodynas.....	83
4.7 pav. Matematika 1 modulio neigiamų aspektų tekstai kai žodyną yra papildęs ekspertas .....	83
4.8 pav. Morfologiškai anotuoto teksto panaudojimas teksto analizei .....	84

## TERMINŲ IR SANTRUMPŲ ŽODYNAS

**Vskus** – vidinė studijų kokybės universitete sistema.

**SQL** – standartinė kalba, skirta pasiekti duomenis iš duomenų bazės valdymo sistemos.

**Business Intelligence Development Studio** – Microsoft sukurta programinė įranga, veikianti kaip Microsoft Visual Studio programos dalis, skirta kurti verslo analitikos projektus.

**Microsoft Visual Studio** – galinga programinės įrangos kūrimo aplinka, kurios dėka galima kurti internetinius puslapius ir taikomąsias programas kompiuteriams ir išmaniesiems telefonams.

**.NET** – Microsoft sukurtas karkasas, suteikiantis galimybę naudotis daugybe bibliotekų, kurių pagalba galima lengvai kurti tokias įvairias programinės įrangos dalis, kaip formas, duomenų bazių komponentus ir panašiai.

**OLAP kubai** – daugiadimensinės duomenų struktūros, kuriose saugomi duomenys iš daugelio duomenų šaltinių, apibrėžiami savo matais ir dimensijomis.

**C#** - Microsoft sukurta objektiškai orientuota programavimo kalba, naudojama kaip viena iš .NET karkaso programavimo kalbų.

**DMX** – užklausų kalba, kuri naudojama paimti duomenis iš duomenų gavybos modelių ar struktūrų bei manipuluoti jais.

**XML** – bendros paskirties duomenų struktūrų bei jų turinį aprašanti kalba, dažniausiai skirta užtikrinti lengvą keitimąsi duomenimis tarp sistemų.

**ASP** – Microsoft išleistas internetinių taikomųjų programų karkasas, skirtas kurti dinامينius turinio internetinius sprendimus.

## IVADAS

Kiekviena organizacija siekia sėkmingai vykdyti savo veiklą. Tokį patį tikslą įgyvendina ir Kauno technologijos universitetas. Jo veiklos efektyvumą lemia tokie kriterijai, kaip studijuojamų dalykų dėstymo vertinimas, studentų mokymosi rezultatai, darbdavių atsiliepimai apie universitetą pabaigusius absolventus, vertinimas, kaip gerai yra supažindinami studentai su nauja medžiaga, ir daugybė kitų faktorių.

Dažnai universitete girdimas klausimas, kokios kokybės yra studijos. Norint tai sužinoti, reikia atlikti kokybės analizę. Ji daugelyje organizacijų paprastai atliekama, nenaudojant pažangių verslo analitikos metodikų, todėl jos rezultatuose gali įsivelti klaidų, ir neatsispindėti tikra situacija. Norint to išvengti, galima naudoti įvairius naujus metodus, kaip pavyzdžiui, duomenų gavyba. Šio metodo dėka galima iš anksto numatyti galimas problemas, tokias, kaip labai žemi studentų vidurkiai, ir imtis priemonių jas pašalinti.

Norint pasinaudoti šiomis galimybėmis, reikia pasinaudoti informacija, kurią turi universitetas: akademinė studentų informacija, kurioje matomi asmeniniai studento duomenys, modulių įvertinimai, stojamųjų brandos egzaminų balai. Taip pat reikia pasinaudoti studentų pildomomis universitete dėstomų modulių kokybės įvertinimo anketų rezultatais. Šiose anketose yra atviro tipo klausimų, kuriuose studentai nurodo modulio teigiamus, neigiamus aspektus bei parašo pasiūlymus, kaip pagerinti dėstomą dalyką. Turint šiuos anketų duomenis, yra galimybė analizuoti juose atviro tipo klausimus, pasinaudojant teksto gavyba.

Šio darbo tikslas – pagerinti studijų kokybę. Tai galima pasiekti sukūrus sprendimą, kuris padėtų vartotojui, naudojantis pažangiomis informacinėmis technologijomis, tirti turimus duomenis ir gauti dėsningumus, kurių dėka galima padaryti atitinkamas išvadas apie universiteto studijų kokybę ir visą veiklą.

Minėtuosius tyrimus galima atlikti, pasinaudojant Microsoft SQL Server duomenų bazių valdymo sistema. Joje yra sukurtas ne tik pats duomenų bazių kūrimo įrankis, bet ir specialios dalys vadinamos paslaugomis (*angl. Services*), kurios kiekviena skirta tam tikrai veiklai su duomenimis atlikti: analizės paslaugos skirtos saugoti tokius analizės objektus, kaip duomenų gavybos struktūra, modelis, OLAP kubas ar dimensija. Kita paslauga skirta transformuoti, perkelti ar kitaip modifikuoti duomenis.

Vartotojo sąsajai kurti galima pasinaudoti ASP.NET karkasu. Jo pagalba patogiai kuriami internetiniai puslapiai. Šią technologiją patogiu naudoti, nes ji yra sukurta taip pat Microsoft kompanijos, todėl labai gerai suderinta su anksčiau minėta SQL Server duomenų bazių valdymo sistema.

Šias technologijas reikia priderinti su universiteto naudojamu Vidiniu studijų kokybės užtikrinimo sistemos aprašu, kuriame yra įvardyti universitete vykstantys veiklos procesai,

padaliniai, atsakingi už studijų kokybę, ir kiti dokumentai, pagrindžiantys vykstančius veiklos procesus. Taip pat reikia atsižvelgti į tik ką (2013 metais) prasidėjusią Kauno technologijos universiteto mokslo ir studijų atsinaujinimo programą, kurios dėka pradėtas vykdyti universiteto veiklos reorganizavimas.

Peržiūrėjus visus šiuos dokumentus, nuspręsta kurti programinę įrangą trimis vartotojų tipams: Administratoriui, Koordinatoriui ir Senato studijų komisijos nariams. Kiekvienas iš šių vartotojų gali atlikti sprendime tokias funkcijas, kaip duomenų gavybos struktūrų ar modelių kūrimas, apmokymas, naujai paduotų duomenų prognozavimas, studijų kokybės anketų tekstų analizė bei ataskaitos formavimas ir peržiūra.

Sprendimas neturėtų sudaryti galimybių vartotojams, kurie neturi kompetencijos, kurti verslo analitikos sprendimus. Darbuotojai turėtų matyti jiems skirtus analizės rezultatus, naudojantis interneto naršyklės pagalba, nenaudojant specialių sprendimų kūrimo priemonių. Sukurtoje sąsajoje yra tik tie elementai, kurie yra būtini atlikti funkciją, kurią vykdo tam tikra sprendimo dalis.

Norint sėkmingai sukurti anksčiau aprašytą sprendimą, reikėjo išanalizuoti nemažai literatūros. Pirmiausia reikėjo išsiaiškinti, kaip Kauno technologijos universitete vykdoma studijų kokybės stebėseną. Tam reikėjo išnagrinėti įvairius dokumentus, kurie reglamentuoja universiteto darbą[1, 2, 3, 7, 8]. Kaip buvo paminėta anksčiau, šiuo metu vyksta universiteto veiklos reorganizavimas. Todėl teko susipažinti su šios programos dokumentais.[4, 5, 6].

Taip pat reikėjo išsiaiškinti, kaip duomenų gavyba yra taikoma užsienyje atliktuose tyrimuose. Tam buvo susipažinta su straipsniais, parašytais užsienio universitetuose[11, 12, 13]. Nors ir pasirinkta naudoti Microsoft produktą, reikėjo išsiaiškinti, ką siūlo ir kitos kompanijos. Tam buvo analizuojama jų pateikta informacija[14, 16, 17].

Kuriant sprendimą reikėjo išsiaiškinti, kaip naudoti daugybę naujų technologijų, kurias sukūrė Microsoft, tokių, kaip DMX užklausų kalba, AMO biblioteka, SQL Server integravimo paslaugų transformacijos ir kitomis. Tam reikėjo nagrinėti literatūrą apie šių technologijų naudojimą, aiškintis, kaip spręsti susidariusias tam tikras technologines problemas ir kitus panašius klausimus[15, 23, 22, 24, 25].

Visas darbas sudarytas iš keturių didelių dalių. Pirmoje analizės dalyje yra nagrinėjama, kaip Kauno technologijos universitete yra atliekama studijų kokybės ir stebėsenos procesas, surandami jo svarbiausi padaliniai ir jų darbuotojai bei jiems iškylančios problemos. Palyginami skirtingų kūrėjų duomenų gavybos produktai bei peržvelgiami užsienio universitetų atlikti darbai.

Antrame skyriuje yra pateikiami reikalavimai, kurie keliami sprendimui. Čia išryškėja, kokie veikėjai sistemoje, kurias funkcijas atliks. Pateikiami panaudojimo atvejai, kurie aprašo visas sistemoje vykdomas funkcijas, specifikacijas. Taip pat yra pateikiami nefunkciniai reikalavimai, kuriuos sistema turi atitikti. Šioje dalyje išryškėja duomenų esybės, kurios vėliau pavirs į duomenų

bazės lentelės. Galiausiai yra aprašomi privalumai, kodėl reikia naudoti DMX užklausų kalbą, o ne SQL Server pakete esančias grafines vartotojo sąsajos priemones.

Trečioje dalyje aprašoma, kokia architektūra yra realizuota visa sistema: parodoma, iš kokių internetinės sąsajos dalių susideda bei pateikiama duomenų bazės schema. Apžvelgiama AMO biblioteka, kurios pagalba galima kurti ar atlikti pakeitimus su SQL Server analizės paslaugų serverio objektais, ir DMX užklausų kalbos užklausos struktūra, kurios dėka galima gauti prognozės rezultatus iš apmokyto duomenų gavybos modelio. Taip pat aprašyta metodika, naudojama atlikti teksto gavybą, naudojantis Microsoft SQL Server integravimo paslaugų srautais arba tiesiogiai ieškant frazių iš žodyno tekste. Pabaigoje pateikti nurodymai, kaip dalykinės srities ekspertas turėtų įvertinti žodyne esančias frazes.

Paskutinėje ketvirtoje dalyje yra aprašyti atlikti eksperimentai su sprendimu. Pirmiausia patikrinta, ar AMO bibliotekos pagalba tikrai įmanoma sukurti teisingą duomenų gavybos modelį ir struktūrą. Toliau patikrinta, kaip tiksliai modelis sugeba atlikti prognozę. Išbandyti trys modeliai, prognozuojantys studento būseną universitete. Toliau yra pateikti eksperimentai su teksto gavyba. Nurodyta, kiek laiko ekspertas užtrunka įvertindamas frazes, pateikta, kaip jos surandamos universiteto studijų anketų atsakymų tekstuose. Tai atliekama pirmiausia, kai naudojama SQL Server integravimo paslaugos, toliau, kai ieškamos frazės sudarytos jau minėtosios paslaugos žodyne tiesiogiai, ir galiausiai, kai žodyną papildė ekspertas.

# **1. UNIVERSITETO STUDIJŲ KOKYBĖS IR STEBĖSENOS PROCESŲ ANALIZĖ**

## **1.1. Tyrimo objektas, sritis ir problema**

Tyrimo objektas - studijų kokybės analizė, kuri atliekama, siekiant įvertinti, kaip gerai yra organizuojamos studijos.

Tyrimo sritis - Kauno technologijos universiteto studijų kokybės veiklos procesų kompiuterizavimas, naudojant duomenų gavybos priemones.

Tyrimo problema yra ta, kad, pareikalavus Lietuvos Respublikos valdžios institucijoms ar aukštesnei universiteto valdžiai atlikti tam tikrą analizę, ji atliekama, nenaudojant modernių duomenų gavybos priemonių. Todėl įsivelia klaidų, gali nepavykti gauti reikalaujamų rezultatų, ir patys rezultatai vaizduojami nepatogia forma, kurią yra sudėtinga interpretuoti. Šias problemas galima išspręsti, naudojant duomenų gavybos algoritmus ir modernias ataskaitų kūrimo priemones.

## **1.2. Tyrimo tikslas ir uždaviniai**

Tyrimo tikslas - pagerinti studijų kokybės analizės procesą. Norint pasiekti šį tikslą, reikia:

1. Išanalizuoti duomenų gavybos įrankius ir pasirinkti vieną tinkamą.
2. Sukurti informacinę sistemą, kurioje būtų galima atlikti abiturientų ir studentų studijų kokybės analizę, pritaikius duomenų gavybos algoritmus.
3. Sukurti priemones patogiai peržiūrėti analizės rezultatus.

## **1.3. Studijų kokybės ir stebėsenos procesų analizės tikslas**

Analizės tikslas – surasti kurie Kauno technologijos universiteto struktūroje esantys padaliniai yra atsakingi už studijų kokybę ir stebėseną bei kurioms pareigybėms šiuose padaliniuose priskirta atlikti šias funkcijas. Taip pat išanalizuoti dabar taikomą studijų kokybės ir analizės procesą ir parodyti, kas pasikeis naudojant duomenų gavybos priemones. Galiausiai palyginama keletas galimų skirtingų produktų, su kuriais galima atlikti duomenų analizę.

## **1.4. Pasirinkti studijų kokybės ir stebėsenos procesų analizės metodai**

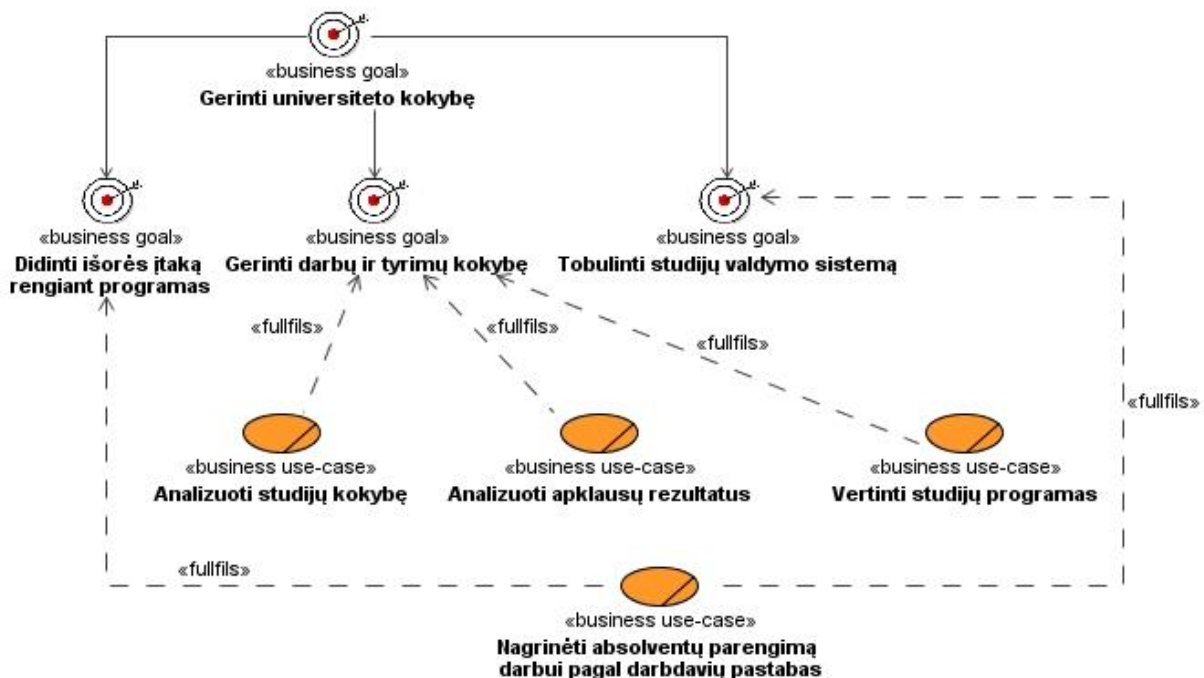
Surasti vartotojus naudojama dokumentų turinio analizė, peržiūrint dokumentus, kuriuose pateikta Kauno technologijos universiteto pareigybių aprašai, tyrimo objekto analizei nagrinėjamas universiteto Vidinės studijų kokybės užtikrinimo sistemos (VSKUS) aprašas. Sprendimų analizėje bus palyginti skirtingų kūrėjų duomenų gavybai atlikti siūlomi produktai bei dabar sukurti ir naudojami duomenų gavybos sprendimai.

## 1.5. Studijų kokybės analizė universitete

Atliekant analizę Kauno technologijos universitete yra vadovaujama Vidinės studijų kokybės užtikrinimo sistemos aprašu. Šis aprašas buvo parengtas 2010 metais ir remiasi Europos Sąjungos aukštojo mokslo dokumentais ir. Pagal aprašą[1] pagrindiniai universiteto tikslai turi būti:

- Didinti išorinių asmenų įtaką, kuriant naujas arba tobulinant esamas studijų programas.
- Gerinti atliekamų mokslinių tyrimų darbų kokybę ir didinti jų apimtį bei atnaujinti tyrimų bazę.
- Atliekant studijų analizę, nuolat tobulinti studijų valdymo sistemą.

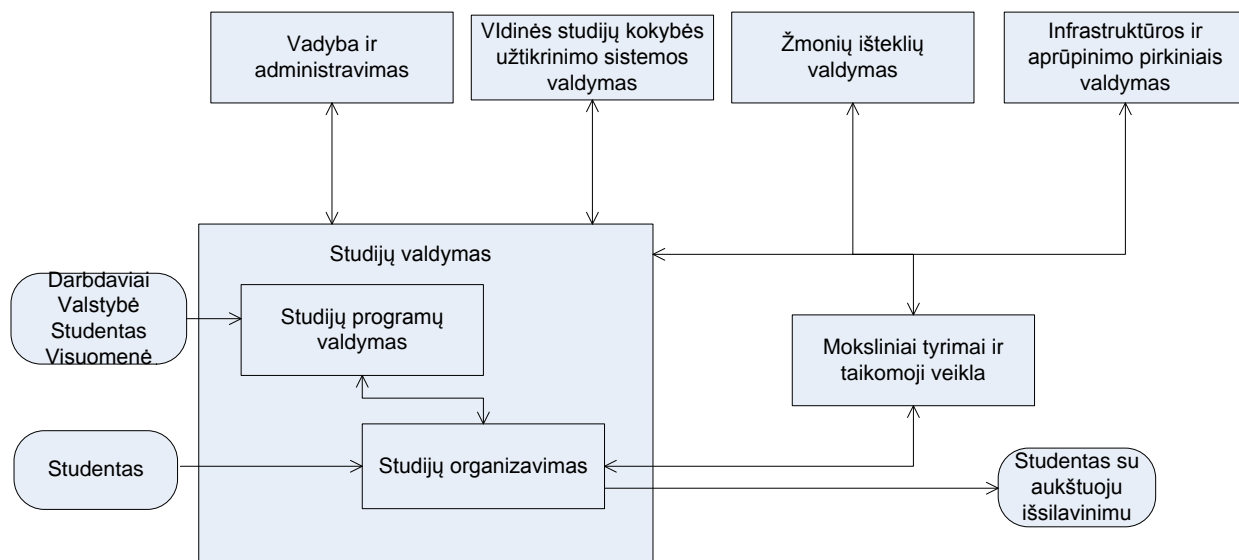
Apibendrintai šiuos tikslus galima parodyti tikslų modeliu, kuris pavaizduotas 1.1 paveiksle, kartu parodant, kurie panaudojimo atvejai šiuos tikslus išpildo. Modelyje rodomi tik tie tikslai, kurie yra aktualiausi ir vykdomi katedrų arba Strateginio planavimo ir kokybės departamento Kokybės skyriaus.



1.1 pav. Universiteto studijų kokybės tikslai pagal Vidinį studijų kokybės užtikrinimo sistemos aprašą

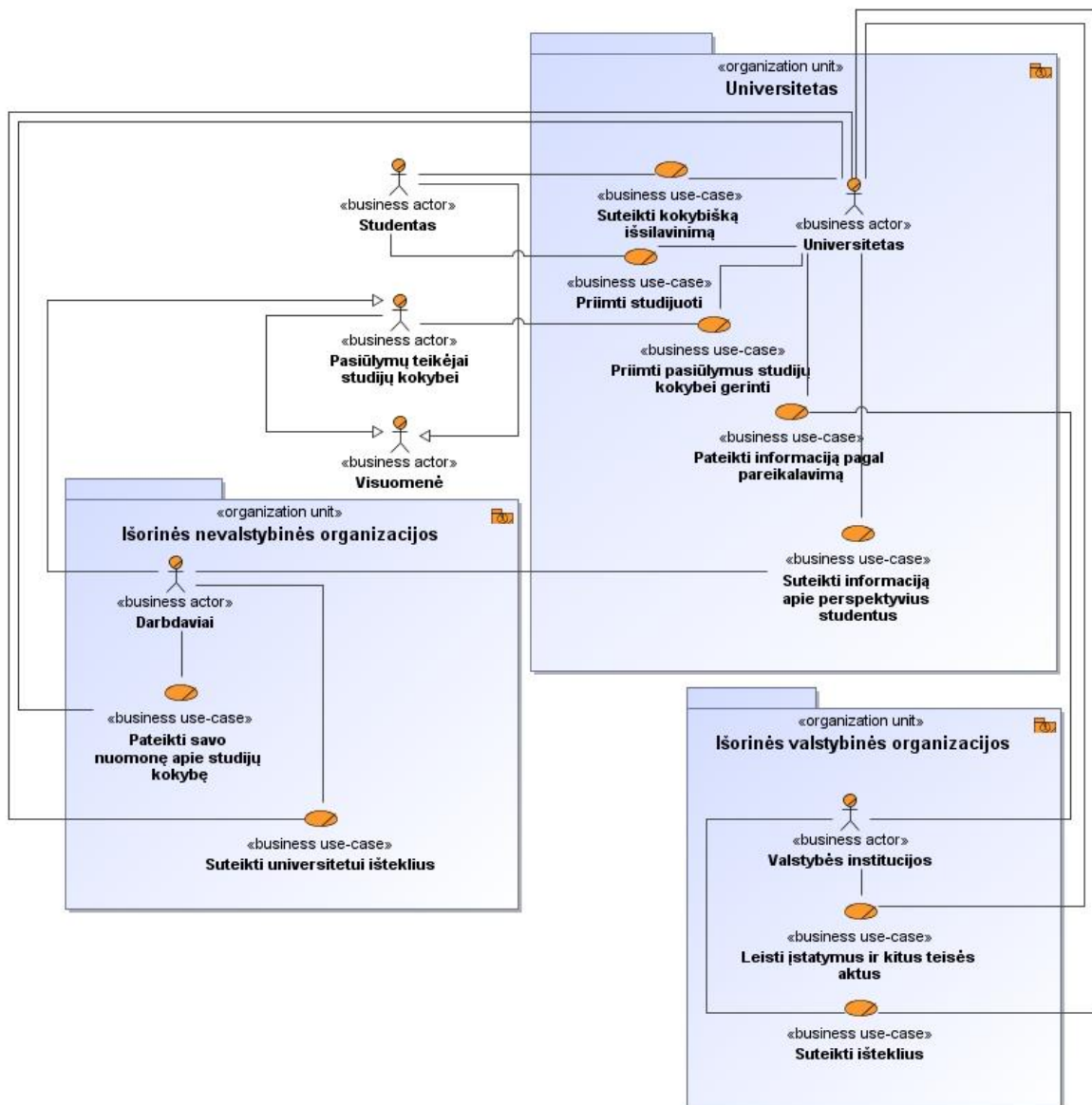
Viena iš aprašo dalių yra viso universiteto procesų žemėlapis, kuriame galima matyti, kaip tarpusavyje siejasi visos universiteto atliekamos veiklos. Šis žemėlapis pavaizduotas 1.2 paveiksle.





**1.2 pav.** Procesų žemėlapis pagal vidinį studijų kokybės užtikrinimo sistemos aprašą

Studijų programų valdymo procesui įtakos turi darbdaviai, valstybinės įstaigos, studentas ir visuomenė, o studijų organizavimui - tik studentas. Visų procesų galutinis rezultatas – studentas, turintis aukštąjį išsilavinimą. Šių išorinių organizacijų įtaką universitetui galima pavaizduoti veiklos sąveikos modeliu. Universitetas iš įvairių valstybinių ar nevalstybinių organizacijų gali gauti paramą, taip pat sužinoti jų nuomonę apie studijų kokybę. Studentai įstoja į universitetą, jame mokosi ir galiausiai gauna aukštojo mokslo diplomą. Iki tol, kol studentas pabaigia studijas, yra keičiamasi informacija su darbdaviais apie perspektyvius studentus bei sužinoma, ką pasiekė jau baigę studijas absolventai. Taip pat universitetas turi laikytis visų teisės aktų, kuriuos leidžia Lietuvos Respublikos valstybinės institucijos. Visus šiuos ryšius su išoriniais veikėjais galima pavaizduoti veiklos sąveikų modeliu, kuris pavaizduotas 1.3 paveiksle.



**1.3 pav.** Universiteto veiklos sąveikų modelis su išorinėmis organizacijomis ir visuomene

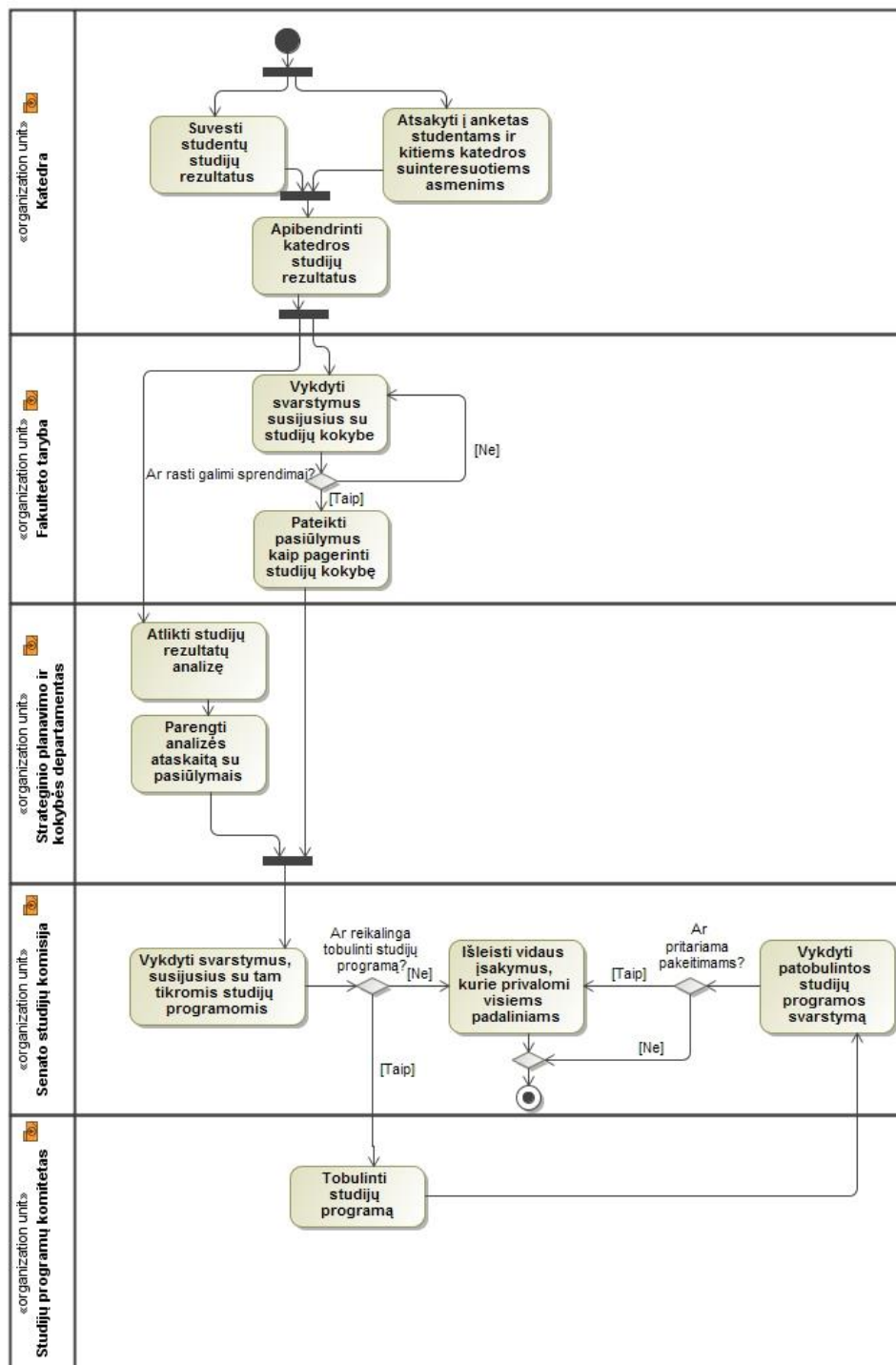
Iš procesų žemėlapiu, kuris buvo pateiktas anksčiau, svarbiausia šio darbo kontekste procesų grupė yra Vidinės studijų kokybės užtikrinimo sistemos valdymas. Šios grupės procesai yra:

- Studijų kokybės politikos ir tikslų nustatymas ir vertinimas.
- Studijų procesų pasiekimų vertinimas.
- Studijų procesų analizė ir tobulinimas (pavyzdinės praktikos skleidimas).
- Studijų proceso dokumentacijos vedimo ir duomenų įrašų fakultetuose valdymo stebėseną.
- Suinteresuotų šalių apklausų organizavimas ir rezultatų analizė.

Pagal Kauno technologijos universiteto Statutą[2] už pirmąjį procesą atsakingas yra universiteto Senatas, kuris turi atlikti studijų kokybės užtikrinimo sistemos kontrolę, stebėti kaip ji įgyvendinama. Taip pat studijų kokybės vertinimą atlieka ir Strateginio planavimo ir kokybės departamento Kokybės skyrius bei katedros. Dokumentavimo tvarkymu užsiima fakultetai, Strateginio planavimo ir kokybės departamento Kokybės skyrius. Esant reikalui Rektoratas gali

sudaryti darbo grupę, kuri atlieka patikrinimą visų padalinių, kurie atsakingi už dokumentų tvarkymą. Tokie padaliniai yra Informacijos technologijų departamentas, Studijų departamentas, fakultetų raštinės ir archyvas. Gali būti tikrinama dokumentacijos vedimo tvarka fakultetuose ir akademinės sistemos duomenų bazės įrašų patikimumas[3]. Visi anksčiau paminėti universiteto padalinių pavadinimai yra pakeisti po pradėto reorganizavimo pagal universiteto rektoriaus įsakymą[4].

Pagal universiteto statutą[2] ir Vidinės studijų kokybės užtikrinimo sistemos aprašą[1], vykdant studijų procesų analizę ir tobulinimą, pirmiausia apibendrinami visi turimi rezultatai katedrose, po to fakulteto taryba teikia pasiūlymus Senato studijų komisijai, kaip reikėtų tobulinti studijų programas ir studijų procesus. Tuo pačiu metu Strateginio planavimo ir kokybės departamento Kokybės skyrius atlieka studijų rezultatų analizę ir ją taip pat pateikia Senato studijų komisijai. Jis gali Studijų programų komitetui nurodyti patobulinti studijų programą arba išleisti teisės aktą, keičiantį kai kuriuos studijų proceso kai kuriuos elementus. Šis procesas pavaizduotas 1.4 paveiksle.



1.4 pav. Studijų procesų analizės ir tobulinimo procesas

Šis procesas ateityje gali keistis, nes universitete nuo 2013 metų pradžios pradėta vykdyti Kauno technologijos universiteto mokslo ir studijų atsinaujinimo programa[5]. Šios programos prirėikė todėl, kad:

- Paskutinius kelis metus sparčiai mažėja studijuojančių studentų skaičius.
- Neoptimaliai išnaudojama turima universiteto infrastruktūra, turimos patalpos.
- Mažėja jaunų mokslininkų skaičius.
- Informacinės sistemos labiausiai pritaikytos administracijai, bet ne studentams ar dėstytojams.

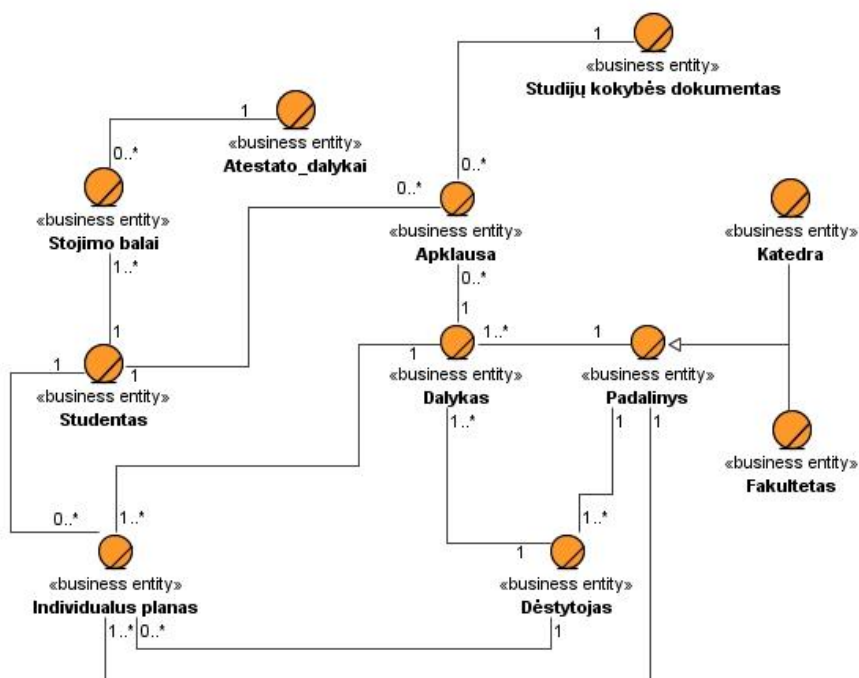
Pagal šią programą bus vykdoma tolesnė universiteto strategija 2012-2020 metų laikotarpyje.

Visa programa apims tris kryptis:

- Sieks, kad studijos būtų patrauklesnės.
- Įvykdyt universiteto struktūros pertvarką.
- Pagerins universiteto infrastruktūros efektyvumą.

Šios programos įgyvendinimo laikotarpyje bus atliekami padalinių atnaujinimo ar reorganizavimo, naujų tvarkų, nuostatų ir kitų svarbių teisės aktų rengimo darbai. Šios programos vykdymo planą parengė rektorius specialiai sudaryta darbo grupė[6].

Anksčiau pateikto dabartinio studijų proceso metu yra atliekami veiksmai su studentų duomenimis, jų įvertinimais mokykloje ir universitete, naudojamosi studentų atsakytomis anketomis bei informacija, kurie dėstytojai kokius modulius dėstė. Atlikus analizę sudaroma ataskaita, kuri apibendrintai pavadinta studijų kokybės dokumentu. Šie veiklos objektai yra parodyti 1.5 paveiksle.



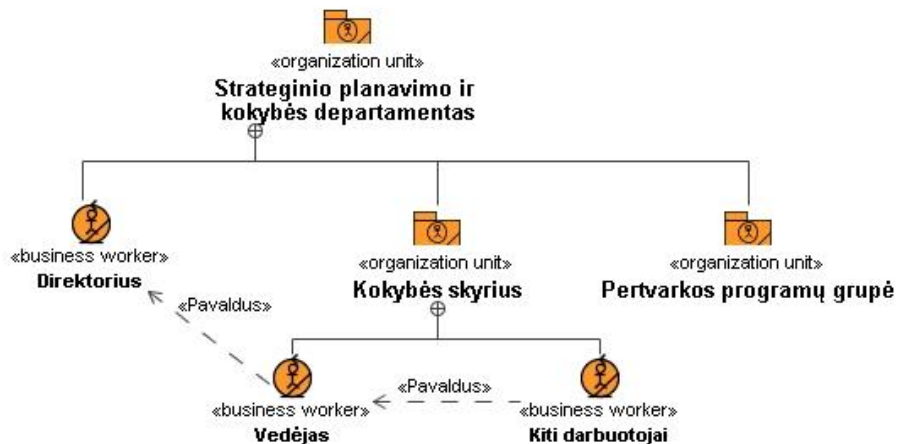
1.5 pav. Studijų kokybės ir stebėsenos veiklos proceso veiklos objektai

## 1.6. Vartotojai dalyvaujantys studijų kokybės ir stebėsenos procese

### 1.6.1. Vartotojų aibė, tipai ir savybės

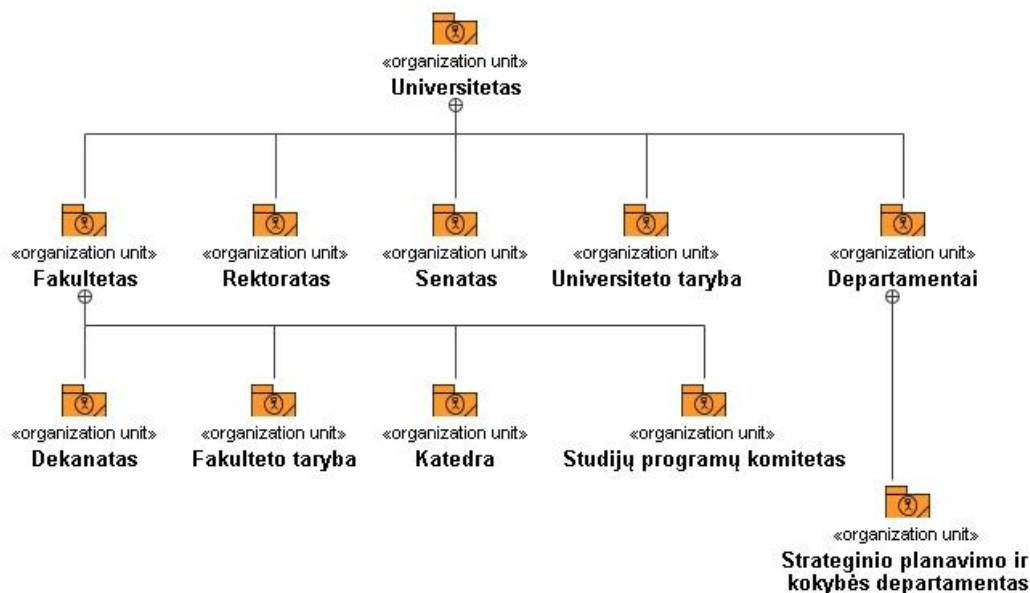
Vienas iš Kauno technologijos universiteto padalinių yra Strateginio planavimo ir kokybės departamentas, kuris inicijuoja ir koordinuoja Universiteto veiklos kokybės gerinimo veiklas ir projektus, organizuoja ir vykdo Universiteto bakalauro, magistrantūros ir doktorantūros studijų kokybės stebėseną, koordinuoja vidinės studijų kokybės vertinimo sistemos veiklą[7]. Šis departamentas sudarytas iš Kokybės skyriaus ir Pertvarkos programų grupės. Kadangi 2012 metų balandžio mėnesį dar nebuvo paskelbti naujai reorganizuotų padalinių pareigybių aprašymai, toliau vadovaujamosi tokia padalinio struktūra, kokia buvo iki reorganizacijos. Anksčiau buvusi Strategijos

įgyvendinimo ir kokybės užtikrinimo tarnyba yra sudaryta iš skyriaus vedėjo, kuris pavaldus tarnybos direktoriui, ir kitų darbuotojų. Šios pareigybės su nauju departamento pavadinimu pavaizduotos 1.6 paveiksle.



1.6 pav. Strateginio planavimo ir kokybės departamento pareigybės

Vidinės studijų kokybės užtikrinimo sistemos apraše yra minimi ir kiti universiteto padaliniai. Kai kurie iš padalinių priima sprendimus, susijusius su tuo, kaip turės atlikti kiti padaliniai savo veiklą. Kaip pavyzdį tokio padalinio, galima laikyti Senatą. Kiti, kaip ir jau minėtasis Strateginio planavimo ir kokybės departamento Kokybės skyrius, atlieka studijų kokybės įvertinimą. Visi padaliniai, kurie atlieka funkcijas, susijusias su studijų kokybės užtikrinimu, pavaizduoti 1.7 paveiksle.



1.7 pav. Universiteto padaliniai susiję su studijų kokybės užtikrinimu

### 1.6.2. Vartotojų tikslai ir problemos

Visi darbuotojai, kurie dirba Strateginio planavimo ir kokybės departamento Kokybės skyriuje turi atlikti funkcijas, aprašytas nuostatuose, kai jie bus parengti. Nuostatai bus sudaryti pagal

dabar egzistuojančios tarnybos[7] bei skyriaus nuostatus[8]. Kadangi juose nėra tiksliai nurodyta, kurie darbuotojai kokias funkcijas atlieka, jos pateiktos sąrašu:

- Vykdo Strateginio planavimo ir kokybės departamento direktoriaus nurodymus.
- Rengia Universiteto lygmens veiklos ataskaitas Švietimo ir mokslo ministerijai ir Statistikos departamentui.
- Inicijuoja ir koordinuoja Universiteto veiklos kokybės gerinimo veiklas ir projektus.
- Organizuoja ir vykdo Universiteto bakalauro, magistrantūros ir doktorantūros studijų kokybės stebėseną; rengia su šia veikla susijusias ataskaitas.
- Koordinuoja vidinės studijų kokybės vertinimo sistemos veiklą.
- Rengia fakultetams ir kitiems padaliniam reikalingų veiklos ataskaitų dokumentų formas.
- Nustatyta tvarka rengia ir teikia rektoriui veiklos vykdymo ataskaitas.

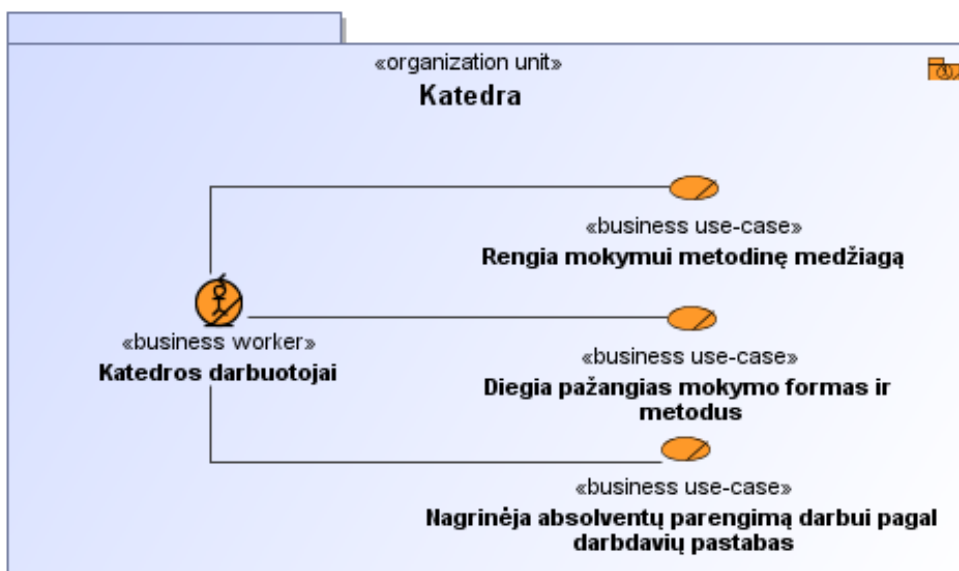
Specialistai, kurie dirba Strateginio planavimo ir kokybės departamento Kokybės skyriuje, šiuo metu nenaudoja pažangių verslo analitikos priemonių, todėl susiduria su problemomis, kurios matomos 1.1 lentelėje.

**1.1 lentelė.** Vartotojų problemų aprašymas

<b>Problema</b>	<b>Kaip viskas vyksta dabar, problemos aprašymas</b>
Analizės metu negaunama informacija apie ateitį.	Analizė atliekama praėjus semestru su jau praėjusio laikotarpio duomenimis, išskritus daliai studentų
Nenaudojami pažangūs duomenų gavybos algoritmai, todėl įsivelia įvairių netikslumų	Naudojant standartines priemones nepamatomi įvairūs dėsningumai duomenyse, kuriuos gali pastebėti duomenų gavybos algoritmai
Nenaudojamos vaizdžios SQL Serverio ataskaitos	Ataskaitos sukuriamos, naudojantis tokiomis paprastomis priemonėmis, kaip Microsoft Excel skaičiuoklė
Atliekant analizę vienas iš pagrindinių analizuojamų objektų yra studentų atsakytos anketos.	Anketų rezultatuose naudojama tik klausimų, kurie vertinami penkiabalėje sistemoje, duomenys, skaičiuojant kiekvieno iš klausimų vidurkį[9][10]

Sukūrus sprendimą bus galimybė sužinoti pasirinktų studentų savybių reikšmės ateityje. Todėl bus galima imtis priemonių gerinti situaciją anksčiau nei pasibaigus semestru. Analizės rezultatus bus galima peržiūrėti, pasinaudojus ASP.NET internetinių puslapių pagalba. Šis sprendimas gerai papildys anketas, kurios dabar yra naudojamos analizuojant studijų kokybę.

Iš likusių padalinių svarbiausias padalinys yra katedra, kadangi pagal Vidinės studijų kokybės užtikrinimo sistemos (VSKUS) aprašą[1] ji: vertina studijų kokybę pagal įdarbinimo įstaigų pastabas, ruošia mokomąją literatūrą bei diegia pažangias mokymo formas ir metodus. Šias veiklas galima parodyti panaudojimo atvejų modeliu, kuris matomas 1.8 paveiksle.



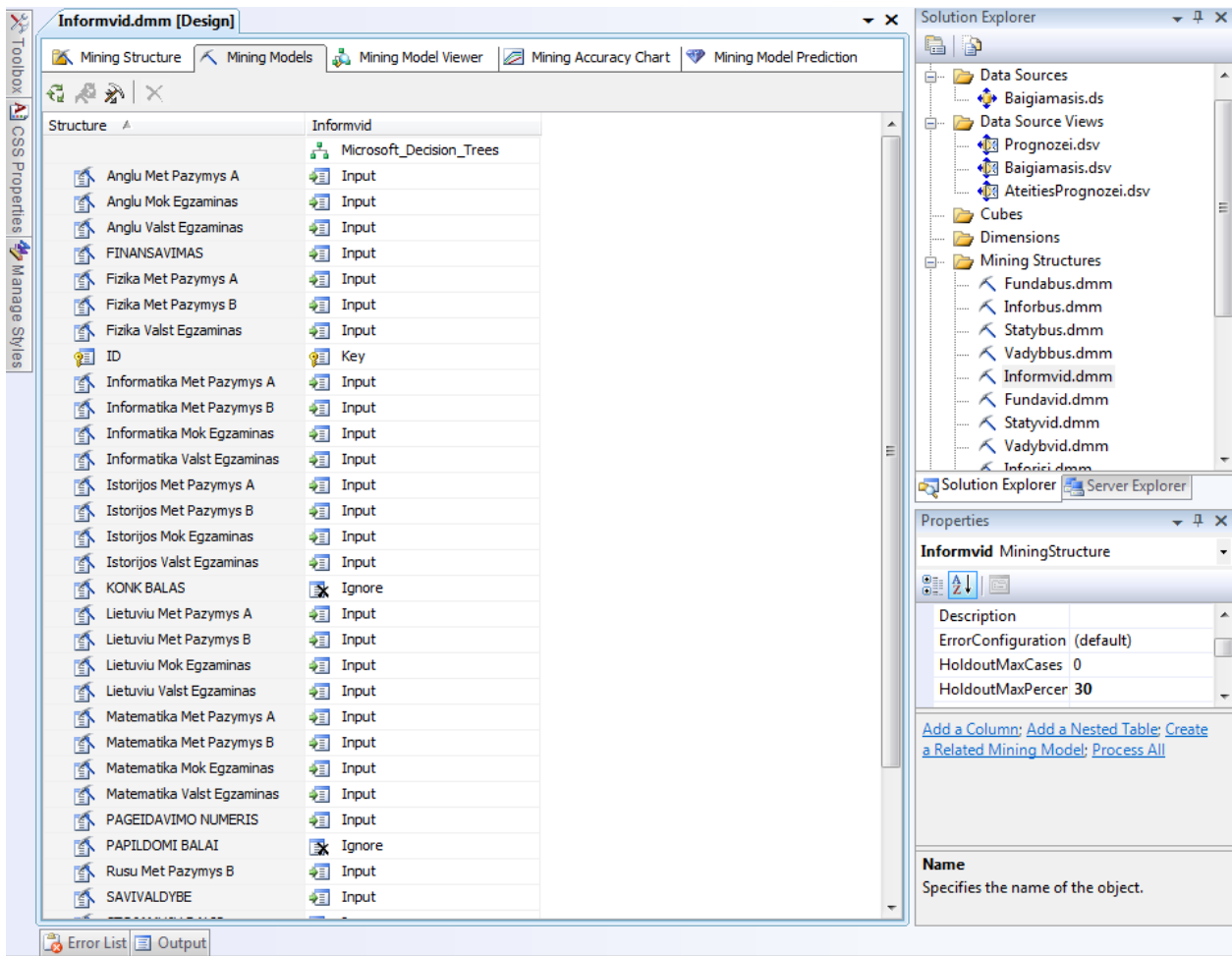
1.8 pav. Katedros studijų kokybės gerinimo panaudojimo atvejai

## 1.7. Esamų sprendimų analizė

### 1.7.1. Šiuo metu atlikti darbai, naudojant duomenų gavybą

Iki atliekant šį tyrimą universitete jau buvo mėginimų tirti turimą akademinę informaciją. Buvo naudojami naujai priimamų į universitetą abiturientų duomenys ir pagal žinomus ankstesnių metų duomenis spėjama, ar abiturientas iškris, ir kokį vidurkį jis turės. Darbų pradžioje buvo kuriama tik su Microsoft SQL Server Business Intelligence Development Studio paketu, kuriame esančiomis priemonėmis sukuriami duomenų gavybos modeliai, ir vėliau peržiūrimi rezultatai. Šis sprendimas buvo nepatogus, kadangi vartotojui reikėjo naudotis Microsoft Visual Studio programine įranga, skirta kurti ir plėtoti verslo analitikos sprendimus, tarp kurių yra ir duomenų gavyba, bei kurti kitus sprendimus. Eiliniam vartotojui buvo sudaromos galimybės sugadinti sprendimą arba veiksmus atlikti neteisingai. Situacija, kai vartotojas gali pasiekti modelio atributus, kas kelia pavojų sprendimo veikimui, matoma 1.9 paveiksle.





1.9 pav. Paprastam vartotojui matoma duomenų gavybos modelio struktūra

Šiame sprendime buvo galimybė vaizdžiomis priemonėmis peržiūrėti pasirinktų spėjamų atributų rezultatus, matyti nuo kokių kitų atributų jie priklauso. Taip pat buvo galima peržiūrėti dėsningumus duomenyse, juos sugrupavus grupėmis. Buvo galima surasti taisykles, kurios galioja tarp skirtingų duomenų stulpelių. Duomenų gavybos modeliai ir struktūros buvo kuriami vedlių pagalba, naudojant vieną iš algoritmų.

Vėliau buvo naudojamas kitas sprendimas - apjungti jau minėtą galimybę su PHP programavimo kalba. Kadangi PHP kalba nėra pati tinkamiausia naudoti SQL Server duomenų bazių valdymo sistemą, kilo įvairiausių nesuderinamumo problemų, tokių, kaip atributų vardų ilgio apribojimai lentelėse, lietuviškų simbolių neteisingas interpretavimas. Taip pat buvo neįmanoma išnaudoti visų duomenų bazių valdymo sistemos ir verslo analitikos galimybių. Tokių, kaip pavyzdžiui, naudojantis sukurta sąsaja, atnaujinti informaciją apie studentus. PHP kalba sukurtoje vartotojo sąsajoje buvo galima įkelti abiturientų duomenis bei atlikti vieną iš kelių analizių. Kadangi rezultatai vaizduojami lentelėmis, nebuvo galimybės vaizdžiai peržiūrėti rezultatus. Norint juos peržiūrėti, reikėjo papildomai naudoti Business Intelligence Development Studio programą, kas nebuvo patogiu. Pačiame verslo analitikos projekte buvo realizuoti abiturientų duomenų

sugrupavimas pagal panašias savybes, taisyklių suradimas tarp abiturientų duomenų stulpelių, bandyta spėti, kiek abiturientų studijuos po kelių metų.

Užsienio universitetuose, pasinaudojant duomenų gavyba, yra atlikta daugybė tyrimų, kuriais sprendžiami kitokio tipo uždaviniai. Niujorko valstijos Bufalo miesto universitete buvo atliekami tyrimai, kurių dėka testuose apie energiją prognozuojamas studentų žinių lygis, įvertinama ar jie sugebės atlikti tam tikras užduotis[11]. Panaudoti realūs kelių metų skirtingų studijų duomenys. Pasinaudota sprendimo medžių ir linijinės tiesinės funkcijos metodu. Šių tyrimų dėka norėta sužinoti, kokia tikimybė, kad studentas sugebės vykdyti užduotis, suskirstytas į keletą temų, tokių kaip veiklos, energijos išsaugojimo ir panašiai. Šios temos dar suskirstytos pagal tai, ar aptariami dalykai iš kasdieninės veiklos ar ne, ir galiausiai pagal studentų išsilavinimą. Taip skirstant gautas sprendimo medis arba išskirstymo elementai, naudojami linijinėje funkcijoje.

Japonijos Meiji universitete buvo atliekamas tyrimas, kuriuo bandyta nustatyti elektros perdavimo sistemos apkrovimo prognozę trumpajam laikotarpiui[12]. Buvo pasinaudojama duomenų gavyba, kaip priemone tai atlikti. Sudarytas regresijos medis, kuris yra labai panašus į paprastą sprendimo medį, bei neuroninis tinklas. Gauta, kad apkrova priklauso nuo savaitės dienos, maksimalaus elektros suvartojimo dieną bei temperatūros lauke. Sprendimų medis sudarytas, pasinaudojant šiais kriterijais.

Indijos mokslininkai duomenų gavybą pritaikė tirti finansiniams duomenims[13]. Buvo pasinaudota grupavimo algoritmu, duomenis suskirstant į smulkias grupes. Nustatyta, kad jis tinka, norint priimti įvairius finansinius sprendimus. Vienas iš sprendimo uždavinių yra atlikti rizikos faktorių analizę, pasinaudojant duomenų gavyba bei statistiniais modeliais. Gavyba labai padeda, kai reikia turėti tam tikras žinias apie galimą tam tikro finansinio kriterijaus būklę ateityje. Finansiniams duomenims tirti taip pat naudojamas laiko serijų algoritmas.

Nors paminėti atlikti tyrimai pritaikyti kitokiose dalykinėse srityse, juos galima pritaikyti ir norint pagerinti studijų kokybę. Algoritmai gali būti panaudoti spręsti panašaus pobūdžio uždavinius, tokius kaip studento mokymosi rezultatų semestro gale radimas, studentų apklausų tyrimas, dėsningumų, kurie yra sunkiai žmogaus pastebimi, suradimas.

### **1.7.2. Šiuo metu siūlomi duomenų gavybos sprendimai**

Atliekant analizes Kauno technologijos universitete, buvo vykdomi duomenų gavybos tyrimai, naudojant Microsoft SQL Server duomenų bazių valdymo sistemą, kuri išleista 2008 metais. Ši duomenų bazių valdymo sistema universitete yra naudojama mokymo tikslais neviename modulyje. Tačiau be Microsoft kompanijos produkto duomenų gavybą galima atlikti ir kitų kompanijų gaminiais. Pagal Microsoft pateiktą duomenų bazių sistemų palyginimą[14] nuspręsta, kad didžiausi SQL Server konkurentai yra Oracle ir IBM produktai. Oracle produktas yra pavadintas Data Miner. Pati Oracle duomenų bazių valdymo sistema, kurios dalis yra šis sprendimas, yra

naudojama akademinėje informacinėje sistemoje, kuria naudojami universiteto darbuotojai ir studentai, tačiau duomenų gavybos įrankis joje nenaudojamas. Tuo tarpu IBM duomenų bazių valdymo sistema, jos duomenų gavybos įrankis iš viso nėra panaudoti. Toliau 1.2 lentelėje yra pateikiami šių trijų įmonių įrankių, skirtų atlikti duomenų gavybą, palyginimai.

**1.2 lentelė.** Duomenų gavybos sprendimų palyginimas

Savybė	Microsoft Business Intelligence studio [15]	Oracle Data Miner [16]	IBM DB2 Warehouse [17]
Kokiais būdais galima atlikti duomenų gavybą	Naudojant DMX užklausų kalbą arba naudojantis Business Intelligence Studio grafinėmis kūrimo priemonėmis.	Naudojantis Data Miner grafine aplinka, kurioje kuriamos duomenų srautų sekos kurti duomenų gavybos procesą. Galima naudoti PL/SQL paketus ir standartinio SQL funkcijas.	Naudojant Design Studio grafinę aplinką, kurioje sukuriama duomenų saugyklos projektas. Tai atlikus pasinaudojama duomenų srautais vykdyti duomenų transformacijas, ir galiausiai sukuriama duomenų gavybos srautas. Taip pat galima naudoti SQL skriptus.
Galimų algoritmų skaičius	9	9	10
Ar buvo atliekami studentų tyrimai universitete	Buvo	Nebuvo	Nebuvo
Ar yra galimybė gauti nemokamai	Kauno technologijos universiteto studentams yra galimybė gauti ją mokymosi tikslams	Nėra	Nėra

Pagal lentelėje matomą informaciją galima pastebėti, kad daugiausia įtakos tam, kad yra pasirinktas Microsoft produktas, turi jo pasiekiamumas studentams ir tai, kad su juo, skirtingai nei su kitais produktais, jau buvo bandymai atlikti tyrimus.

### 1.7.3. Sprendimai, naudojantys lietuvių kalbos atpažinimą

Šiame skyrelyje bus trumpai paminėti keli sprendimai, naudojantys lietuvių kalbos atpažinimą. Palyginus su anglų kalbos atpažinimu, tokių sprendimų yra labai mažai. Viena didžiausių tuo užsiimančių organizacijų yra „Tildė IT“ grupė. Pagal informaciją pateiktą jos internetiniame puslapyje[18] jie užsiima:

- Programinės įrangos lokalizavimu.
- Dokumentacijos vertimu.
- Lietuvių kalbos palaikymo programinės įrangos kūrimu.

Ši organizacija yra sukūrusi vieną iš geriausiai žinomų programinės įrangos paketų Tildės Biuras, kurio dėka užtikrinamas lietuvių kalbos veikimas asmeniniuose kompiuteriuose. Šios programinės įrangos naujausioje versijoje Tildės Biuras 2012 yra sekančios galimybės[19]:

- Lietuvių kalbos rašybos ir gramatikos tikrintuvas. Sinonimų žodynas, automatinis žodžių kėlimas į kitą eilutę. Jis veikia su populiariausiais ofiso paketais.

- „Tildės biuro žodynas“. Žodynai verčiantys vieną kalbą į kitą. Taip pat yra terminų žodynai. Galimybė išgirsti žodžių tarimą, pasižiūrėti gramatinę analizę bei angliškų žodžių transkripciją, Visi žodynai integruojami į MS Word.
- Klaviatūros įrankiai ir išdėstymai. Naujas klaviatūros išsidėstymas Russian translit kuomet užrašius lotyniškai rusišką raidę taip kaip ji tariama gaunama raidė parašyta kirilica. Taip pat yra dokumentų šablonai ir paveikslėlių galerija.
- Šriftų rinkiniai Europos kalboms bei kirilicai.

Bene vienintelį sukurta produktą, skirtą lietuvių kalbos morfologiniam nagrinėjimui, galime paminėti Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centro sukurta morfologinį anotatorių[20]. Jo dėka galima gauti morfologiškai anotuotą ir vienareikšmintą tekstą, kai parenkamas labiausiai tikėtinas atvejis. Tuomet gaunami veiksmažodžiai bendratyje, daiktavardžiai vardininke ir panašiai. Tokia gauta forma vadinama lema. Be jos dar gaunama informacija, kurioje matoma žodžių linksniai, giminės, nuosakos, kalbos dalys ir panaši morfologinė informacija.

Šio sprendimo dėka bandyta išspręsti problemą su lietuvių kalbos morfologiniu daugiareikšmiškumu, kai vienas žodis turi kelias lemas arba gali turėti kelias skirtingas gramatines atžymas. Po to, kai buvo pritaikyti statistiniai modeliai ir pusiau automatinis tekstynas, pasiektas įrankio tikslumas yra apie 94 procentai.

Norint pasinaudoti šiuo įrankiu, reikia internetiniame puslapyje į pateiktą sąsają įrašyti sakinius bei, paspaudus mygtuką „Rezultatas puslapyje“, gauti anotuotą informaciją. Rezultatą galima gauti ekrane arba, paspaudus kitą mygtuką „Rezultatas faile“, gauti tekstinę bylą, kurią galima išsisaugoti asmeniniame kompiuteryje.

## **1.8. Duomenys, pagal kuriuos reikia kurti sprendimą**

Visų duomenų gavybos algoritmų apmokymui bei duomenų gavybos analizei atlikti naudojami realūs Kauno technologijos universiteto duomenys. Visus šiuos duomenis galima suskirstyti į dvi dalis:

- Duomenys apie studentų mokymąsi universitete bei įvertinimus stojimo metu.
- Studentų užpildytų studijų modulių kokybės įvertinimo anketų duomenys.

Pirmosios kategorijos duomenys - tai 2008 metais mokęsi Informatikos fakultete studentai, apie kuriuos informacija patalpinta Lietuvos aukštųjų mokyklų asociacijos bendrajam priėmimui organizuoti priėmimo duomenyse. Šie duomenys yra paimti iš universiteto akademinės informacinės sistemos. Naudojami duomenys apie:

- Studento stojimo įvertinimus.
- Studento asmeninė informacija.
- Studento individualių planų informacija.
- Moduliai bei dėstytojai, kurie dėstė nagrinėjamiems studentams.

- Taip pat naudojami papildomi duomenys, tokie kaip padalinių, pareigybių sąrašai, kuriuos galima laikyti klasifikatoriais.

Antrosios kategorijos duomenys - tai anketų informacija. Anketas studentai pildė 2011 metų gruodžio pradžios - 2012 metų sausio pabaigos laikotarpiu. Visi šie duomenys yra nuasmeninti - užtikrinta, kad nebūtų galimybės sužinoti, kuris studentas, kokius atsakymus pateikė į anketos klausimus. Visa informacija buvo pateikta vienoje lentelėje, kurioje buvo atsakymai į klausimus bei papildoma informacija, kuri nurodo respondento charakteristikas: kiek jam metų, kokia jo lytis ir panašiai.

Klausimai anketoje yra dviejų tipų. Pirmojo tipo klausimuose reikia pažymėti vieną iš kelių pateiktų atsakymo variantų. Pateikiamas teiginys ir reikia nurodyti, kaip stipriai patvirtiname ar paneigiame šį teiginį. Antrojo tipo klausimai yra atviro tipo. Juose studentas laisva forma parašo atsakymą. Šie klausimai yra svarbiausi anketų analizei, kadangi jos naudojamos atlikti teksto gavybą. Iš viso yra trys klausimai: nurodykite dalyko teigiamus aspektus, nurodykite neigiamus aspektus bei kaip siūlytumėte patobulinti modulį.

Kadangi anketą pildyti buvo neprivaloma, nemaža dalis studentų jos visai nepildė. Kiti studentai atsakinėjo tik į tuos klausimus, kuriuose reikia pasirinkti teisingą atsakymą. Gautuose duomenyse yra visų fakultetų studentų atsakymai, tačiau, kadangi pirmosios kategorijos naudojami duomenys yra iš Informatikos fakulteto, naudinga ir čia naudoti to paties padalinio studentų atsakymus. Toliau yra pateikta, kiek atsakymų buvo pateikta iš viso bei kiek pateikta atsakymų į atviro tipo klausimus visuose fakultetuose ir konkrečiai Informatikos fakultete. Ši informacija parodyta 1.3 lentelėje.

**1.3 lentelė.** Atsakymų į anketos klausimus skaičiai 2011 m. gruodžio mėn – 2012 m. sausio mėn laikotarpiu

	<b>Visi fakultetai</b>	<b>Informatikos fakultetas</b>
<b>Iš viso atsakyta</b>	8574	729
<b>1 klausimas atsakytas</b>	2352	198
<b>2 klausimas atsakytas</b>	2031	215
<b>3 klausimas atsakytas</b>	1737	185

## **1.9. Siekiamas sprendimas, gerinantis studijų kokybę**

Pagrindinis siekiamas tikslas - sukurti duomenų gavybos algoritmus naudojančią sistemą, kuri pagerins studijų kokybės ir stebėsenos procesą, naudojant Microsoft SQL Server duomenų bazių valdymo sistemą su visomis paslaugomis bei aplinką, pasiekiamą interneto naršyklės pagalba, taip, kad šie abu elementai turėtų geresnį suderinamumą, nei PHP kalbos atveju. Aplinka bus sukurta, naudojant ASP.NET karkasą, kuris taip pat sukurtas Microsoft kompanijos, todėl suderinamumas turėtų būti geras.

Sukūrus sistemą ji bus įtraukta į studijų kokybės stebėsenos procesą, tačiau visiškai nepakeis metodų, kurie naudojami dabar. Tai bus priemonė gauti skaičius ateityje vykstančių įvykių ar

pamatyti dėsniumus, kurie paprastai nėra pastebimi. Pagal šiuo metu pradėtą vykdyti universiteto mokslo ir studijų atsinaujinimo programą, kuriamas sprendimas galėtų padėti valdyti studijas – tobulinti studijų programas, taip pat prisidėtų prie akademinės informacinės sistemos modernizavimo.

### **1.10. Rizika, kuri kyla, kuriant duomenų gavybos sprendimus**

Didžiausia problema, kuri kyla kuriant sprendimą – problemos su skirtingų SQL Serverio komponentų integravimu. Didelė rizika kyla dėl to, kad nėra patirties, kaip teisingai integruoti duomenų bazių valdymo sistemą su jos ataskaitų kūrimo paslauga. Ataskaitų paslaugoms reikalingas tinkamai sukonfigūruotas šių paslaugų serveris. Jį parengti darbui, taip pat sukurti ataskaitas taip, kad teisingai veiktų su .NET karkasu yra sudėtinga, todėl dažnai apsimoka ataskaitas kurti, naudojant tik .NET priemones. Pasinaudojant įvairiais ASP komponentais bei stiliumis bylomis, grafinę išvaizdą galima realizuoti beveik tokią pačią, kaip ir kuriant ataskaitų serverio pagalba. Tačiau šiuo atveju nereikalingas joks papildomas serveris.

Taip pat dažniausiai nebūna iki galo aišku, koku būdu atlikti duomenų paruošimą duomenų gavybai. Ar naudoti vaizdinius, ar kurti daugiadimensinius kubus. Antruoju variantu yra patogiau sudarinėti naujas analizes, tačiau yra rizika, kad kils problemų, panaudojant MDX kalbą, kuri reikalinga gauti duomenis iš kubų. Taip kubams reikia pasirinkti tinkamas dimensijas bei tinkamus duomenų peržiūros pjūvius.

Atliekant analizes paprastai reikia turėti duomenis, kurie bus tiriami. Čia taip pat yra įvairiausių galimybių. Galima panaudoti prisijungimą prie kitos išorinės sistemos, tačiau yra didelė rizika, kad tos informacinės sistemos savininkai nenorės tokią prieigą suteikti. Kitas kelias yra pasinaudoti bylomis, kurios įkeliamos į sistemą. Šis būdas taip pat turi trūkumų: naudojant Excel ar Access bylas, kyla problemų su duomenų formatais. Dėl šios priežasties susigadina informacija, ją perkėlus į duomenų bazę.

Kadangi lietuviškam tekstui tirti naudojama programinė įranga sukurta kitose valstybėse yra didelė rizika, kad nebus galima teisingai atlikti tyrimus. Galima situacija, kai neteisingai arba visai nesurandami žodžių linksniai, asmenys, laikai, gali kilti problemų dėl to, kad produktai nesugebės surasti žodžių šaknų. Tokiu atveju gali tekti naudoti alternatyvią metodiką, norint atlikti modulių įvertinimą pagal anketos atsakymo tekstą.

### **1.11. Rezultato kokybės kriterijai**

Pagrindinis tyrimo rezultatas - sukonstruotas sprendimas, pagerinantis universiteto studijų kokybės ir stebėsenos procesą. Sprendimas yra informacinė sistema, kurios kokybę galima analizuoti charakteristikomis pagal ISO 9126 standartą[21]. Produkto kokybę gali būti tikrinama patikimumo, naudojamumo, palaikomumo ir funkcionalumo aspektais.

Įvykus klaidai produktas turi nepatekti į būseną, kai vartotojas nebegali nieko atlikti dėl to, kad sistema pateko į „mirties tašką“. Įvykus klaidai turi būti galimybė tęsti darbą toliau. Visas surastas klaidas yra galimybė ištaisyti.

Naudojimasis produktu turi būti patogus, suprantamas kiekvienam vartotojui. Vos po pusės valandos mokymosi, vartotojas turi sugebėti atlikti sistemos funkcijas. Todėl vartotojo sąsaja turi būti patogi ir lengvai perprantama.

Sukurtas sprendimas turi būti nesudėtingai keičiamas, pridėdamos naujos dalys. Kadangi naudojamas .NET karkasas, tai atlikti nebus labai sudėtinga, tereikės tik sukurti naują meniu pasirinkimą į naujai sukurtą formą sistemoje. Pakeitus tam tikras dalis sistemoje, jos turi nepakenkti kitoms dalims. Šio kriterijaus bus laikomasi, naudojant duomenų bazės valdymo sistemos serverio procedūras.

Sistemos atliekamos funkcijos turi atitikti funkcinius ir nefunkcinius reikalavimus, kurie yra išskelti sprendimui. Gaunamas naudojimosi produktu rezultatas turi būti nuspėjamas ir neprieštarauti studijų dalykinės srities faktams, kurie yra nepaneigiami. Funkcionalumas taip pat turi neleisti pasiekti sistemoje informacijos asmenims, kurie neturi teisių ją matyti.

Taip pat sprendimo veikimą galima įvertinti, atsižvelgiant į duomenų gavybos algoritmų tikslumą. Ar gauti rezultatai, kurie gaunami naudojant duomenis, ir kurių dėsningumai jau yra žinomi neprieštarauja realiai situacijai.

## **1.12. Analizės išvados**

1. Surasta, kad Kauno technologijos universitetas studijų kokybę ir stebėseną vykdo pagal vidinę studijų kokybės užtikrinimo sistemą. Pagal šios sistemos aprašą jos valdymas įtakoja studijų valdymą, tuo tarpu studijų valdymo rezultatai ir vykstanti veikla keičia šios studijų kokybės sistemos valdymą.
2. Šiuo metu atliekama universiteto reorganizacija pagal mokslo ir studijų atsinaujinimo programą. Pagal šią programą turi pasikeisti universiteto organizacinė struktūra, tapti efektyviau išnaudojama infrastruktūra ir studijų programos atrodyti patrauklesnės visuomenei. Kuriamas sprendimas galėtų būti suintegruotas su universiteto dabar naudojama akademinė sistema.
3. Pastebėta, kad svarbiausias universiteto padalinys, susijęs su studijų kokybe, yra Strateginio planavimo ir kokybės departamento Kokybės skyrius. Todėl šio padalinio darbuotojai yra pagrindiniai vartotojai, susiduriantys su tiriamą problema.
4. Universiteto veiklą įtakoja ir kai kurios išorinės organizacijos. Jas galima skirstyti į valstybines ir nevalstybines. Pastarųjų vadovai yra potencialūs studentų darbdaviai, tad jų nuomonė yra svarbi studijų kokybei, kuri gali būti matuojama pagal tai, kaip universiteto absolventai sugeba dirbti darbo vietose.

5. Kadangi dabar atliekama analizė su duomenimis, kurie yra jau praėjusių įvykių informacija, gaunami rezultatai nerodo, kas atsitiks ateityje. Ji apibendrina tai, kas jau įvyko, todėl negalima pamatyti galimos studentų situacijos ateityje.
6. Iki dabar atlikti tyrimai neturėjo jokios gerai suderintos vartotojo sąsajos su duomenų gavybos modeliais. Daugelį veiksmų reikėdavo daryti, naudojant Business Intelligence Development Studio paketo aplinką, dėl ko buvo galimybė sugadinti duomenų gavybos modelius.
7. Sukurtas sprendimas su PHP programavimo kalba prastai suderinamas su SQL Server duomenų bazių valdymo sistema, todėl nebuvo galimybės sistemą tobulinti dėl kylančių suderinamumo problemų.

## **2. KURIAMOS SISTEMOS, GERINANČIOS KOKYBĖS IR STEBĖSENOS PROCESUS, REIKALAVIMAI**

### **2.1. Kuriamo sprendimo reikalavimų specifikacija**

#### **2.1.1. Kuriamo duomenų gavybos sprendimo funkcionalumo aprašymas**

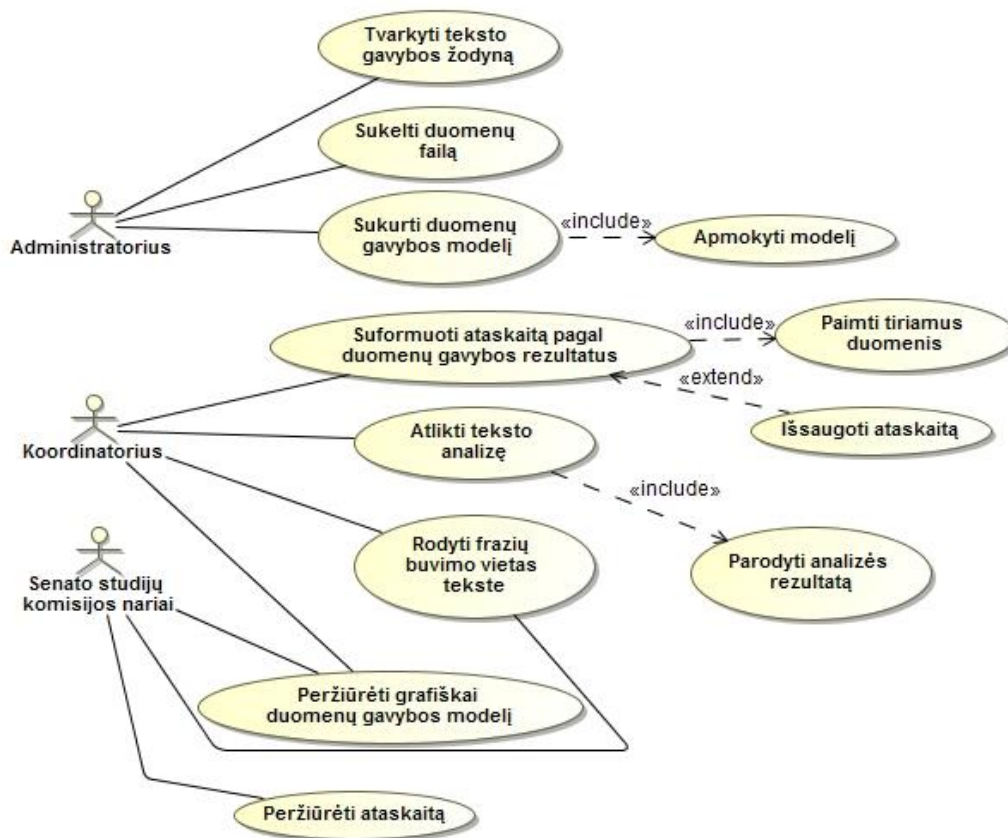
Pagal atliktą studijų kokybės ir stebėsenos proceso analizę nuspręsta, kad sistemoje turi būti du vartotojų tipai. Vienas skirtas studijų kokybės ir stebėsenos skyriaus koordinatoriui, kitas - universiteto Senato studijų komisijos nariams. Taip pat reikalingas ir papildomas vartotojas administratorius, kuris turi įtraukti naujus modelius į sistemą.

Pastarasis ne tik sukuria duomenų gavybos modelius ir struktūras, bet ir gali sukelti duomenis, kurie turi būti tiriami, pasinaudojant duomenų gavyba. Nuspręsta, kad tam labiausiai tinka XML bylos, kurios geriausiai perduoda duomenis tarp skirtingų sistemų. Kuriant anksčiau paminėtas struktūras ir modelius, pastarieji privalo būti apmokomi. Šį procesą sistema atlieka automatiškai. Teksto gavyboje yra naudojami žodynai, kuriuos taip pat administruoja administratorius.

Koordinatorius turi turėti galimybę atlikti visas analizes, kurios yra sistemoje. Analizių metu yra sukuriama ataskaita, kuri vėliau parodoma senato studijų komisijos nariams. Atliekant teksto gavybą, turi būti galimybė matyti, kur yra kiekviena iš konkrečių frazių, lemianti studijų modulio įvertinimą. Po analizės visada parodomas jos rezultatas. Kuo įvertinimas didesnis, tuo modulis yra geresnės kokybės. Kai įmanoma yra galimybė peržiūrėti duomenų gavybos modelio grafinį vaizdą.

Senato studijų komisijos nariai, kaip ir koordinatorius, turi galimybę peržiūrėti duomenų gavybos grafinį vaizdą ir matyti studijų dalyko įvertinimą bei kaip tas įvertinimas yra sudarytas, pasinaudojant teksto gavyba. Tik šis vartotojo tipas gali pamatyti Koordinatoriaus suformuotą ataskaitą. Visos aprašytos funkcijos pavaizduotos panaudojimo atvejų diagramoje, kuri matoma 2.1 paveiksle.





**2.1 pav.** Funkcijos, kurias turi turėti kuriamas sprendimas

Toliau visi panaudojimo atvejai yra pateikti lentelėse, kuriose matoma detalesnė informacija apie juos. Pateikiama funkcijos tikslas, kas ją vykdo, kada ji vykdoma bei kaip vykdoma. Pirmasis panaudojimo atvejis „Tvarkyti teksto gavybos žodyną“ detaliau pateiktas 2.1 lentelėje.

2.1 lentelė. Panaudojimo atvejo „Tvarkyti teksto gavybos žodyną“ aprašymas

<b>PA</b> „Tvarkyti teksto gavybos žodyną“	
<b>Tikslas.</b> Tvarkyti sąrašą frazių, kurios naudojamos teksto gavybai	
<b>Aprašymas.</b> Administratorius gali tvarkyti frazių sąrašus. Šie sąrašai yra skirti kiekvienam iš studijų modulių įvertinimo anketos klausimų. Galima pridėti, redaguoti ir naikinti frazę.	
<b>Prieš sąlyga</b>	Administratorius prisijungęs savo paskyra
<b>Aktorius</b>	Administratorius
<b>Sužadinimo sąlyga</b>	Administratorius nori atlikti pakeitimus frazių žodyne, kuris naudojamas atliekant teksto gavybą.
<b>Veiklos taisyklės</b>	
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>
	<b>Apima PA</b>
	<b>Specializuoja PA</b>
<b>Pagrindinis įvykių srautas</b>	
1. Meniu pagalba vartotojas pasirenka, kurio klausimo žodyną nori tvarkyti.	Parodomas pasirinkto klausimo frazių sąrašas
2. Vartotojas pažymi frazę, kurios pakeitimą nori atlikti	Parodoma pasirinktos frazės informacija
3.a) 1. Norint pridėti naują frazę, paspaudžiamas pasirinkimas „Naujas“ arba norint redaguoti paspaudžiama „Redaguoti“	Sistema parodo atitinkamą formą, kurioje reikia nurodyti prašomą informaciją
3.a) 2. Vartotojas nurodo frazę, jos įvertinimą ir paspaudžia pasirinkimą „Įterpti“, kai sukuriama nauja, arba pasirinkimą „Atnaujinti“, kai redaguojama frazė	Sistema įrašo pakeitimus į duomenų bazę
3.b) 1. Norint panaikinti frazę, paspaudžiamas pasirinkimas „Trinti“	Sistema ištrina pasirinktą frazę
4. Sistema baigia PA	
<b>Po sąlyga</b>	Atliktas vienas iš frazių administravimo veiksmų
<b>Pastabos</b>	

Panaudojimo atvejais „Sukelti duomenų failą“, kurio dėka XML byla įkeliamą į sistemą, pateiktas 2.2 lentelėje.

2.2 lentelė. Panaudojimo atvejo „Sukelti duomenų failą“ aprašymas

<b>PA</b> „Sukelti duomenų failą“	
<b>Tikslas.</b> Į sistemą įkelti XML failą iš kurio vėliau bus nuskaitomi duomenys.	
<b>Aprašymas.</b> Administratorius gali į sistemą įkelti XML bylą iš savo asmeninio kompiuterio	
<b>Prieš sąlyga</b>	Administratorius prisijungęs savo paskyra
<b>Aktorius</b>	Administratorius
<b>Sužadinimo sąlyga</b>	Administratorius nori į sistemą įkelti XML bylą
<b>Veiklos taisyklės</b>	
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>
	<b>Apima PA</b>
	<b>Specializuoja PA</b>
<b>Pagrindinis įvykių srautas</b>	
1. Vartotojas pasirenka meniu pasirinkimą „Įkelti XML bylą“	Sistema parodo formą, kurioje galima pasirinkti failą asmeniniame kompiuteryje
2. Vartotojas pasirenka bylą ir paspaudžia mygtuką „Vykdyti“	Sistema patikrina failo tipą ir įkelia jį į sistemą

3. Sistema baigia PA	
<b>Po sąlyga</b>	Įkelta XML byla į sistemą
<b>Alternatyvūs scenarijai</b>	
1. Failo dydis ir tipas netenkina reikalavimo	Parodomas klaidos pranešimas
<b>Pastabos</b>	

Duomenų gavybos struktūra ir modelis sukuriama pasinaudojant funkcija, kuri aprašyta panaudojimo atveju „Sukurti duomenų gavybos modelį“. Šis atvejis aprašytas 2.3 lentelėje.

2.3 lentelė. Panaudojimo atvejo „Sukurti duomenų gavybos modelį“ aprašymas

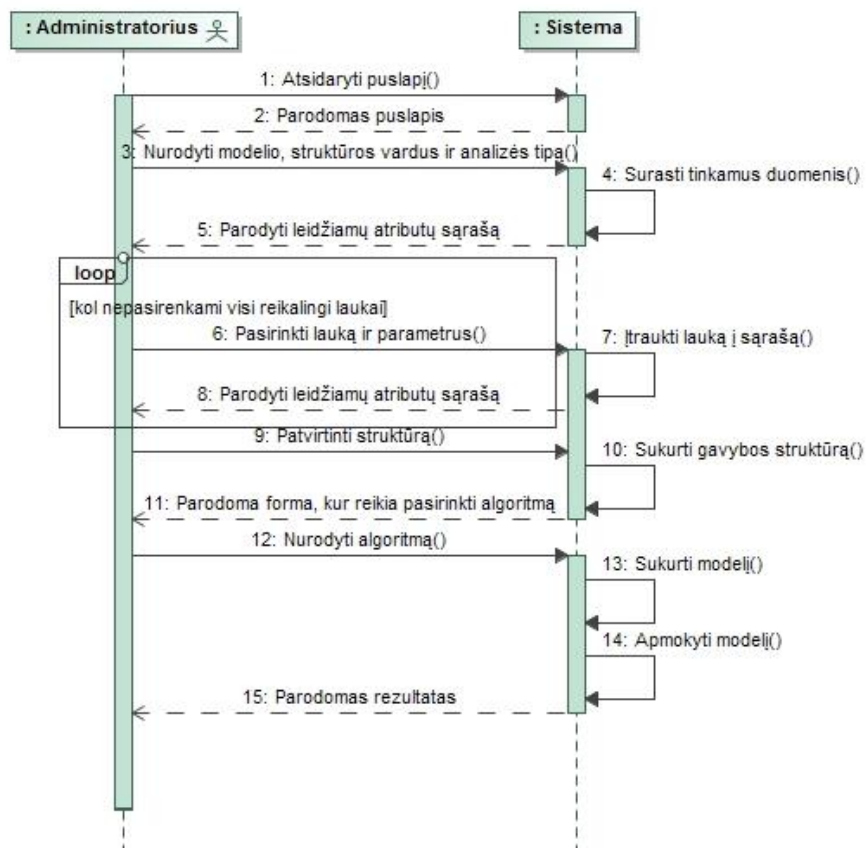
<b>PA „Sukurti duomenų gavybos modelį“</b>		
<b>Tikslas.</b> Sukurti duomenų gavybos struktūrą ir modelį pagal nurodytus kriterijus.		
<b>Aprašymas.</b> Administratorius gali sukurti duomenų gavybos struktūrą ir modelį, nuroydamas parametrus tokius kaip, iš kokių laukų jie turi susidėti.		
<b>Prieš sąlyga</b>	Administratorius prisijungęs savo paskyra	
<b>Aktorius</b>	Administratorius	
<b>Sužadinimo sąlyga</b>	Administratorius nori sukurti duomenų gavybos struktūrą ir modelį.	
<b>Veiklos taisyklės</b>		
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>	
	<b>Apima PA</b>	Apmokyti algoritimą
	<b>Specializuoja PA</b>	
<b>Pagrindinis įvykių srautas</b>		
1. Vartotojas pasirenka meniu pasirinkimą „Sukurti modelį“	Sistema parodo formą, kurioje reikia nurodyti struktūros ir modelio pavadinimus bei kokio tipo analizę reikės vykdyti	
2. Vartotojas nurodo struktūros ir modelio pavadinimus bei analizės tipą, kuri bus vykdoma	Sistema parodo formą, kurioje reikia nurodyti, iš ko susidės duomenų gavybos struktūra	
3.a) Vartotojas nurodo, kokį lauką iš duomenų bazės naudoti struktūroje, koks jo bus tipas. Šis žingsnis kartojamas tiek kartų, kiek reikės laukų struktūroje		
3.b) Vartotojas patvirtina struktūrą		
3.c) Vartotojas paspaudžia mygtuką „Sukurti struktūrą“	Sistema sukuria duomenų gavybos struktūrą bei parodo formą, kurioje reikia nurodyti, kokį algoritimą taikysime duomenų gavybai	
4. Vartotojas nurodo duomenų gavybos algoritimą ir paspaudžia mygtuką „Patvirtinti“	Sukuriamas duomenų gavybos modelis bei vykdoma „Apmokyti modelį“ PA	
5. Sistema baigia PA		
<b>Po sąlyga</b>	Sukurtas apmokytas duomenų gavybos modelis	
<b>Alternatyvūs scenarijai</b>		
1. Nenurodyta viename iš laukų informaciją	Parodomas klaidos pranešimas, kurio pagalba vartotojas gali pataisyti duomenis formoje	
<b>Pastabos</b>		

Vykdam prieš tai aprašytą panaudojimo atvejį, reikia apmokyti duomenų gavybos modelį. Tam yra naudojama funkcija „Apmokyti modelį“ sistemoje. Šios funkcijos panaudojimo atvejis parodytas 2.4 lentelėje

2.4 lentelė. Panaudojimo atvejo „Apmokyti modelį“ aprašymas

<b>PA</b> „Apmokyti modelį“	
<b>Tikslas.</b> Apmoko duomenų gavybos modelį.	
<b>Aprašymas.</b> Apmoko duomenų gavybos modelį, po to, kai jis yra sukuriamas „Sukurti duomenų gavybos modelį“ panaudojimo atvejo pagalba	
<b>Prieš sąlyga</b>	Sukurta duomenų gavybos struktūra ir modelis
<b>Aktorius</b>	Sistema
<b>Sužadinimo sąlyga</b>	Sukurta duomenų gavybos modelis, kuris privalo būti apmokytas, norint jį toliau naudoti.
<b>Veiklos taisyklės</b>	
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>
	<b>Apima PA</b>
	<b>Specializuoja PA</b>
<b>Pagrindinis įvykių srautas</b>	
1. Modelis, kuris sukuriamas yra apmokomas	Sistema apmoko duomenų gavybos modelį
2. Sistema baigia PA	
<b>Po sąlyga</b>	Apmokytas duomenų gavybos modelis
<b>Pastabos</b>	

Visi veiksmai atliekami modelio sukūrimo metu pavaizduoti sekų diagramos pagalba. Joje matoma tai, kas aprašyta panaudojimo atvejų „Sukurti duomenų gavybos modelį“ ir „Apmokyti modelį“ lentelėse. Ši diagrama matoma 2.2 paveiksle.



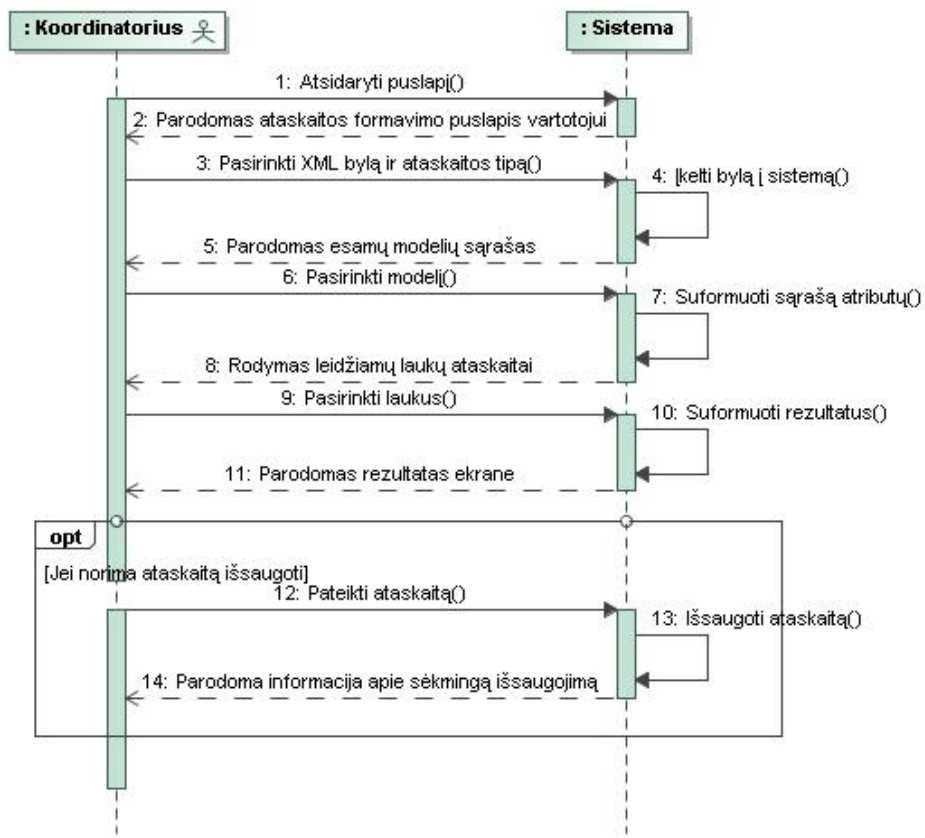
2.2 pav. Duomenų gavybos modelio sukūrimo sekų diagrama

Kai sukurtas bent vienas duomenų gavybos modelis, galima gauti ataskaitą su analizės rezultatu. Tai aprašyta panaudojimo atveju „Suformuoti ataskaitą pagal duomenų gavybos rezultatus“. Šis atvejis detalizuotas 2.5 lentelėje.

**2.5 lentelė.** Panaudojimo atvejo „Suformuoti ataskaitą pagal duomenų gavybos rezultatus“ aprašymas

<b>PA</b> „Suformuoti ataskaitą pagal duomenų gavybos rezultatus“		
<b>Tikslas.</b> Suformuoti ataskaitą pagal apmokytą duomenų gavybos modelį.		
<b>Aprašymas.</b> Koordinatorius gali sukurti ataskaitą, pagal jos kūrimo metu nurodytus parametrus su tiriamais duomenimis esančiais XML faile		
<b>Prieš sąlyga</b>	Koordinatorius prisijungęs savo paskyra	
<b>Aktorius</b>	Koordinatorius	
<b>Sužadinimo sąlyga</b>	Koordinatorius nori atlikti prognozę su duomenimis XML faile pagal anksčiau sukurtą duomenų gavybos modelį	
<b>Veiklos taisyklės</b>		
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>	Išsaugoti ataskaitą
	<b>Apima PA</b>	Paimti tiriamus duomenis
	<b>Specializuoja PA</b>	
<b>Pagrindinis įvykių srautas</b>		
1. Vartotojas pasirenka meniu pasirinkimą „Suformuoti ataskaitą“	Sistema parodo formą, kurioje reikia pasirinkti ataskaitos tipą ir bylą, iš kurios bus keliami duomenys. Toliau vykdoma panaudojimo atvejo „Paimti tiriamus duomenis“ veiksmas.	
2. Vartotojas pasirenka, pagal kurį duomenų gavybos modelį kurti ataskaitą	Sistema išskviečia formą, kurioje galima pasirinkti, kokius laukus rodyti ataskaitoje.	
3. Vartotojas pažymi norimus matyti laukus ataskaitoje ir spaudžia mygtuką „Vykdyti“	Sistema parodo lentelę su formoje nurodytais laukais	
4. Jei vartotojas nori išsaugoti ataskaitą, įrašo jos pavadinimą ir paspaudžia mygtuką „Išsaugoti“	Vykdoma panaudojimo atvejo „Išsaugoti ataskaitą“ veiksmas	
4. Sistema baigia PA		
<b>Po sąlyga</b>	Suformuota ataskaita	
<b>Pastabos</b>		

Atlikus šio panaudojimo atvejo veiksmus, galima išsaugoti gautą ataskaitą. Tam reikia vykdyti panaudojimo atvejo „Išsaugoti ataskaitą“ veiksmus. Šis panaudojimo atvejis bus detalizuotas vėliau 2.12 lentelėje. Abu šie panaudojimo atvejai yra pavaizduoti sekų diagramoje, kuri matoma žemiau esančiame 2.3 paveiksle.



2.3 pav. Ataskaitos suformavimo ir išsaugojimo sekų diagrama

Ataskaitos formavimo metu yra nuskaitoma XML duomenų byla. Ši funkcija aprašoma panaudojimo atveju „Paimti tiriamus duomenis“. Šio panaudojimo atvejo specifikacija pateikta 2.6 lentelėje.

2.6 lentelė. Panaudojimo atvejo „Paimti tiriamus duomenis“ aprašymas

<b>PA</b> „Paimti tiriamus duomenis“	
<b>Tikslas.</b> Sistema paima duomenis iš XML bylos.	
<b>Aprašymas.</b> Kai Koordinatorius formuoja ataskaitą, sistema nuskaityti tiriamuosius duomenis iš XML bylos ir sukelia juos į sistemą	
<b>Prieš sąlyga</b>	Koordinatorius pradėjo ataskaitos formavimo funkciją
<b>Aktorius</b>	Koordinatorius
<b>Sužadinimo sąlyga</b>	Pradėta ataskaitos formavimo funkcija
<b>Veiklos taisyklės</b>	
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>
	<b>Apima PA</b>
	<b>Specializuoja PA</b>
<b>Pagrindinis įvykių srautas</b>	
1. Vartotojas nurodo ataskaitos tipą ir bylą bei paspaudžia mygtuką „Vykdėti“	Sistema perskaityti duomenis XML faile ir perkelia juos į duomenų bazę. Toliau atliekami panaudojimo atvejo „Suformuoti ataskaitą pagal duomenų gavybos rezultatus“ veiksmai
2. Sistema baigia PA	
<b>Po sąlyga</b>	Nuskaityti duomenys iš XML bylos
<b>Pastabos</b>	

Norint pasinaudoti integravimo paslaugų transformacijomis teksto gavybai, reikia paleisti paketą, kuris atlieka gavybą. Šie veiksmai parodyti panaudojimo atveju „Atlikti teksto analizę“. Šis panaudojimo atvejis parodytas 2.7 lentelės pagalba.

2.7 lentelė. Panaudojimo atvejo „Atlikti teksto analizę“ aprašymas

<b>PA</b> „Atlikti teksto analizę“		
<b>Tikslas.</b> Atlikti teksto analizę integravimo paslaugų pagalba.		
<b>Aprašymas.</b> Koordinatorius gali atlikti teksto analizę, pasinaudodamas integravimo paslaugomis, bei gauti studijų dalyko įvertinimą pagal anketos atsakymų tekstus		
<b>Prieš sąlyga</b>	Koordinatorius prisijungęs savo paskyra	
<b>Aktorius</b>	Koordinatorius	
<b>Sužadinimo sąlyga</b>	Koordinatorius nori atlikti teksto studijų modulių kokybės įvertinimo anketos analizę.	
<b>Veiklos taisyklės</b>		
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>	
	<b>Apima PA</b>	Parodyti analizės rezultatą
	<b>Specializuoja PA</b>	
<b>Pagrindinis įvykių srautas</b>		
1. Vartotojas pasirenka „Atlikti teksto gavybą“ meniu pasirinkimą	Sistema parodo formą, kurioje paspaudus mygtuką atliekama analizė	
2. Vartotojas paspaudžia mygtuką „Vykdėti“	Sistema įvykdo transformacijas, kurios yra integravimo paslaugų paketo srautuose. Po to vykdoma „Parodyti analizės rezultatą“ panaudojimo atvejo veiksmai	
3. Sistema baigia PA		
<b>Po sąlyga</b>	Atlikta teksto analizė	
<b>Pastabos</b>		

Kai atlikti veiksmai paketo srantuose, parodomas teksto analizės rezultatas. Šie veiksmai aprašomi panaudojimo atvejo pagalba „Parodyti analizės rezultatą“. Šis panaudojimo atvejis aprašytas 2.8 lentelėje.

**2.8 lentelė.** Panaudojimo atvejo „Parodyti analizės rezultatą“ aprašymas

<b>PA</b> „Parodyti analizės rezultatą“	
<b>Tikslas.</b> Parodyti atliktos teksto gavybos rezultatą ekrane.	
<b>Aprašymas.</b> Po atliktos analizės internetiniame puslapyje parodoma studijų dalyko įvertinimai pagal atsakymus į atvirus klausimus.	
<b>Prieš sąlyga</b>	Pradėta vykdyti teksto analizę
<b>Aktorius</b>	Koordinatorius
<b>Sužadinimo sąlyga</b>	Koordinatorius pradėjo atlikti studijų anketų atsakymų teksto analizę.
<b>Veiklos taisyklės</b>	
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>
	<b>Apima PA</b>
	<b>Specializuoja PA</b>
<b>Pagrindinis įvykių srautas</b>	
1. Po to, kai atlikta teksto analizė, sistema parodo rezultatų lentelę	Sistema suformuoja lentelę, kurioje matomi studijų dalykai bei jų įvertinimai, kurie suskaičiuoti analizės metu
2. Sistema baigia PA	
<b>Po sąlyga</b>	Parodomi teksto gavybos rezultatai ekrane
<b>Pastabos</b>	

Norint sužinoti, kaip išsidėstę frazės tekstuose galima pasinaudoti funkcija sistemoje, kurios veiksmai aprašomi panaudojimo atveju „Rodyti frazių buvimo vietas tekste“. Šis panaudojimo atvejis pateiktas 2.9 lentelėje.

**2.9 lentelė.** Panaudojimo atvejo „Rodyti frazių buvimo vietas tekste“ aprašymas

<b>PA</b> „Rodyti frazių buvimo vietas tekste“	
<b>Tikslas.</b> Parodyti, kaip žodyno frazės yra išsidėstę tekstuose.	
<b>Aprašymas.</b> Koordinatorius ir Senato studijų komisijos nariai gali peržiūrėti, kur apklausos atsakymo tekste yra kiekviena iš frazių, kuri egzistuoja frazių žodyne.	
<b>Prieš sąlyga</b>	Koordinatorius arba Senato studijų komisijos narys prisijungia savo paskyra
<b>Aktorius</b>	Koordinatorius, Senato studijų komisijos nariai
<b>Sužadinimo sąlyga</b>	Koordinatorius arba Senato studijų komisijos narys nori peržiūrėti, kur tekste yra frazių žodyno žodžiai bei pamatyti įvertinimą.
<b>Veiklos taisyklės</b>	
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>
	<b>Apima PA</b>
	<b>Specializuoja PA</b>
<b>Pagrindinis įvykių srautas</b>	
1. Vartotojas pasirenka meniu pasirinkimą „Rodyti frazes tekste“	Sistema parodo formą, kurioje pasirinkus parametrus parodoma frazių vieta atsakymų tekstuose



2. Vartotojas pasirenka modulį, kokio klausimo nori matyti tekstus bei koku metodu nori nagrinėti tekstą	Sistema nuspalvina tekstą pagal pasirinktus parametrus
3. Sistema baigia PA	
<b>Po sąlyga</b>	Nuspalvintos žodyno frazės atsakymų tekstuose pasirinktu metodu.
<b>Pastabos</b>	

Sukūrus duomenų gavybos modelį, galima peržiūrėti jo grafinį vaizdą. Šios funkcijos veiksmi aprašomi panaudojimo atvejo „Peržiūrėti grafiškai duomenų gavybos modelį“ pagalba. Šis panaudojimo atvejis pateiktas 2.10 lentelėje.

**2.10 lentelė.** Panaudojimo atvejo „Peržiūrėti grafiškai duomenų gavybos modelį“ aprašymas

<b>PA „Peržiūrėti grafiškai duomenų gavybos modelį“</b>	
<b>Tikslas.</b> Peržiūrėti duomenų gavybos modelio grafinį vaizdą.	
<b>Aprašymas.</b> Koordinatorius arba Senato studijų komisijos nariai gali peržiūrėti duomenų gavybos modelio grafinį vaizdą. Šis vaizdas yra gaunamas toks pat, kaip matoma naudojantis Business Intelligence Studio programos pagalba.	
<b>Prieš sąlyga</b>	Koordinatorius arba Senato studijų komisijos nariai prisijungę savo paskyra
<b>Aktorius</b>	Koordinatorius, Senato studijų komisijos nariai
<b>Sužadinimo sąlyga</b>	Koordinatorius arba Senato studijų komisijos nariai nori pamatyti duomenų gavybos modelio grafinį vaizdą.
<b>Veiklos taisyklės</b>	
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>
	<b>Apima PA</b>
	<b>Specializuoja PA</b>
<b>Pagrindinis įvykių srautas</b>	
1. Vartotojas pasirenka meniu pasirinkimą „Grafinio modelio peržiūra“	Sistema parodo puslapį, kuriame reikia pasirinkti duomenų gavybos modelį, kurio vaizdą norime matyti
2. Vartotojas pasirenka duomenų gavybos modelį	Sistema parodo modelio grafinį vaizdą.
3. Sistema baigia PA	
<b>Po sąlyga</b>	Atlikta studentų analizė
<b>Pastabos</b>	

Senato studijų komisijos nariai gali peržiūrėti ataskaitą, parengtą koordinatoriaus. Ši funkcija aprašoma panaudojimo atveju „Peržiūrėti ataskaitą“. Jo specifikacija pateikta 2.11 lentelėje.

**2.11 lentelė.** Panaudojimo atvejo „Peržiūrėti ataskaitą“ aprašymas

<b>PA „Peržiūrėti ataskaitą“</b>	
<b>Tikslas.</b> Pamatyti Koordinatoriaus suformuotą ataskaitą.	
<b>Aprašymas.</b> Senato studijų komisijos narys gali peržiūrėti suformuotą ataskaitą, kuri anksčiau buvo suformuota Koordinatoriaus	
<b>Prieš sąlyga</b>	Senato studijų komisijos narys prisijungęs savo paskyra
<b>Aktorius</b>	Senato studijų komisijos narys
<b>Sužadinimo sąlyga</b>	Senato studijų komisijos narys nori peržiūrėti ataskaitą.
<b>Veiklos taisyklės</b>	

<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>	
	<b>Apima PA</b>	
	<b>Specializuoja PA</b>	
<b>Pagrindinis įvykių srautas</b>		Sistemos reakcija ir sprendimai
1. Pasirenkamas meniu pasirinkimas „Rodyti ataskaitą“		Sistema parodo formą, kurios pagalba galima pasirinkti kokią ataskaitą norime matyti
2. Vartotojas pasirenka, kokią ataskaitą nori matyti ir paspaudžia mygtuką „Rodyti“		Sistema parodo ataskaitą ekrane
3. Sistema baigia PA		
<b>Po sąlyga</b>		Ekrane parodoma suformuota ataskaita

Kai Koordinatorius suformuoja ataskaitą, yra galimybė ją išsaugoti. Tokiu atveju vykdoma funkcija, kurią aprašo panaudojimo atvejis „Išsaugoti ataskaitą“. Šis panaudojimo atvejis specifikuotas 2.12 lentelėje

**2.12 lentelė.** Panaudojimo atvejo „Išsaugoti ataskaitą“ aprašymas

<b>PA „Išsaugoti ataskaitą“</b>		
<b>Tikslas.</b> Išsaugoti ataskaitą, kurią vėliau galės peržiūrėti Senato studijų komisijos nariai.		
<b>Aprašymas.</b> Koordinatorius gali išsaugoti anksčiau suformuotą ataskaitą, kurią galės pasižiūrėti Senato studijų komisijos narys		
<b>Prieš sąlyga</b>	Suformuota ataskaita	
<b>Aktorius</b>	Koordinatorius	
<b>Sužadinimo sąlyga</b>	Koordinatorius suformavo ataskaitą ir nori ją išsaugoti	
<b>Veiklos taisyklės</b>		
<b>Susiję panaudojimo atvejai</b>	<b>Išplečia PA</b>	
	<b>Apima PA</b>	
	<b>Specializuoja PA</b>	
<b>Pagrindinis įvykių srautas</b>		Sistemos reakcija ir sprendimai
1. Nurodomas ataskaitos vardas ir paspaudžiamas pasirinkimas „Išsaugoti“		Sistema išsaugoja ataskaitą ir parodo informaciją apie išsaugojimą
2. Sistema baigia PA		
<b>Po sąlyga</b>		Išsaugota ataskaita, kurią galės peržiūrėti Senato studijų komisijos narys

### 2.1.2. Kuriamo sprendimo studijų analizei nefunkciniai reikalavimai

Sistema turi būti realizuota, naudojant Microsoft SQL Server 2008 duomenų bazių valdymo sistemą. Duomenų gavybos modeliai saugojami Analizės paslaugų serveryje, ataskaitos formuojamos pasinaudojant ASP.NET karkaso suteikiamų komponentų pagalba.

Pati sistema veiks ją įdiegus į IIS serverį. Naudojamas .NET karkasas, objektiškai programuojama C# kalba. Vartotojas sistemą pasiekia, pasinaudojant interneto naršykle. Ji gali būti bet kuri dabar naudojama populiariausia naršyklė.

Sistemos stilių formuoti naudojama CSS stilių kalba, visi puslapiai turi turėti vieningą dizainą ir panašų elementų išdėstymą.

Naudojantis sistema turi būti kuo mažesnė galimybė sugadinti duomenų gavybos modelių struktūrą, todėl duomenų gavybos modelių grafinės peržiūros galimybės turi būti įkeltos į .NET puslapį tam, kad nereikėtų naudotis Business Intelligence Studio programa.

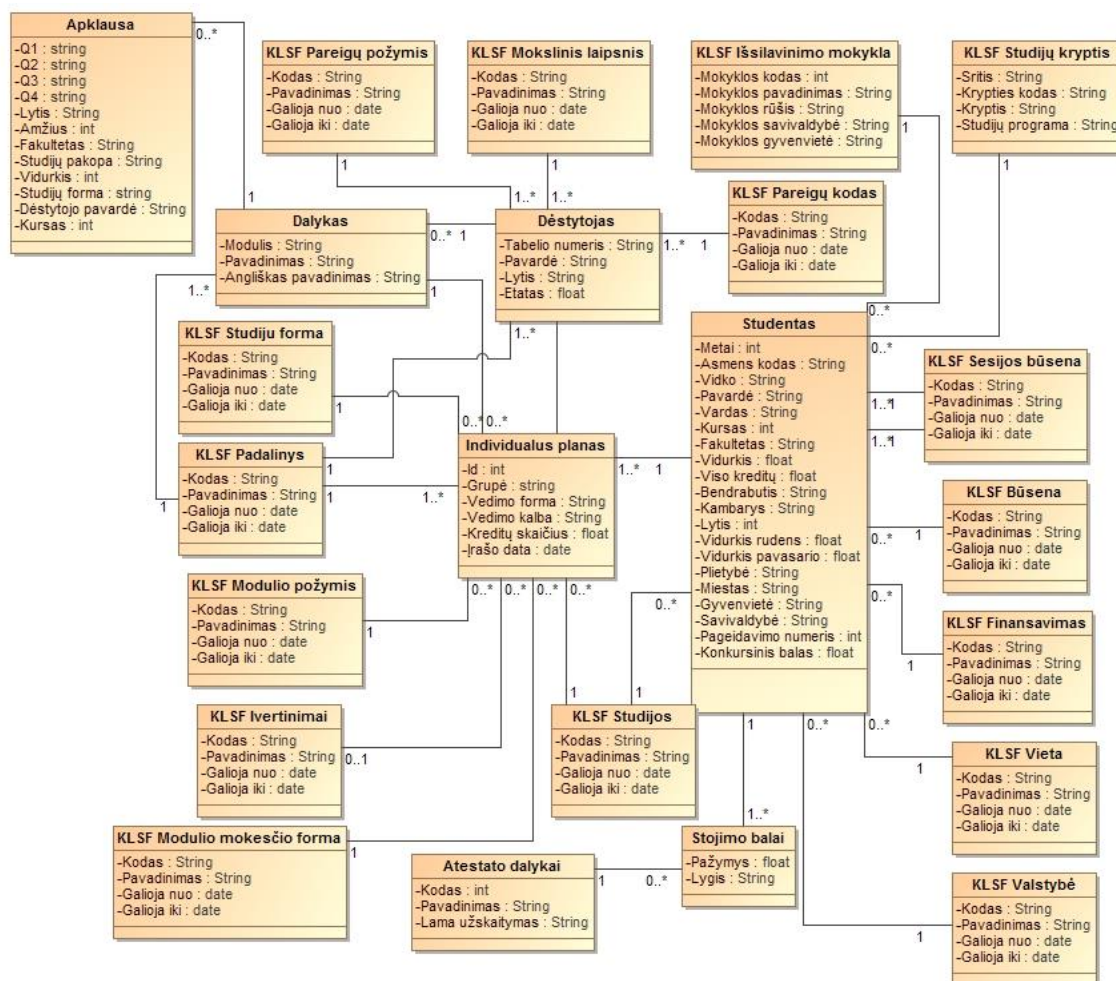
Duomenų gavybos rezultatus turi būti galimybė peržiūrėti ne vien tekistine forma, bet ir peržiūrėti modelio grafinį vaizdą, jeigu toks vaizdas egzistuoja algoritme.

Ataskaitos turi būti formuojamos naudojant APS.NET karkaso galimybes. Tai daroma tam, kad būtų geras suderinamumas su visomis kitomis sprendimo dalimis bei nereikėtų papildomai konfigūruoti ataskaitų paslaugų serverio.

Visi duomenys vedami iš formų puslapiuose turi būti patikrinami tam, kad nebūtų galima pavogti konfidencialios informacijos iš sistemos.

## 2.2. Analizuojamos dalykinės srities modelis

Sukurti vaizdinius bei apmokyti duomenų gavybos modelius bus naudojami duomenys iš duomenų bazės. Dalykinės srities modelis, iš kurio vėliau gauta duomenų bazės schema, sudaryta iš esybių, kurios atvaizduoja studentų duomenis universitete ir duomenis iš mokyklos, taip pat informaciją, kas dėstė dalykus universitete ir kokio padalinio dėstytojai yra dėstę tuos dalykus. Šios esybės ir ryšiai tarp jų pavaizduoti 2.4 paveiksle.



2.4 pav. Dalykinės srities modelis

Kai kurie iš duomenų yra sukelti klasifikatoriuose. Tokie klasifikatoriai paprastai turi savo kodą, pavadinimą ir datas nuo kada iki kada galioja. Tokiu būdu nereikia pagrindinėse lentelėse, tokiose kaip studentas, saugoti keletą laukų.

### 2.3. Duomenų gavybos praplėtimo kalbos naudojimas

Duomenų gavybos informacijai pasiimti iš duomenų gavybos modelių bei ją atvaizduoti turi būti naudojama duomenų gavybos praplėtimo (angl. *Data mining extensions (DMX)*) kalba. Kai kurie veiksmai su struktūromis, modeliais ir duomenimis gali būti atliekami ir nepasinaudojant šia kalba. Galima pasinaudoti Business Intelligence Studio programos vartotojo sąsajos suteikiamomis galimybėmis. Tačiau naudojantis tik grafinėmis priemonėmis visų veiksmų atlikti nėra galimybės.

Taip pat DMX užklausų kalba gali būti lengvai naudojama, kuriant įvairius duomenų gavybos taikymus. Visas šios užklausų kalbos galimybes galima suskirstyti į 7 kategorijas:

- Duomenų šaltinio atvejų peržiūros veiksmai.
- Veiksmai su turinio peržiūra.
- Veiksmai, naudojant sprendimo medžių algoritmą.
- Veiksmai, naudojant laiko serijų algoritmą.
- Veiksmai su grupavimo algoritmu.
- Veiksmai su asociacijos taisyklių ir sekų grupavimo algoritmais.
- Duomenų apibrėžimo funkcijos.

Su kiekvienu algoritmu yra nemažai galimybių. Kai kurios iš jų galimos ir nenaudojant DMX kalbos. Toliau yra pateikta 2.13 lentelė, kurioje parodytos kai kurios iš visų kategorijų savybių. Šalia nurodyta, ar tą savybę galima panaudoti ir su grafine vartotojo aplinka, ar tik pasinaudojant duomenų gavybos užklausų kalbą.

2.13 lentelė. Duomenų gavybos užklausų kalbos panaudojamumo palyginimas su vartotojo sąsaja

Tipas	Savybė	Pasinaudojant grafine aplinka	Naudojant DMX
Duomenų šaltinio atvejų peržiūros veiksmai	Peržiūrėti duomenų šaltinio turinį	Galima, kai naudojama paprasta lentelė	Galima naudojant lenteles, sudarytas iš kitų lentelių
	Peržiūrėti duomenis, kurie naudojami kaip apmokymo duomenys	Negalima	Galima
	Peržiūrėti tik tam tikrus duomenų atvejus	Negalima	Galima
	Peržiūrėti tik testinius duomenis	Negalima	Galima
	Peržiūrėti duomenis tik vienos duomenų gavybos grupavimo algoritmo grupės	Galima, jeigu nereikia nurodyti sąlygos, ką rodyti ar rikiuoti	Galima
	Peržiūrėti sprendimo medžio vieno elemento duomenis	Galima, jeigu nereikia nurodyti sąlygos, ką rodyti ar rikiuoti. Sąlyga gali būti tik tokia, kokia užrašyta medžio viršūnėje	Galima

Tipas	Savybė	Pasinaudojant grafine aplinka	Naudojant DMX
	Peržiūrėti sekų modelio duomenis	Galima, tačiau neišskleidus lentelių lentelėse	Galima ir išskleidus
	Panaudoti rikiavimą arba pirmųjų įrašų peržiūrą	Negalima	Galima
Veiksmai su turinio peržiūra	Pamatyti tarp grupių ir kitų besisiejiančių algoritmuose elementų ryšius	Galima labai netiksliai pagal grafinį vaizdą	Galima labai detalai
	Pakeisti grupės pavadinimą	Galima	Galima
	Galima panaudoti where sąlygas ar rikiavimą	Negalima	Galima
	Galima atlikti detalesnę peržiūrą tokią, kaip išskleisti lentelę lentelėje į paprastą lentelę ar atlikti užklausas į detalesnę lentelę	Negalima	Galima
	Peržiūrėti duomenis, naudojant įvairias funkcijas	Negalima	Galima
Sprendimo medžio veiksmai	Atlikti prognozę, naudojant prognozavimo užklausą ir funkcijas	Galima, kai nėra užklausa labai sudėtinga	Galima
	Ar galima naudoti prognozės užklausose sąlygas ir rikiavimus	Galima	Galima
	Singletono užklausa	Negalima	Galima
Veiksmai su laiko serijų algoritmu	Ar galima analizuoti turimus duomenis laiko serijų algoritme	Galima	Galima
	Ar galima analizuoti turimus duomenis laiko serijų algoritme naudojant sąlygas	Negalima	Galima
	Ar galima naudoti užklausas, kurios leidžia matyti, kas vyko prieš tam tikrą laiko tarpą	Galima, tačiau ne visais atvejais	Galima
	Ar galima pranašauti laiko serijoje, kas įvyks po kelių laiko vienetų	Galima, jeigu užklausa nesudėtinga	Galima
	Ar galima rasti nuokrypį nuo teisingų duomenų	Galima, jeigu užklausa nesudėtinga	Galima
	Ar įmanoma surasti vieną parametą jei žinoma, kaip keisis kitas laiko serijų algoritme	Negalima	Galima
Veiksmai su grupavimo algoritmu	Ar galima pamatyti grupavimo algoritme grupės narius	Galima grafiškai	Galima
	Ar galima pamatyti, kokia tikimybė, kad viena grupė siesis su kita	Galima labai apytiksliai pagal grafinį vaizdą	Galima
	Ar galima pažiūrėti, kokia tikimybė, kad tam tikras dalykas bus grupėje	Galima pagal grafinį vaizdą	Galima
	Ar galima pažiūrėti, kaip tikėtina, kad dalykas turintis tam tikrus	Negalima	Galima

Tipas	Savybė	Pasinaudojant grafine aplinka	Naudojant DMX
	parametrus, bus grupėje		
	Ar galima prognozę atlikti ir grupavimo algoritme	Galima su nesudėtinga užklausa ir jei nenaudojamos fiksuotos parametro reikšmės	Galima
Veiksmai su asociacijos ir sekų grupavimo algoritmais	Ar galima asociacijos algoritme rasti daiktų rinkinį	Galima	Galima
	Ar galima pamatyti tik taisykles nerodant rinkinių	Negalima	Galima
	Ar galima rodyti surikiuotas pagal svarbumą arba keletą pirmųjų taisyklių	Negalima	Galima
	Ar galima rodyti taisykles tik su pasirinktu elementu	Galima, bet parodoma ir pašalinių įrašų	Galima
	Ar galima rasti daugiausia kartų pasikartojančius dalykus kartu	Negalima	Galima
	Ar galima vykdyti analizę, kuri parodo, kas esant vienam dalykui labai tikėtina, kad taip pat bus kitas dalykas	Negalima	Galima
	Ar galima sekų algoritme pamatyti tam tikrą skaičių iš eilės vykusių dalykų	Negalima	Galima
Duomenų apibrėžimo funkcijos (DDL)	Sukurti duomenų gavybos struktūrą	Galima kuriant pirmą algoritmą	Galima
	Sukurti duomenų gavybos modelį	Galima	Galima
	Apmokyti duomenų gavybos modelį	Galima	Galima
	Peržiūrėti struktūros duomenis	Galima, jei naudojamos paprastos lentelės, o ne OLAP kubai	Galima
	Ar galima nurodyti, kiek procentų duomenų bus naudojama testuoti modelį	Galima	Galima
	Ar galima nurodyti modelio parametrų reikšmes	Galima	Galima
	Ar galima nurodyti filtrą, kokius duomenis naudoti apmokymui	Negalima	Galima
	Ar galima ištrinti modelius	Galima	Galima
	Ar galima panaikinti iš modelio visus duomenis (panaikinti apmokymą)	Negalima	Galima
	Ar galima pervadinti modelius ir struktūras	Galima	Galima
	Ar galima pasidaryti modelių atsargines kopijas ir vėliau iš jų atkurti	Negalima	Galima

Tipas	Savybė	Pasinaudojant grafine aplinka	Naudojant DMX
	Ar galima sukurti modelį su lentelėmis atribute	Galima	Galima

Dėl to, kad daugelis duomenų gavybos galimybių yra prieinami tik per DMX užklausas ir dėl to, kad realizuojant patogų kreiptis į modelius užklausių pagalba bus naudojama duomenų gavybos praplėtimo kalba.

## 2.4. Reikalavimų analizės apibendrinimas

Išsiaiškinus visus reikalavimus kuriamam produktui, nuspręsta, kad realizuoti jį reikės trimis posistemiais, nes sistemos funkcijos pasiskirsto pagal turimus tris vartotojo tipus. Vienas posistemis atsakingas už analizių atlikimą, antrasis - už duomenų gavybos modelių tvarkymą bei duomenų sukėlimą ir trečiasis - už ataskaitų peržiūrą. Kiekvienam iš šių posistemių numatyta po vieną vartotojo tipą. Sąveika su duomenų gavybos modeliais bus atliekama, pasinaudojant DMX užklausių kalbą.

## 3. UNIVERSITETO STUDIJŲ KOKYBĖS IR STEBĖSENOS PROCESUS GERINANČIOS SISTEMOS PROJEKTAS

Projekto tikslas sumodeliuoti sistemos architektūrą, išsiaiškinti, iš ko turi susidėti visi posistemiai. Kaip, atliekant įvairius veiksmus sistemoje, jos dalys bendrauja tarpusavyje. Pagal tai yra nusprendžiama, kokie komponentai turi būti panaudoti kuriant sprendimą. Dar vienas tikslas yra suprojektuoti duomenų bazės, kurioje bus saugomi studentų duomenys, schemą.

### 3.1. Studijų proceso tyrimo sprendimo reikalingumas

Universitetui kaip ir visoms organizacijoms yra svarbu, kaip kokybiškai yra atliekama veikla. Tai galima sužinoti, tik atliekant įvairius tyrimus ar auditą. Viena iš priemonių, kuri gali būti panaudota šiam tikslui, yra verslo analitika. Šis būdas tampa vis labiau populiariesne priemone įvairiems tyrimams. Jie Kauno technologijos universitete pradėti atlikti, pasinaudojant šiuolaikiškais priemonėmis, tokiomis, kaip duomenų gavyba, realiu laiku veikiančios švieslentės ar daugiamaciai kubai. Šiems sprendimams sukurti reikia pasinaudoti ne viena skirtinga programa iš SQL Server duomenų bazių valdymo sistemos. Tokiu būdu darbas vartotojui tampa sudėtingas, atsiranda didelė klaidos tikimybė.

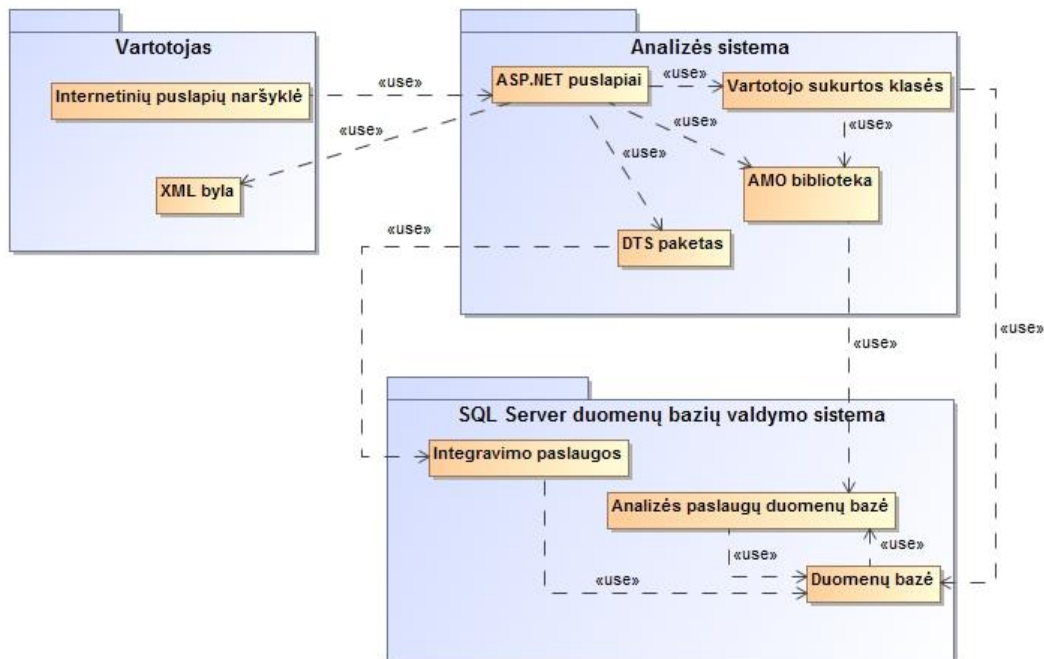
Norint išspęsti šią problemą, reikia turėti priemonę, kuri leistų atlikti analizes paprasčiau, rodant patogią vartotojo sąsają, kurią žymiai paprasčiau suprasti, palyginus su programomis, esančiomis duomenų bazių valdymo paketuose. Tai patogų atlikti, pasinaudojant internetinėmis technologijomis, internetinių puslapių pagalba. Kadangi šie tyrimai atliekami naudojant Microsoft programinę įrangą, patogiausia naudoti ASP.NET karkasą internetinėms formoms kurti. Tokiu būdu vartotojas, naudodamasis interneto naršykle, galės kurti reikiamas struktūras, pamatyti suformuotas ataskaitas. Visi modeliai, esantys analizės paslaugų serveryje, nebus tiesiogiai pasiekiami, todėl sumažės tikimybė juos sugadinti.

Toliau bus aprašyta, iš kokių dalių sudarytas sprendimas, kaip tos dalys sąveikauja tarpusavyje bei kiti sistemos architektūrai ir realizacijai atlikti pasirinkimai.

## 3.2. Panaudoti sprendimai kuriant sistemą

### 3.2.1. Sistemos loginė architektūra

Sistema yra sudaryta iš standartinės trijų lygių architektūros. Numatyta sukurti tris anksčiau minėtus posistemius ir papildomą pagrindinį posistemį, kurio paskirtis nukreipti vartotoją į sistemos dalis. Visi šie posistemiai yra realizuoti asp.net internetinių formų pagalba. Jų klasių failuose yra sukurti veiksmai, atliekantys visą veiklos logiką bei besikreipiantys į programuotojo parašytas klases. Šios klasės kreipiasi į duomenų bazių valdymo sistemą gauti ar pakeisti informaciją joje. Manipuliuoti su analizės objektais panaudota AMO biblioteka. Jos pagalba sukuriami visi reikiami analizės objektai: struktūros modeliai ir kiti elementai analizės paslaugų serveryje. Teksto analizei panaudota integravimo paslaugų paketas. Abi paminėtosios SQL Server paslaugos naudoja duomenų bazės variklį duomenims gauti. Ši architektūra pavaizduota toliau pavaizduotame 3.1 paveiksle.



3.1 pav. Panaudota sistemoje architektūra

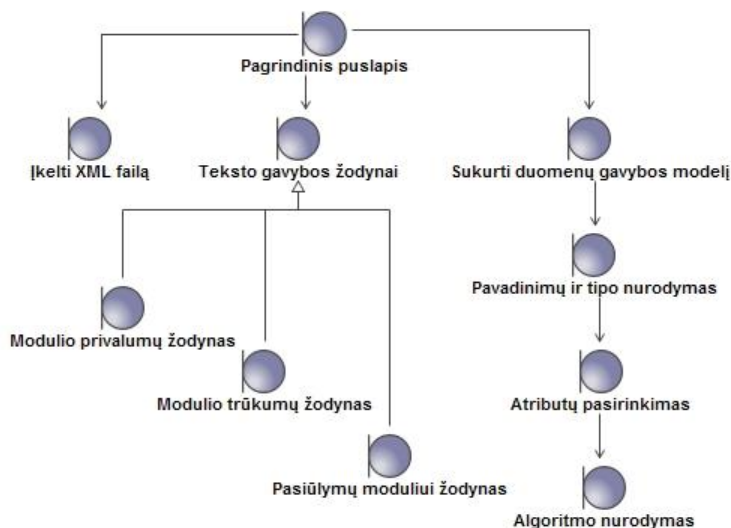
Toliau sekančiuose skyreliuose detaliau bus aprašytos šios pasirinktos architektūros dalys.

### 3.2.2. Vartotojui atvaizduojama internetinė sąsaja

Vartotojas naudosis sistema internetinės naršyklės pagalba. Kadangi sistema kuriama .NET karkaso pagrindu, naršyklėje bus atvaizduojami aspx formato failai. Šiuose puslapiuose lentelių pagalba bus parodomi analizių rezultatai ir kita informacija. Informacija pateikiama, užpildant informaciją pateiktose formose. Kiekvienas vartotojo tipas matys tik tas sistemos funkcijas, kurios jam buvo numatytos reikalavimų dalyje. Kiekvieną iš pasiekiamų veiksmų besinaudojantis sistema pasieks vartotojo meniu pagalba. Kaip vartotojas keliauja iš vienos formos į kitą, patogu parodyti navigavimo planu.

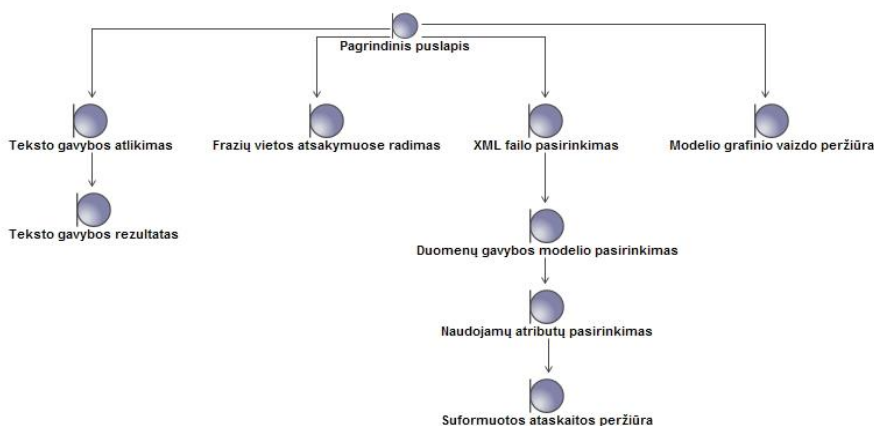


Toliau pateikti kiekvieno iš vartotojų navigavimo planai. Pirmasis vartotojas - administratorius. Jis gali į sistemą įkelti XML bylą, administruoti teksto analizės frazių žodynus bei sukurti analizei reikalingus modelius. Teksto analizės žodynai yra atskiri pagal kiekvieną iš apklausos klausimų. Kuriant modelį reikia pereiti per kelias formas, kurių dėka vartotojas nurodo visus reikiamus parametrus duomenų gavybos struktūros ir modelio sukūrimui. Visi šie puslapiai pavaizduoti navigavimo plane matomame 3.2 paveiksle.



3.2 pav. Administratoriaus navigavimas tarp sistemos puslapių

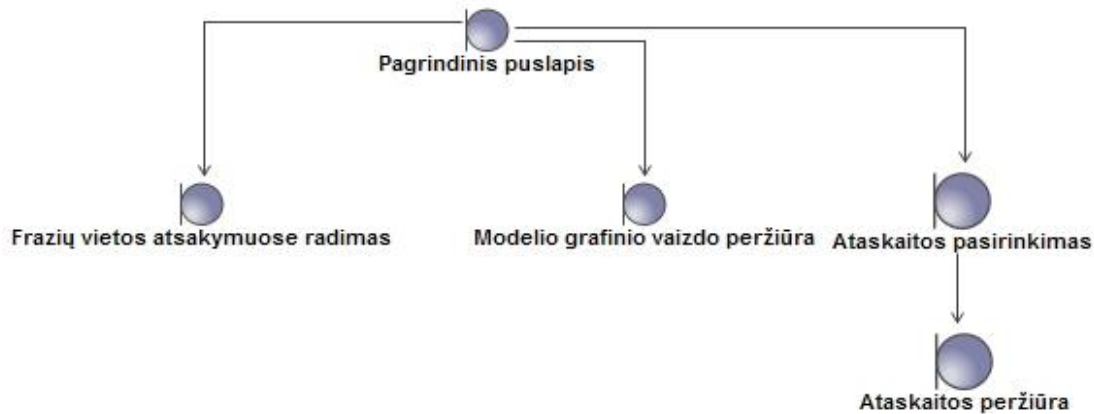
Koordinatoriaus atliekami veiksmai yra: teksto gavybos atlikimas, frazių peržiūrėjimas atsakymų tekstuose, ataskaitos suformavimas bei duomenų gavybos grafinio vaizdo peržiūra. Atliekant teksto gavybą, pirmiausia parodoma viena forma, kurioje reikia patvirtinti atlikimą mygtuko paspaudimu, vėliau parodomas kitas puslapis su suskaičiuotu įverčiu. Formuojant ataskaitą reikia pereiti per pora formų, kuriose nurodomi būtini ataskaitai parametrai: iš kokio XML failo naudosime duomenis, kokį gavybos modelį panaudosime, kokie atributai turėtų būti įtraukti į ataskaitą. Šio vartotojo navigavimo planas pateiktas 3.3 paveiksle.



3.3 pav. Koordinatoriaus navigavimas tarp sistemos puslapių

Trečiojo vartotojo tipo, Senato studijų komisijos nario, atliekami veiksmai šiek tiek sutampa su Koordinatoriaus. Abu šie vartotojai gali pasižiūrėti, kaip frazės yra išsidėstę anketų atsakymų į

klausimus tekstuose, taip pat abu tipai gali peržiūrėti duomenų gavybos modelio grafinį vaizdą. Unikali šiam vartotojo tipui suteikta funkcija yra ataskaitos peržiūra. Pirmiausia Senato studijų komisijos narys pasirenka, kokią ataskaitą nori matyti, tada parodoma pati ataskaita. Šio vartotojo navigavimo planas yra pateiktas 3.4 paveiksle.



3.4 pav. Senato studijų komisijos nario navigavimas tarp sistemos puslapių

Toliau bus aptarta, kaip yra suformuojami duomenų gavybos struktūros, modeliai bei atliekamas jų apmokymas, pasinaudojant aprašytais internetiniais puslapiais.

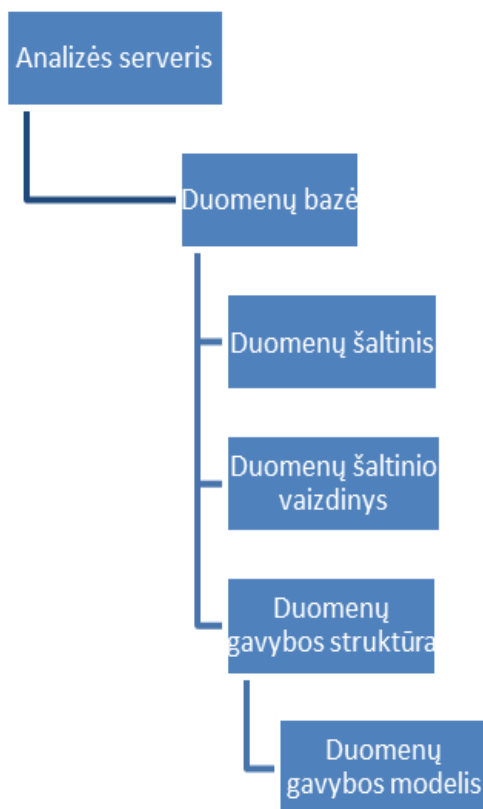
### 3.2.3. Duomenų gavybos struktūros bei modelio suformavimas, pasinaudojant ASP.NET karkasu bei DMX užklausomis

Prieš atliekant bet kokias analizes pasinaudojant duomenų gavyba, reikia parengti duomenis. Kai duomenys yra sukelti į duomenų bazę, reikia sukurti duomenų gavybos struktūrą bei modelį su pasirinktu duomenų gavybos algoritmu. Anksčiau naudotose realizacijose reikėdavo, pasinaudojant SQL Server Business Intelligence Studio programos pagalba, visa tai atlikti. Darbuotojams, kurie nėra informacinių technologijų specialistai, tai atlikti gali būti labai sudėtinga dėl naudojamų priemonių sudėtingumo.

Kadangi duomenų gavybai naudojamos Microsoft technologijos, labai patogu sukurti vartotojo sąsają, pasinaudojant ASP.NET karkasu. Jo pagalba sukurtais puslapiais vartotojas labai paprastai užpildo informaciją pateiktose formose ir, paspausdamas keletą mygtukų, sukuria reikiamą struktūrą ir modelį. Besinaudojantysis sistema turi pasirinkti lenteles ir laukus, kuriuos nori naudoti analizėje. Ši informacija gaunama, pasinaudojant duomenų bazės informacine schema, kurioje yra saugojama informacija apie lenteles ir atributus, iš kurių ji susideda. Informacija iš internetinėje naršyklėje matomų formų įrašoma į laikiną lentelę duomenų bazėje.

Tuo pačiu metu yra formuojamas sukinyš. Jis reikalingas dėl to, kad analizuojant įvertinimus, kuriuos turi studentas, yra kuriamos kelios to pačio studento duomenų eilutės – kiek yra įverčių, tiek ir eilučių. Tuo tarpu sukinyš parodo apibendrintus duomenis. Tokiu atveju modulio identifikatorius tampa lentelės atributu. Tik turint duomenis tokiu pavidalu, galima atlikti korektišką analizę.

Realizacijoje naudojamame ASP.NET karkase yra Analizės objektų valdymo (*angl. Analysis Management Objects*) biblioteka, kurią toliau sutrumpintai vadinsime AMO. Ji turi priemones skirtas sukurti visus Microsoft SQL Server analizės paslaugų objektus. Šie objektai yra skirti ne tik atlikti duomenų gavybą, bet ir analizuoti duomenis, naudojantis OLAP kubus. Pati duomenų gavyba vykdoma naudojantis Microsoft technologijomis, susideda iš objektų. Visi šie objektai, skirti atlikti duomenų gavybą, matomi 3.5 paveiksle.[22]

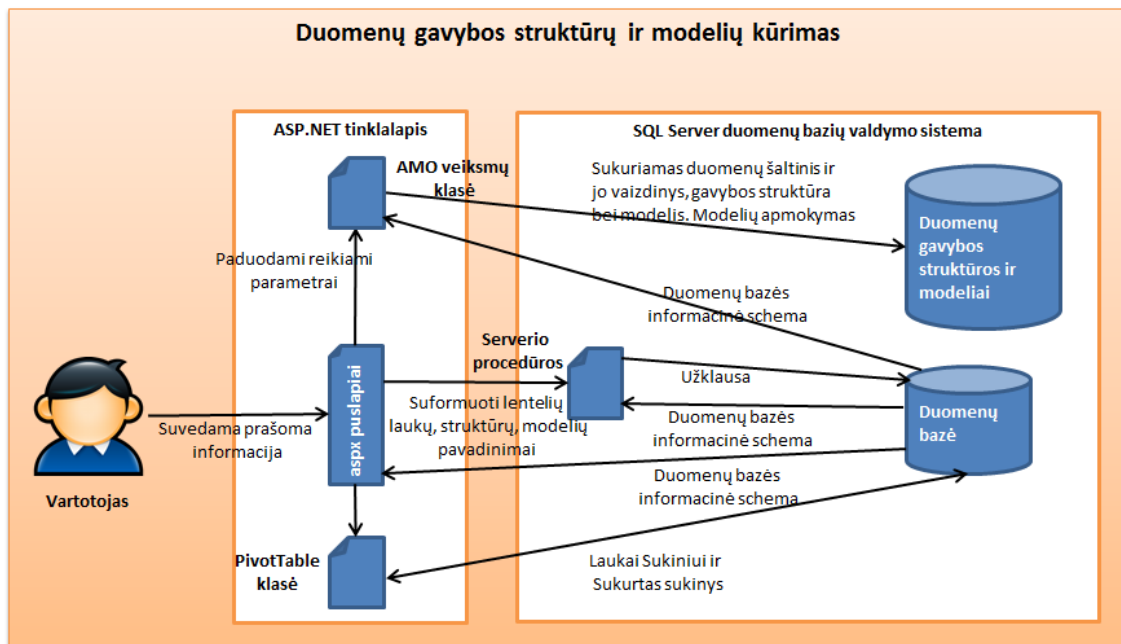


**3.5 pav.** Analizės paslaugų objektai, skirti atlikti duomenų gavybą

Prieš pradėdant bet kokius veiksmus su analizės paslaugomis, reikia nurodyti duomenų šaltinį. Jame reikia nurodyti, kokias lenteles ar vaizdinius naudosime duomenų gavybai. Tam yra sukuriama Duomenų šaltinio vaizdinys. Turint šiuos du objektus, galima sėkmingai kurti duomenų gavybos struktūras ir jose duomenų gavybos modelius.

Anksčiau minėtos AMO bibliotekos pagalba įvykdomi visi kūrimo ar koregavimo veiksmai. Ji sudaryta iš hierarchinės struktūros, kuri tokia pati, kaip buvo parodyta ankstesniame paveikslėlyje. Norint sukurti žemesniame lygyje esančius objektus, privalo būti sukurti aukštesnio lygio objektai. Pavyzdžiui, norint sukurti duomenų gavybos modelį, reikia turėti struktūrą, kadangi į modelio kūrimo konstruktoriaus operaciją reikia nurodyti struktūros pavadinimą.

Grafiškai duomenų gavybos struktūros ir modelio sukūrimą galima parodyti iliustracija, kurioje matomi svarbiausi elementai bei kas perduodama tarp jų. Visa tai matoma 3.6 paveiksle.



3.6 pav. Duomenų gavybos struktūros ar modelio sukūrimas

Antrasis variantas yra naudoti DMX užklausas. Tačiau šis metodas turi trūkumą. Sukūrus struktūrą ir modelius negalima nurodyti duomenų šaltinio bei jo vaizdo. Jie iš anksto turi būti sukurti analizės serveryje. Todėl negalima pilna realizacija pasinaudojant ASP.NET karkasu.

DMX duomenų gavybos struktūros sukūrimo užklausa, kuri turi būti perduodama į DMXCLR.dll klasę, turi būti tokios struktūros, kaip matoma 3.7 paveiksle.

```

Create mining structure [Struktūra]
(
  [VIDKO] long key,
  [VIDURK] text continuous,
  [BUSENA] text discrete
)

```

**Kuriamos struktūros pavadinimas**

**pavadinimas, duomenų tipas, reikšmių tipas**

3.7 pav. Duomenų gavybos struktūros sukūrimo DMX užklausa

Kai sukurta struktūra, toliau reikia sukurti modelį. Jis kuriamas su užklausa, kurioje nurodoma, į kokią struktūrą jį sukurti, kokį algoritmą taikyti bei kurie laukai bus prognozuojami. Tokia užklausa yra sekančios struktūros, matomos 3.8 paveiksle.

```

Alter mining structure [Struktūra]
Add mining model [Modelis]
(
  [VIDKO],
  [VIDURK],
  [BUSENA] predict
)
Using microsoft_clustering

```

**Struktūra, kuriai norime sukurti modelį**

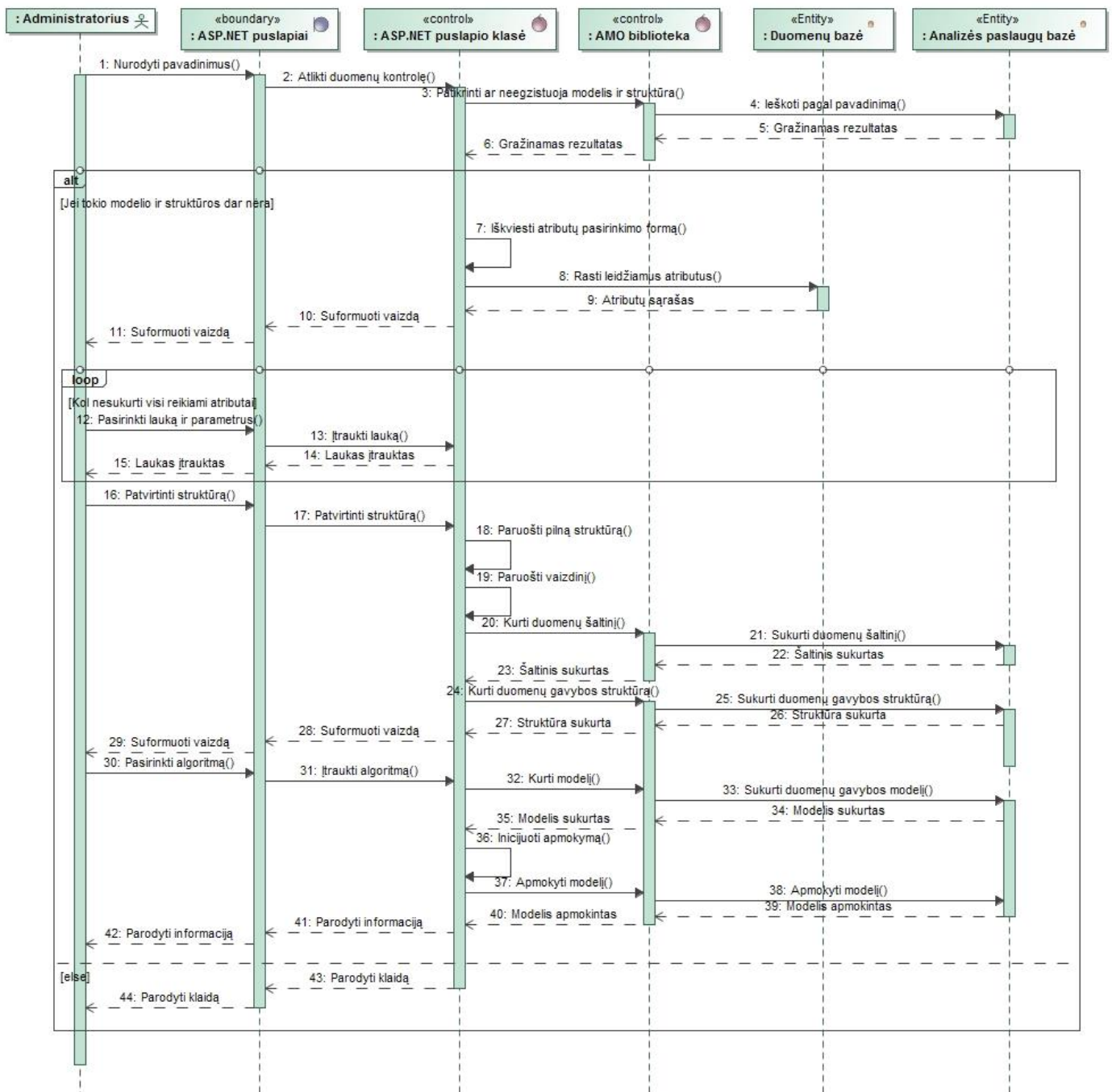
**Norimo sukurti modelio pavadinimas**

**Atributo pavadinimas, prognozuojamumo požymis**

**Naudojamo algoritmo pavadinimas**

3.8 pav. Duomenų gavybos modelio sukūrimo DMX užklausa

Patogumo dėlei tam, kad detaliau pamatyti kaip sukuriama duomenų gavybos struktūros, modeliai ir galiausiai atliekamas jų apmokymas yra pateikta sekų diagrama 3.9 paveiksle.



3.9 pav. Duomenų gavybos objektų sukūrimo sekos diagrama

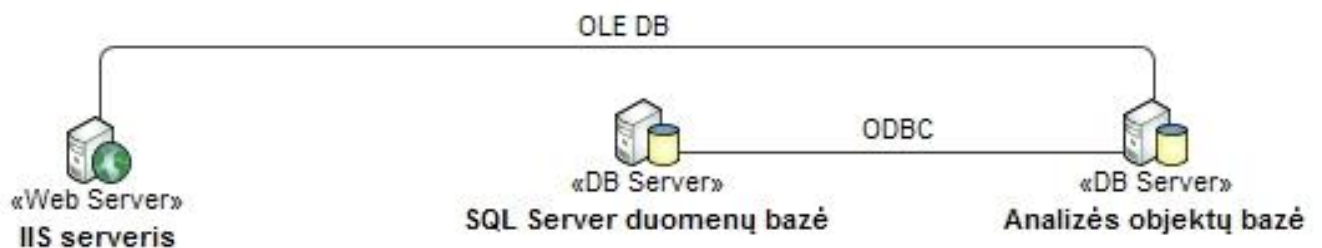
Reikia pastebėti, kad šiame modelyje yra atvaizduotas apibendrintas atvejis, kaip sprendime vyksta duomenų mainai tarp internetinio puslapio, šio puslapio klasės, AMO bibliotekos bei duomenų bazės valdymo sistemos serverių. Norint gauti duomenis iš modelių, pasinaudoti AMO biblioteka nebegalime, kadangi ši biblioteka skirta atlikti analizės objektų pakeitimus. Norint gauti prognozes ir rezultatus, reikia naudoti DMX užklausų kalbą. Kaip tai atliekama sprendime, yra pateikta sekančiame skyrelyje.

### 3.2.4. Duomenų gavybos rezultatų ištraukimas, pasinaudojant DMX užklausų kalba

Su anksčiau aprašyta DMX užklausų kalba galima ištraukti rezultatus, kurie gaunami, paduodant algoritmui tiriamuosius duomenis. Norint perduoti duomenis, reikia prisijungti prie SQL Server analizės paslaugų serverio. Tai galima atlikti dviem būdais. Pirmasis būdas yra pasinaudoti nutolusio serverio prisijungimo galimybe, kai pati duomenų bazių valdymo sistema jungiasi prie analizės serverio. Antrasis būdas yra pačiame ASP.NET karkasu realizuotame interneto tinklapyje pasinaudoti Ole Db prisijungimu.

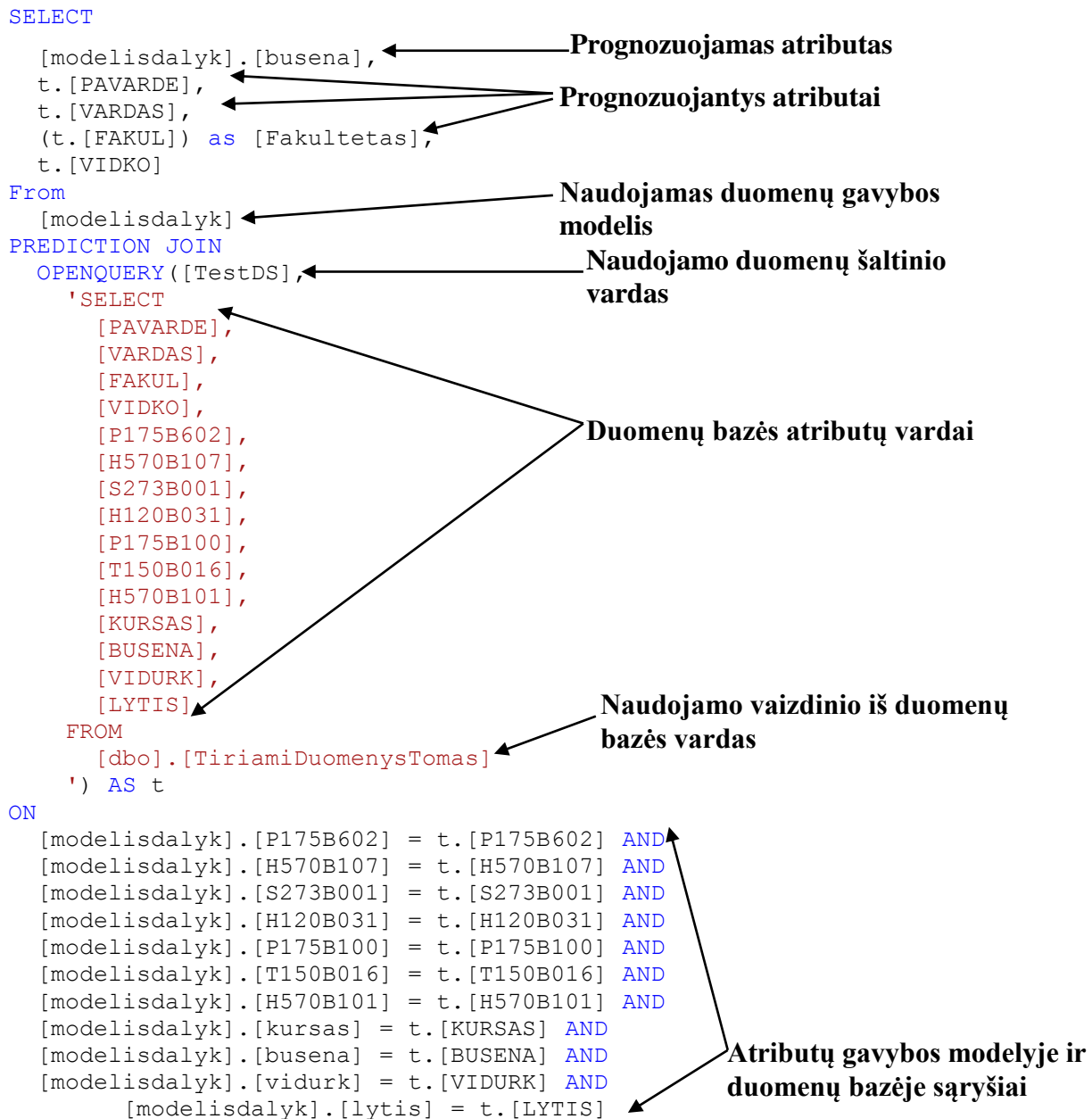
Sprendime pasirinkta naudoti antrąjį būdą, kadangi naudojant pirmąjį užklausa tampa labai komplikotos dėl jos sintaksės. Susidaro problema su kabučių naudojimu – standartinės SQL užklausa apjungtos su DMX užklausa turi būti panaudotos per C# programavimo kalba parašytas funkcijas, tenka naudoti perteklines kabutes, užklausa tampa labai sudėtingomis. Tuo tarpu antrame variante iš karto rašoma DMX užklausa.

Šiuo būdu IIS serveris per Ole Db sąsają perduoda užklausa tiesiai analizės paslaugų serveriui, kuriame ji vykdoma. Duomenų rinkinio gavimo užklausoje visada yra nurodyta, iš kokių duomenų bazės lentelių paimti duomenis duomenų gavybos algoritmui. Abu serveriai (duomenų bazės variklis ir analizės objektų serveris) tarpusavyje keičiasi duomenimis, pasinaudojant ODBC sąsaja. Šie duomenų apsikeitimai parodyti 3.10 paveiksle.



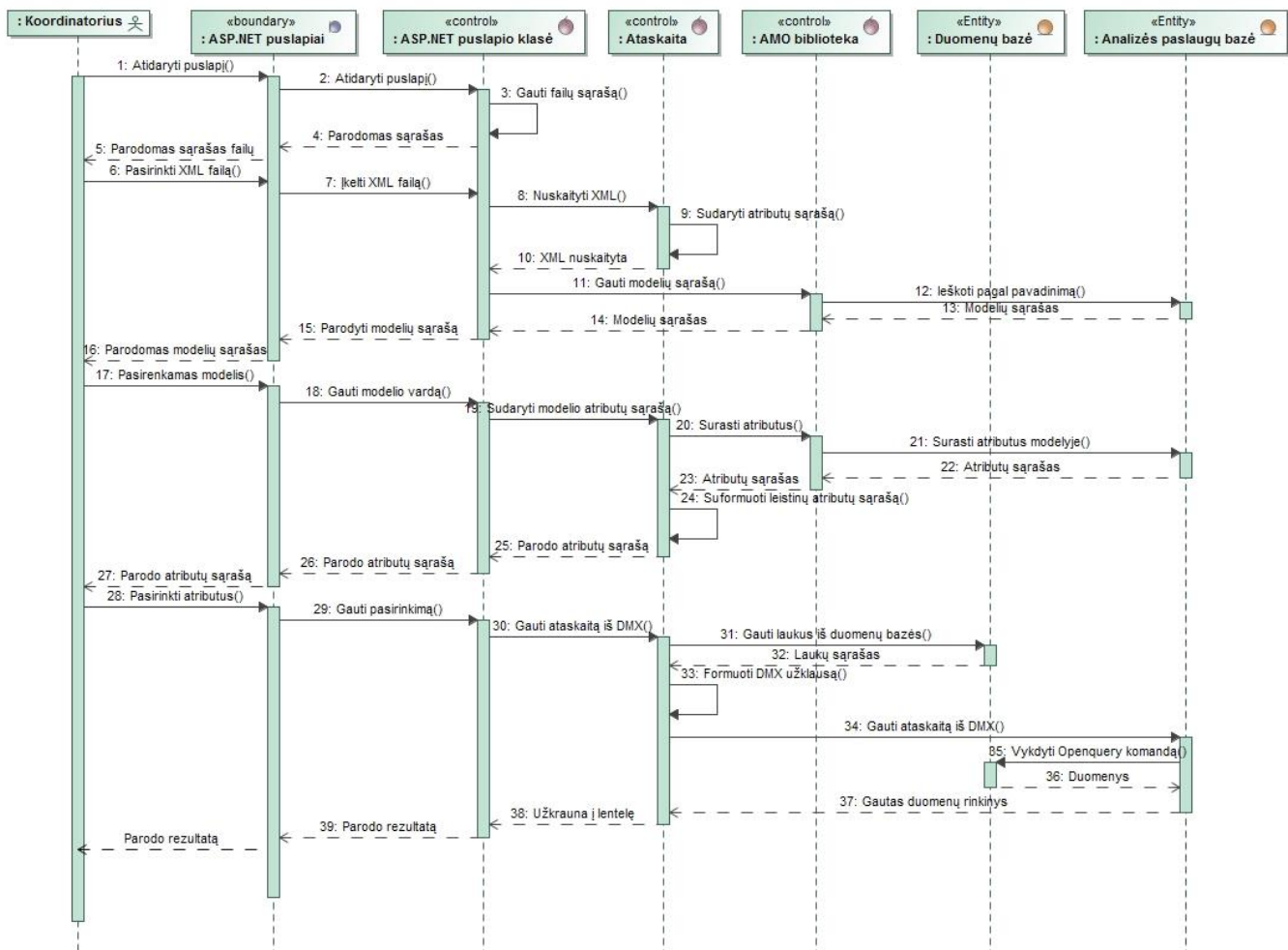
3.10 pav. Apsikeitimas duomenimis, vykdant DMX užklausa

Kadangi sprendime visi modelių atributai yra pasirenkami dinamiškai, reikia sudaryti dinaminę DMX užklausa. Pačią užklausa sudaro keletas dalių, kurias galima suskirstyti pagal tos dalies atliekamą funkcija. Kiekviena iš tų dalių yra suformuojama sprendime realizuotais metodais ir vėliau apjungiami į vieną užklausa. Tokios užklausa pavyzdys pateiktas toliau 3.11 paveiksle.



3.11 pav. Duomenų nuskaitymo iš duomenų gavybos modelio DMX užklausa

Šios DMX užklausa yra panaudotos gauti duomenis Senato studijų komisijos narių ataskaitoms. Jas kuriant pirmiausiai nuskaitymi duomenys iš XML failo. Toliau reikia pasirinkti, kokį duomenų bazės modelį naudosime gauti analizės rezultatus. Atlikus šiuos du veiksmus, reikia patikrinti, kokie atributai yra bendri XML faile ir duomenų gavybos modelyje. Juos suradus vartotojui yra pateikiamas jų sąrašas, kuriame reikia sužymėti varnelėmis tuos, kurie turi būti rodomi ataskaitoje. Po šių parametru nurodymo yra suformuojama anksčiau pavaizduota DMX užklausa. Atliekamas kreipimasis į analizės paslaugų duomenų bazę. Ši pasinaudoja duomenų bazėje esančiais duomenimis. Tokiu būdu gaunamas rinkinys duomenų, kuriuos galima atvaizduoti internetinio puslapio pagalba. Atliekamų veiksmų eiliškumas parodytas sekų diagrama, kuri parodyta 3.12 paveiksle.

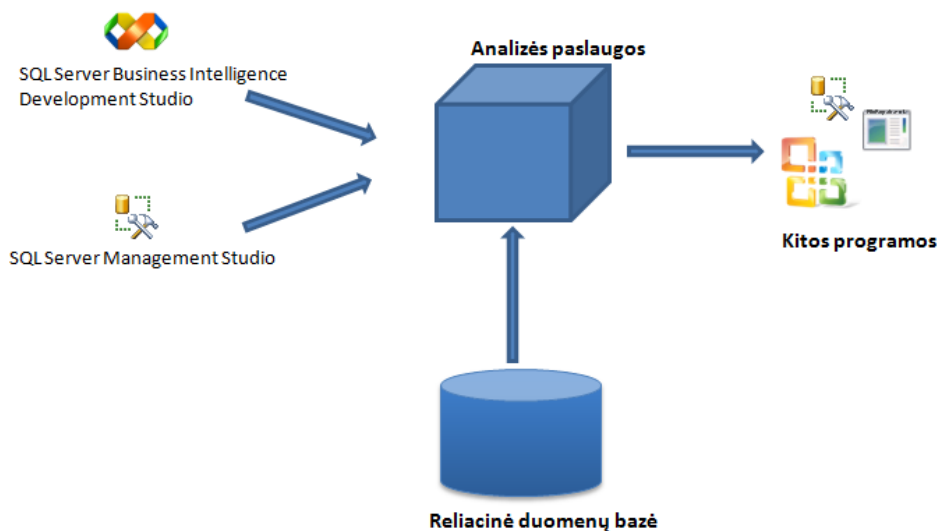


3.12 pav. Duomenų gavybos rezultatų gavimo sekų diagrama

### 3.2.5. Panaudotos duomenų bazės paslaugos

Kaip jau minėta anksčiau, sistemoje naudojama Microsoft SQL Server 2008 duomenų bazių valdymo sistema. Ši sistema turi paslaugas, kurios naudojamos užtikrinti nenutrūkstantį verslo analitikos duomenų analizę, integravimą ir saugojimą. Duomenų gavyba yra viena iš analizės paslaugų galimybių. Taip pat egzistuoja ataskaitų kūrimo ir integravimo paslaugos, kurios yra dalis Microsoft verslo analitikos platformos. Šiame darbe svarbiausios yra analizės paslaugos. Pasinaudojant SQL Server Management Studio arba Microsoft Business Intelligence Studio programomis, yra sukuriamas analizės paslaugų konceptualus modelis. Tai galima atlikti užklausų pagalba arba pasinaudojant grafinėmis kūrimo priemonėmis. Iš reliacinės duomenų bazės gaunami duomenys, kurie užpildo duomenų gavybos modelius ar OLAP kubus. Tuomet iš sukurtų struktūrų kitos taikomosios programos ar kiti Microsoft produktai pasiima analizės rezultatus[23]. Šis analizės paslaugų panaudojimas pavaizduotas 3.13 paveiksle.



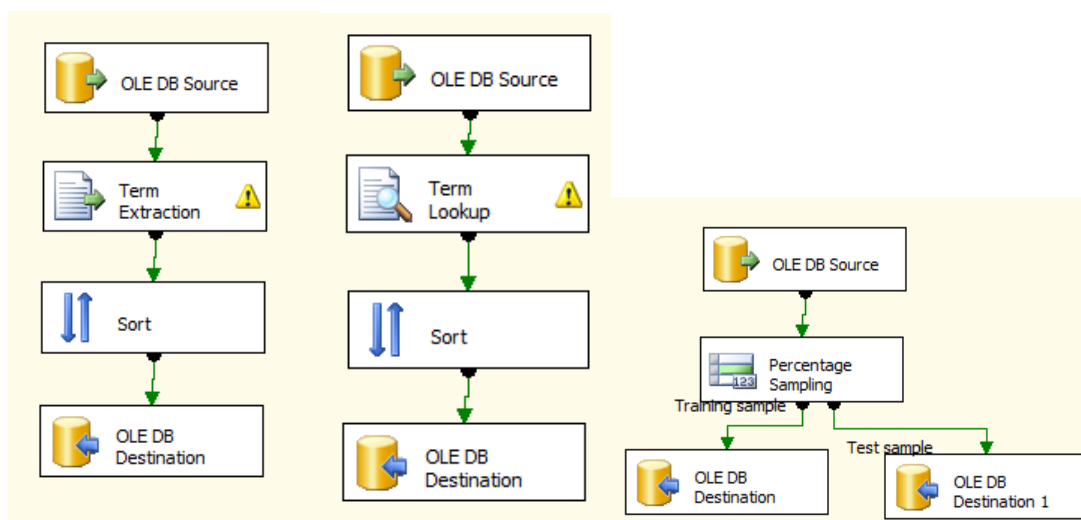


3.13 pav. Analizės paslaugų panaudojimas

### 3.2.6. Teksto gavybos atlikimas, pasinaudojant integravimo paslaugomis

Microsoft integravimo paslaugose kuriant duomenų srautus yra galimybė panaudoti tris elementus, kurie skirti teksto gavybai. Pirmoji yra suradimas pasikartojančių frazių, pagal kurias yra sudaromas žodynas, naudojamas duomenų gavyboje. Antroji funkcija skirta iš anksčiau gauto žodyno sudaryti sąrašą, kuriame matoma, kiek kartų kiekviena iš frazių pasikartoja. Trečioji funkcija padalina visus duomenis į dvi dalis: į skirtą apmokyti algoritmą bei skirtą testuoti, kaip tiksliai algoritmas atlieka prognozę.

Kaip naudojant tokią metodiką atlikti teksto analizę, yra pateikta pavyzdžiu su straipsnių duomenų baze, kuri panaudojama atliekant duomenų gavybą.[24] Šios analizės tikslas, surasti į kurią grupę patenka vienas ar kitas straipsnis. Pateiktoje medžiagoje yra siūloma susikurti tris paketus. Kiekviename iš paketų sukuriamas duomenų srautas, kuris atlieka jame nurodytus veiksmus. Kiekvienas iš paketų atlieka po vieną funkciją, aprašytą anksčiau. Toliau yra iliustracija, kurioje matoma, iš kokių elementų susideda duomenų srautai, kurie naudojami atlikti teksto gavybą. Ji pateikta 3.14 paveiksle.



3.14 pav. Duomenų srautai paketuose, kurie skirti vykdyti teksto gavybą

Šiomis anksčiau aprašytais integravimo paslaugų savybėmis atliekama Kauno technologijos universiteto anketų, kurias užpildė studentai, analizė. Anketos susideda iš klausimų, kuriuose reikia pasirinkti vieną iš pateiktų penkių įvertinimų. Pabaigoje pateikti yra trys atviri klausimai: kokie yra modulio teigiami aspektai, kokie yra modulio neigiami aspektai ir kaip siūlote tobulinti modulį. Būtent šie klausimai ir yra tiriami teksto analizės pagalba.

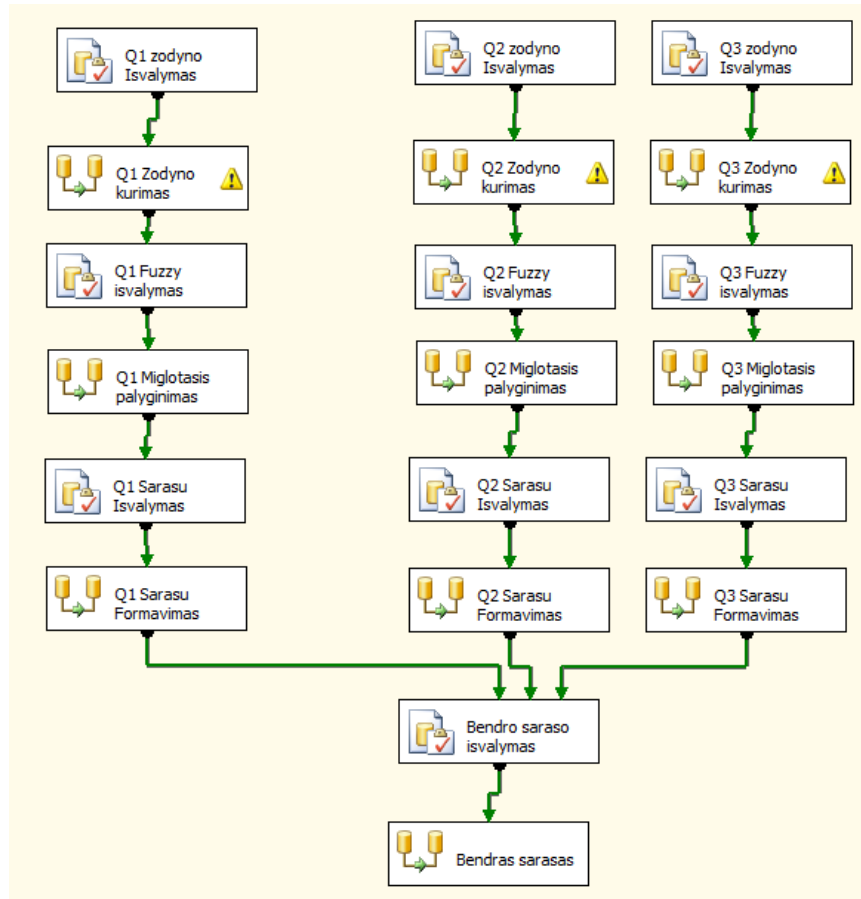
Žodyno frazių suradimas SQL Serveryje vykdomas tik anglų kalba.[25] Ieškomi daiktavardžiai arba daiktavardžių frazės, kuriose gali būti kartu su daiktavardžiu panaudoti būdvardžiai. Naudojant lietuvių kalbą, randami visi žodžiai, kadangi integravimo paslaugos neatpažįsta kalbos dalių. Gavus žodyno frazes reikia įvertinti, kaip – teigiamai ar neigiamai - surasta frazė įtakoja modulio kokybę. Norint tai atlikti su Kauno technologijos universiteto studentų anketomis, reikia dalykinės srities ekspertui gautą žodyną įvertinti.

Šie įvertinimai yra surašomi į lentelę su gautomis žodyno frazėmis. Gautos frazės gali įtakoti modulio įvertinimą teigiamai, neigiamai arba visai neįtakoti. Įtakojimas teigiamai rodo, kad šis dalykas gauna įvertinimą, kuris rodo tam tikrą dalyko teigiamą aspektą. Visai neįtakojama tuo atveju, jeigu gautas žodis ar jų rinkinys neparodo jokios įtakos modulio įvertinimui. Taip gali įvykti tais atvejais, kai gaunama, pavyzdžiui, tik įvardis ar kitos kalbos dalys, pagal kurias negalima surasti įtakos kokybei. Įvertinimas rašomas intervale nuo minus vieno iki plus vieno. Minus vienas reiškia, kad ši frazė stipriai įtakoja dalyką neigiamai, tuo tarpu plus vienas stipriai teigiamai. Pavyzdys su surašytais įvertinimais pateiktas 3.15 paveiksle.

	Tem	Reiksme
1	Abu	0
2	Ačiū	0,1
3	Ačiū Jums	0,1
4	Aiški atsiskaitymų sistema	1
5	aiškiai	0,2
6	Aiškiai	0,2
7	Aiškiai dėstomos paskaitos	1
8	aišku	0,2
9	aiškumas	0,2
10	aktuali	0,2
11	Aktuali	0,2
12	Aktuali informacija	1
13	aktualios temos	1
14	Aktualumas	0,2
15	Anglų	0,1

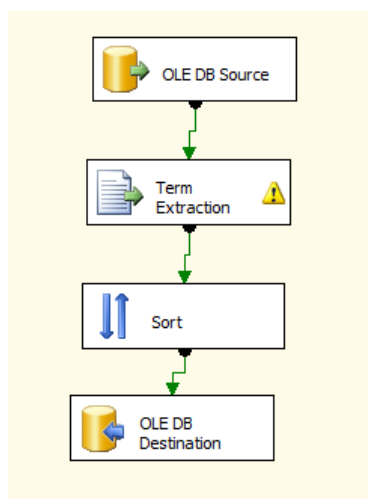
3.15 pav. Dalis įvertintų žodžių

Integravimo paslaugų projekte yra sukurtas vienas paketas, kuriame vykdomi visiems trimis klausimams tie patys veiksmai. Po to, kai suskaičiuojami kiekvieno klausimo modulio įvertinimai, jie yra susumuojami. Tokiu būdu gaunamas bendras modelio įvertinimas. Šis paketo vykdymo kontrolės srautas parodytas 3.16 paveiksle.



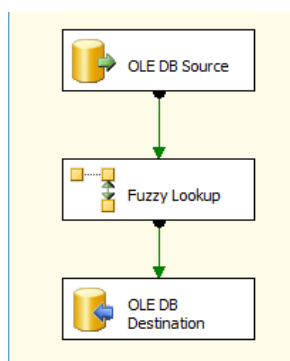
3.16 pav. Teksto gavybos integravimo projekto kontrolės srautas

Prieš kiekvieną duomenų srautą įvykdomos užklauskos, kurių dėka išvalomi rezultatai. Tokiu būdu duomenys nesusimaišo su anksčiau atliktos teksto gavybos rezultatais. Užduotyse matomi sutrumpinimai Q1, Q2 ir Q3 reiškia kiekvieną iš klausimų. Žodyno kūrimo duomenų srautas susideda iš funkcijos, kuri suranda žodžius naujai duotose anketose. Yra laikoma, kad frazė turi kartotis bent du kartus, o didžiausias maksimalus ilgis gali būti dešimt žodžių. Suradus frazes ar žodžius, yra atliekamas rikiavimas pagal įvertinimą bei gauti rezultatai įrašomi į duomenų bazę. Toks srautas yra sukurtas kiekvienam iš trijų anketos klausimų. Iš kokių elementų susideda šis srautas pavaizduota 3.17 paveiksle.



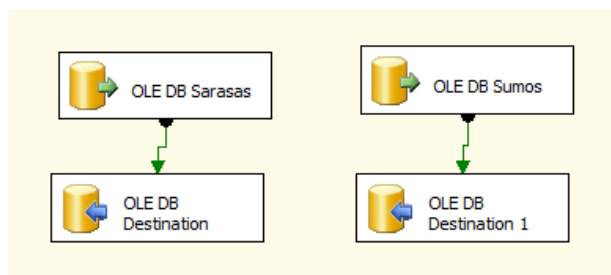
3.17 pav. Žodyno sukūrimo duomenų srautas

Sekantis žingsnis yra sulyginti eksperto sudarytą įvertinimų žodyną su prieš tai aprašyto duomenų srauto surastu sąrašu. Sulyginimui reikia nuskaityti anketų tekstus ir lyginti su eksperto įvertintu sąrašu. Paimamos ne vien visiškai tikslios frazės, bet ir panašios. Surastos frazės įrašomos į duomenų bazę. Visi elementai esantys šiame duomenų sraute matomi 3.18 paveiksle.



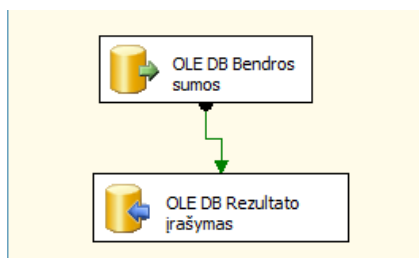
**3.18 pav.** Suradimas, kurie žodžiai egzistuoja iš eksperto įvertinto frazių sąrašo

Sekančiame duomenų sraute yra apjungiamos surastos frazės, kurios yra pateiktose anketose, su eksperto surašytais frazėmis. Tokiu būdu surandami visi žodžiai ar frazės, kurie turi įvertinimus. Tai atliekama, paimant duomenis užklausos pagalba iš lentelių ir perkeltiant į kitą lentelę. Taip pat tuo pačiu metu lygiagrečiai yra suskaičiuojama suma įvertinimų vienam moduliui. Šie įvertinimai - tai suma vieno iš trijų klausimų. Šis duomenų srautas matomas 3.19 paveiksle.



**3.19 pav.** Vieno klausimo įvertinimų radimo moduliams duomenų srautas

Galiausiai reikia sudėti kiekvieno klausimo gautus įvertinimus ir suskaičiuoti vieną įvertinimą kiekvienam moduliui. Tai atliekama SQL užklausos pagalba, kuri vykdoma duomenų nuskaitymo metu. Jos rezultatus reikia įrašyti į duomenų bazę. Tokiu būdu gauname galutinį įvertinimų sąrašą. Galutinio sąrašo gavimo duomenų srautas parodytas 3.20 paveiksle.



**3.20 pav.** Galutinio įvertinimų sąrašo sudarymo duomenų srautas

Dėl lietuvių kalbos nepalaikymo yra realizuota ir alternatyva atlikti teksto gavybą. Jis pateiktas sekančiame skyrelyje.

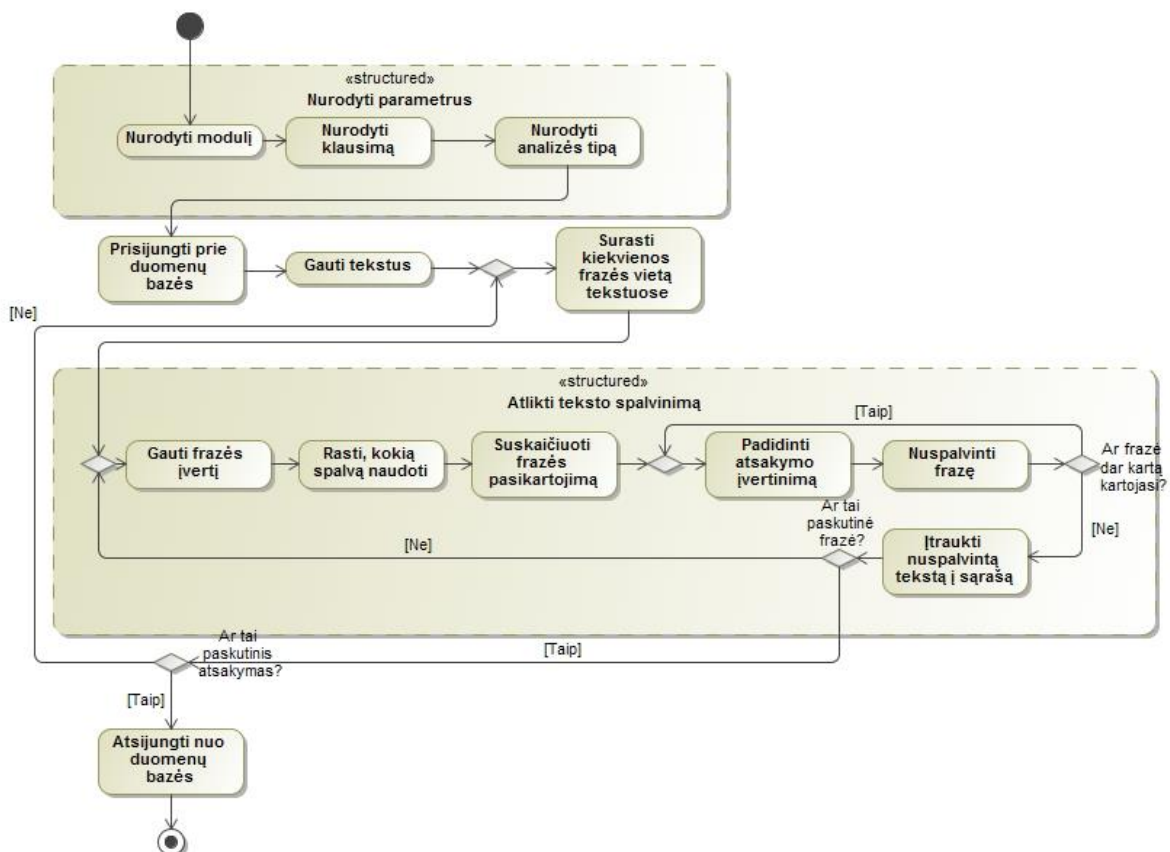
### 3.2.7. Teksto gavyba tiesiogiai ieškant frazių tekste

Universiteto studijų modulių kokybės įvertinimą pagal pateiktus atsakymus į anketos klausimus galima atlikti ir nepasinaudojant integravimo paslaugomis. Turint eksperto įvertintą žodyną galima atlikti kiekvienos frazės paiešką atsakymo tekste. Jei tekstas sutampa, moduliui priskiriame tokį įvertinimą, kuris yra nurodytas prie atitinkamos frazės. Taip atliekame su visomis frazėmis ir galiausiai susumuojame to atsakymo į klausimą bendrą įvertį.

Šis būdas yra realizuotas, pasinaudojant sistemos funkcija, kuri parodo, kuriose teksto vietose yra visos žodyne esančios frazės. Naudojant šią realizacijos dalį, galima pamatyti, kaip tiksliai išsidėsto frazės. Tokiu būdu tampa akivaizdu, kaip teisingai veikia metodika, kuri atlieka teksto gavybą. tačiau šią funkciją galima naudoti ne tik metodikos tikrinimui, bet ir pačiai gavybai.

Pirmiausia reikia pasirinkti tam tikrą vieną studijų modulį, kurio įvertinimą norime pamatyti. Antrasis žingsnis yra pasirinkti, kurio klausimo tekstus norime matyti. Kadangi yra realizuota daugiau nei vienas būdas analizuoti tekstą laukelyje „Pasirinkite, ką rodyti“, reikia pasirinkti analizės tipą. Tai atlikus gaunamas rezultatas – atsakymų tekstai, kuriuose spalvotai nuspalvintos frazės iš žodyno. Šalia kiekvieno nuspalvinto teksto yra pateikta būtent to teksto įvertinimas, kuris gautas sudėjus visų rastų frazių įvertinimus. Po kiekvienu tekstu yra pateikta visų surastų frazių jame sąrašas.

Visas algoritmas, pagal kurį ieškomi studijų modulių įvertinimai, yra pateiktas pasinaudojant veiklos diagrama 3.21 paveiksle.



3.21 pav. Frazių spalvinimo atsakymų tekste algoritmas

Norint sėkmingai taikyti vieną iš paminėtų būdų atlikti teksto gavybą, reikia turėti teiginių vertinimo skalę. Šios skalės sudarymo principas paaiškintas sekančiame skyrelyje.

### 3.3. Frazijų įvertinimo rekomendacijos

Į informacinę sistemą įrašomi žodžiai arba kelių žodžių junginiai, kuriuos toliau vadinsime frazėmis. Pagal frazės reikšmę, kokią įtaką ji turi nagrinėjamui studijų dalykui, ji yra įvertinama arba teigiamai arba neigiamai. Jei jokios įtakos frazė nedaro, įvertinimas yra nulis. Įvertinimas svyruoja nuo minus vieno iki plus vieno. Galimos naudoti reikšmės yra tokios, kokios matomos 3.1 lentelėje.

3.1 lentelė. Galimi frazių įvertinimai

Įvertinimas	Reikšmė
-1	Frazė daro akivaizdžią neigiamą įtaką nagrinėjamui dalykui
-0,5	Frazė daro neigiamą įtaką
-0,2	Pagal tai, kuris klausimas atsakinėjamas, galima spėti, kad labai tikėtina jog veikia neigiamai
-0,1	Pagal tai, kuris klausimas atsakinėjamas, galima spėti, kad labiau tikėtina jog veikia neigiamai
0	Frazė nedaro jokios įtakos nagrinėjamui dalykui arba neaiški frazės įtaka pagal jos kontekstą
0,1	Pagal tai, kuris klausimas atsakinėjamas, galima spėti, kad labiau tikėtina jog veikia teigiamai
0,2	Pagal tai, kuris klausimas atsakinėjamas, galima spėti, kad labai tikėtina jog veikia teigiamai
0,5	Frazė daro teigiamą įtaką
1	Frazė daro akivaizdžią teigiamą įtaką

Pasakymas „Neigiama įtaka“ reiškia, kad tam tikro dėstytojo skaitomas dalykas yra vertinamas neigiamai dėl tam tikro aspekto ir turi būti siekiama šį trūkumą pašalinti. Surinkus labai žemą įvertinimą, gali būti svarstoma dalyką pašalinti iš studijų programos ar imtis kitų priemonių.

Tuo tarpu pasakymas „Teigiama įtaka“ parodo, kad dalyke yra tam tikras teigiamas aspektas, kuris yra pastebėtas studentų. Surinkus aukštą teigiamą įvertinimą, galima svarstyti apie dėstytojo paskatinimą ar kitus veiksmus.

Nulinis įvertinimas labai dažnai yra gaunamas, kai yra naudojama frazė sudaryta iš vieno žodžio. Visi kiti įvertinimai gaunami pagal tai, kokios kalbos dalys yra naudojamos frazėje ir pagal frazės kontekstą. Labai dažnas atvejis tokių kalbos dalių yra daiktavardžio naudojimas su būdvardžiu arbarieveiksmiu. Tokio atvejo pavyzdžiai gali būti tokie, kaip informatyvios skaidrės, konkrečiai dėstomos paskaitos. Kai kurios kalbos dalys niekada neduoda jokio įvertinimo. Vienas iš tokių variantų yra tik daiktavardžio panaudojimas be jokių kitų kalbos dalių. Tinkamas pavyzdys yra frazė „Dėstytojas“.

Žinant, kad įvertinimas yra teigiamas arba neigiamas, reikia parinkti vieną iš kelių galimų pasirinkimų. Tai atliekama pagal frazės kontekstą. Kai kurios iš tokių frazių nevisai tiksliai nusako situaciją. Pavyzdžiui frazė „Įdomu“ nepasako, kas tiksliai dalyke yra įdomu. Kadangi toks tekstas yra klausime apie teigiamus aspektus, galime suprasti, kad jis teigiamai lemia įvertinimą. Tačiau dėl ne visai aiškaus konteksto jam duodamas pusės balo įvertinimas.

Jeigu turime tokią frazę, kurios kontekstas yra visai neaiškus, bet naudoti priešingai, atrodo, nelabai logiška, skiriamas vienos arba dviejų dešimtųjų įvertinimas. Tokios frazės pavyzdys yra „Ačiū“. Akivaizdu, jog labai mažai tikėtina, kad dėkojama būtų klausime „Nurodykite teigiamus aspektus“, turint omenyje neigiamą mintį. Taip pat mažiausias teigiamas įvertinimas skiriamas, kai kyla labai didelė abejonė, ar tai veikia teigiamai, ar visai neveikia įvertinimo.

Klausime „Ką siūlytumėte pakeisti dalyke“ radus frazę „Pakeisti dėstytoją“, kuri kartojasi daug kartų, įvertiname ją minus vienetu, nes tokia frazė parodo tai, kad dėstytojas turi kažkokių trūkumų dėstyje, kuriuos pastebi studentai. Pilnas įvertinimas duodamas, kai teksto turimas kontekstas labai stipriai įtakoja dalyko įvertinimą.

### 3.4. Tyrimų atlikimas, pasinaudojant sukurta sistema

Sukurtame sprendime svarbiausi ir labiausiai sudėtingi veiksmai yra duomenų gavybos struktūros ir modelio sukūrimas bei ataskaitos suformavimas. Pastarasis veiksmas parodo duomenų gavybos rezultatus. Toliau bus aprašyta, kaip šiuos veiksmus atlikti pasinaudojant sistema.

#### 3.4.1. Duomenų gavybos modelių ir struktūrų kūrimas, naudojantis sprendimu

Viena iš kuriamo sprendimo savybių, nurodyta reikalavimų analizėje, yra galimybė sukurti duomenų gavybos objektus, naudojantis pačia sistema. Tokiu būdu vartotojas galės keisti tik tokius nustatymus, kurie yra numatyti kūrėjo. Kūrimas vyksta keliais žingsniais, kurių metu pateiktose formose reikia užpildyti prašomus duomenis. Šios formos yra ASP.NET karkaso pagalba ir C# programavimo kalba sukurtuose internetiniuose puslapiuose.

Duomenų gavybos struktūros ir modelio sukūrimas atliekamas, pasirinkus vartotojo meniu pasirinkimą Sukurti modelį. Pirmajame kūrimo žingsnyje pasirodžiusioje formoje reikia nurodyti, kokios rūšies analizę norime vykdyti su kuriu duomenų gavybos modeliu. Vienas iš pavyzdžių gali būti studentų rezultatai pagal stojimo ar universitete gautus įvertinimus. Formoje taip pat reikia įrašyti norimą struktūros ir modelio pavadinimą. Ši forma su užpildytais pavyzdiniais duomenimis matoma 3.22 paveiksle.

DUOMENŲ GAVYBOS TAIKYMAS Sveiki Vardenis [Atsijungti](#)

STUDIŲ PROCESĖ

Pirmasis puslapis Įkelti XML bylą Teksto gavyba Sukurti modelį

Šioje formoje pirmiausiai nurodykite kokio tipo analizę norite vykdyti. Tai nurodžius į atitinkamus laukus užrašykite kokiu vardu norite pavadinti duomenų gavybos struktūrą ir kokiu vardu duomenų gavybos modelį. Jei tokie vardai jau bus parinkti, Jums bus parodytas klaidos pranešimas.

**DUOMENŲ GAVYBOS MODELIO SUKŪRIMAS**

Kokią analizę vykdyti: Studento rezultatai pagal stojimo balus

Struktūros pavadinimas: stojimostруктура Modelio pavadinimas: stojimomodelis

Toliau

3.22 pav. Analizės tipo bei modelio ir struktūros pavadinimų nurodymas

Paspaudus mygtuką „Toliau“, pasirodo antroji forma. Šioje formoje galima sukurti norimą duomenų gavybos struktūrą pagal tam tikrų lentelių laukus. Jį pasirinkus reikia nurodyti, kokiu vardu

norime pavadinti lauką. Taip pat reikia pasirinkti, ar jis yra prognozuojamas, ar lemiantis prognozę. Nurodžius prašomą informaciją, spaudžiamas mygtukas „Pridėti lauką“. Struktūros pirminis laukas pridedamas automatiškai pagal tai, koks analizės tipas bus vykdomas. Jei tai yra analizė, kuri tiria universiteto įvertinimus, papildomai formoje galima pasirinkti, kokio modulio įvertinimą prognozuosime. Kai yra sudedami visi norimi laukai, spaudžiamas mygtukas „Patvirtinti struktūrą“. Tai atlikus galima paspausti mygtuką „Sukurti struktūrą“. Kaip atrodo vaizdas pridėjus keletą laukų, parodyta 3.23 paveiksle.

**DUOMENŲ GAVYBOS TAIKYMAS**
Sveiki Vardenis [Atsijungti](#)

**STUDIJŲ PROCESĖ**

Pirmasis puslapis
Įkelti XML bylą
Teksto gavyba
Sukurti modelį

Ši forma skirta pasirinkti iš kokių atributų susidės duomenų gavybos struktūra ir modelis. Pirmiausiai turite pasirinkti vieną iš leidžiamų lentelių. Pasirinkus lentelę lauke "Pasirinkti lauką" Jums bus galima rinktis vieną iš atributų. Nurodykite atributo vardą, kuris bus matomas ataskaitose. Toliau reikia nuspresti ar šis laukas bus prognozuojamas ar prognozuojantis. Galiausiai pasirinkite lauko diskretizavimą. Kai surašysite visą lauko informaciją paspauskite "Pridėti lauką" mygtuką. Laukų galima pridėti tiek kiek pageidaujama. Bent vienas laukas turi būti prognozuojamas. Kai baigsite rinktis laukus paspauskite "Patvirtinti struktūrą" ir "Sukurti struktūrą" mygtukus.

### DUOMENŲ GAVYBOS MODELIO SUKŪRIMAS

Pasirinkti lentelę:

Pasirinkti lauką:

Būsimo pavadinimas:

Bus prognozuojamas:  Bus lemiantis:  Bus tik prognozuojamas:

Diskretizavimas:

Lentelė	Laukas	LaukoTipas	Pavadinimas modelyje	Diskretizavimas	Spejamas	Lemiantis	TikSpejamas
Studentas	VIDKO	text	VIDKO	Key	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Studentas	LYTIS	int	lytis	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Studentas	FINANS	nvarchar	finansavimas	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Studentas	VIDISL_MOK_KODAS	decimal	mokykloskodas	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Studentas	PAGEIDAVIMO_NUMERIS	decimal	pageidavimo_nr	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Studentas	SAVIVALDYBE	nvarchar	savivaldybe	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Studentas	BUSENA	nvarchar	busena	Discrete	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

**3.23 pav.** Sistemos vaizdas pridėjus laukų į struktūrą

Paskutiniame žingsnyje reikia nurodyti, koks bus mūsų naudojamos analizės tipas. Šie tipai yra parinkti pagal algoritmų, kurie egzistuoja SQL Serverio analizės paslaugose, paskirtį. Pasirinkus tipą parodomi visi algoritmai, kurie gali būti naudojami pagal padarytą pasirinkimą. Atlikus pasirinkimą paspaudžiamas mygtukas Patvirtinti. Tada sukuriamas modelis bei atliekamas jo apmokymas. Vaizdas, kuris matomas tai atlikus, pavaizduotas 3.24 paveiksle.



## STUDIJŲ PROCESE

Pirmasis puslapis

Įkelti XML bylą

Teksto gavyba

Sukurti modelį

Ši forma skirta pasirinkti kokiu algoritmu norėsite atlikti duomenų gavybą. Pirmiausiai pasirinkite analizės tipą. Jį pasirinkus "Pasirinkti algoritmą" sąrašė bus rodoma visi šiam analizės tipui leidžiami algoritmai. Pasirinkus paspauskite mygtuką "Patvirtinti". Tada bus pradėtas duomenų gavybos modelio apmokymas.

## DUOMENŲ GAVYBOS MODELIO SUKŪRIMAS

Pasirinkti analizės tipą: Sugrupavimas pagal kriterijus

Pasirinkti algoritmą: Grupavimo algoritmas

Patvirtinti

Modelis sukurtas ir apdorotas sėkmingai

## 3.24 pav. Pasirinkimas algoritmo, kuriuo kuriamas modelis

Norint įsitikinti, kad modelis ir struktūra sukurti sėkmingai, juos galime peržiūrėti, pasinaudojant Business Intelligence studijos pagalba. Šiuo atveju be formoje nurodytų laukų dar yra sukuriami dalykų laukai, kurie skirti saugoti įvertinimus. Dalis sąrašo matoma 3.25 paveiksle. Galime pastebėti, kad visi anksčiau pasirinkti laukai čia yra.

Structure	stojimodelis
80040	Input
80229	Input
80240	Input
80270	Input
80273	Input
90090	Input
90155	Input
90195	Input
90225	Input
90245	Input
90250	Input
90265	Input
90335	Input
90375	Input
90400	Input
90645	Input
90655	Input
90670	Input
90745	Input
90755	Input
90768	Input
90770	Input
90810	Input
90925	Input
90930	Input
90975	Input
91645	Input
busena	PredictOnly
finansavimas	Input
lytis	Input
mokykloskodas	Input
pageidavimo_nr	Input
savivaldybe	Input
VIDKO	Key

3.25 pav. Laukai, kurie buvo kuriami anksčiau formose

Sukurtas ir apdorotas gavybos modelis yra naudojamas prognozei, kurios rezultatas yra atvaizduojamas ataskaitoje Senato studijų komisijos nariams. Apie šios ataskaitos sukūrimą aprašyta sekančiame skyrelyje.

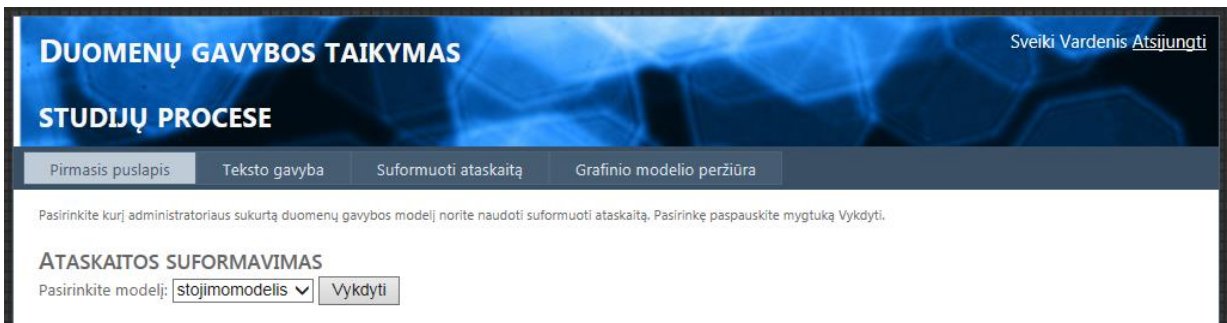
### 3.4.2. Duomenų gavybos gautų modelių rezultatų peržiūra naudojantis sprendimu

Kai yra sukurtas bent vienas duomenų gavybos modelis, sistemoje galima suformuoti ataskaitą. Prognozei reikia turėti duomenų rinkinį, kurį reikia iširti. Jie perduodami į sistemą XML bylos pagalba. Jas įkelia sistemos administratorius. Tik, kai šis veiksmas yra atliktas, koordinatorius gali formuoti ataskaitą.

Pirmajame formavimo žingsnyje reikia pasirinkti anksčiau paminėtą bylą. Rodomos visos anksčiau įkeltos bylos. Taip pat reikia pasirinkti, kokio tipo tai bus failas – ar su priėmimo, ar su studijų, ar pagal modulius, ar pagal dėstytojus. Paspaudus mygtuką „Vykdėti“, duomenys yra paaimami iš bylos ir perkeliami į duomenų bazę. Vaizdas, kaip atrodo ši forma interneto naršyklėje, yra pavaizduotas 3.26 paveiksle.

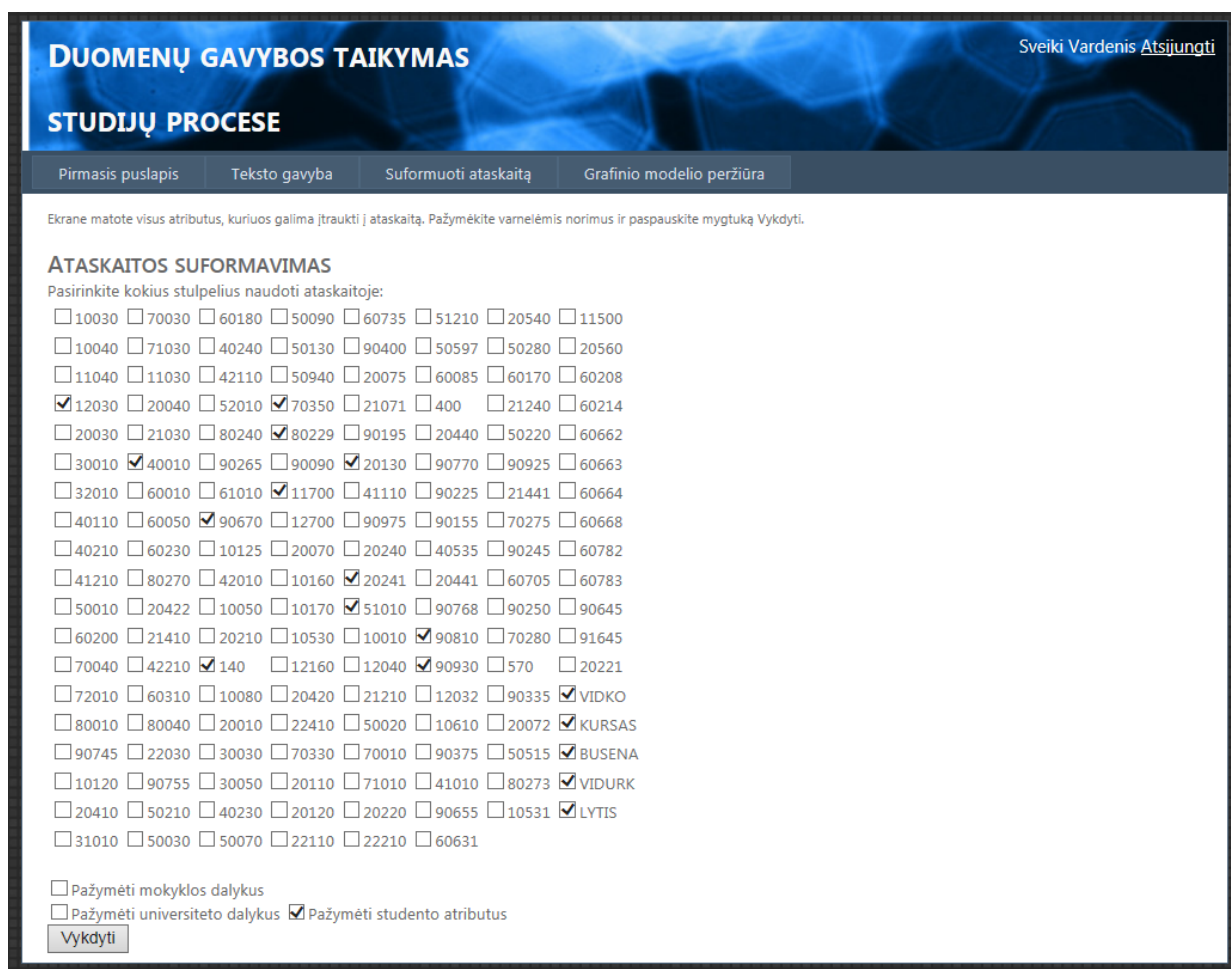
3.26 pav. XML bylos pasirinkimo formos vaizdas naršyklėje

Antrame žingsnyje reikia pasirinkti, kokį duomenų gavybos modelį panaudosime, norėdami tirti universiteto studentų duomenis. Visi formos pasirinkimo sąrašė matomi modeliai yra patalpinti SQL Server duomenų bazių valdymo sistemos analizės serveryje. Į jį jie įtraukiami, pasinaudojant anksčiau aprašyta funkcija sistemoje. Pasirinkus formoje, kuri matoma interneto naršyklėje, vieną iš pasirinkimų, reikia paspausti mygtuką „Vykdėti“. Vaizdas, kaip atrodo šis ataskaitos formavimo žingsnis, parodytas 3.27 paveiksle.



3.27 pav. Duomenų gavybos modelio pasirinkimo formos vaizdas naršyklėje

Toliau sistema palygina atributus tiriamuosiuose duomenyse su atributais, kurie yra duomenų gavybos modeliuose. Tai atlikus gaunamas atributų sąrašas, kuris yra leidžiamas rinktis vartotojui. Visi šie atributai yra parodomi interneto naršyklėje formos pagalba. Kiekvienas atributas yra laukas, kurį galima pažymėti varnele. Tie atributai, kurie yra pažymėti varnele, yra įtraukiami į formuojamą ataskaitą Senato studijų komisijos nariams. Atlikus šį pasirinkimą, reikia paspausti mygtuką „Vykdyti“. Vaizdas, kaip atrodo ši forma internetinėje naršyklėje, parodytas 3.28 paveiksle.



3.28 pav. Atributų pasirinkimo formos vaizdas internetinėje naršyklėje

Interneto naršyklėje parodoma lentelė su analizės rezultatais. Juose matomi visi atributai su reikšmėmis, kurie buvo nurodyti ankstesniame žingsnyje. Norint pateikti ataskaitą Senato studijų komisijos nariams, reikia į apačioje esantį lauką „Ataskaitos pavadinimas“ įrašyti pavadinimą, kuriuo norime išsaugoti ataskaitą ir paspausti mygtuką „Išsaugoti“. Vaizdas, kaip atrodo ataskaita interneto naršyklėje, parodytas 3.29 paveiksle.

## STUDIJŲ PROCESE

Pirmasis puslapis

Teksto gavyba

Suformuoti ataskaitą

Grafinio modelio peržiūra

Ekrane matote pateiktą ataskaitą. Jei norite šią ataskaitą išsaugoti tam, kad ją galėtumėte peržiūrėti senato nariai įrašykite į laukelį Ataskaitos pavadinimas norimą pavadinimą bei paspauskite mygtuką Išsaugoti.

## ATASKAITOS SUFORMAVIMAS

12030	40010	90670	140	70350	80229	11700	20130	20241	51010	90810	90930	VIDKO	KURSAS	VIDURK	LYTIS	120301	30010	40110	40210	4
24												10505	2	6	1	24	8	6	7	5
7												10652	1	5	1	7	7	4	4	
4												10812	1	5	1		5	5	4	
10		8	10	6								14658	2	6	2			10	8	
6					5							15306	2	6	2		6	7		
9												15429	2	6	2		10	9	10	
10												15433	2	10	2		10	10	9	
												15451	2	8	2		10	10	9	
10												15698	2	7	2					
8												15703	2	6	2		7		8	
7					5							15729	2	5	2		10	6	7	
7						7						15904	2	6	2		8	10	10	

3.29 pav. Analizės rezultatų lentelė pateikta interneto naršyklės pagalba

Visi duomenys, kurie naudojami apmokyti duomenų gavybos algoritmus, yra saugojami duomenų bazėje. Jos schema bei aprašymas yra pateiktas sekančiame skyrelyje.

### 3.5. Tyrimams naudojamos duomenų bazės schema

Kaip ir kitos informacinės sistemos, šis sprendimas naudoja duomenų bazę. Joje saugoma informacija apie studentus. Informacija yra reikalinga atlikti studentų arba apklausų, kurias jie atsakė, analizei. Analizei naudojama tokia informacija, kaip egzaminų ir dalykų įvertinimai mokykloje, dalykų joje sąrašai. Taip pat informacija apie studentą universitete – dalykai, kuriuos mokėsi universitete, jų vertinimai, kas tuos dalykus dėstė, kokiam padaliniui dėstytojas priklauso ir panaši informacija. Kas saugoma duomenų bazės lentelėse, pateikta tolesnėse lentelėse. Pirmoji lentelė Apklausų, kurioje saugojama studentų modulių įvertinimo anketos rezultatai. Šios lentelės atributai parodyti 3.2 lentelėje.

3.2 lentelė. Apklausų lentelė

Lauko pavadinimas	Paskirtis
Q1	Apklauso atsakymai į atviro tipo klausimus, atsakytus studentų
Q2	
Q3	
Q4	
Lytis	Studento, kuris atsakinėjo į anketą, lytis
Amžius	Studento, kuris atsakinėjo į anketą, amžius
Fakultetas	Studento, kuris atsakinėjo į anketą, fakultetas, kuriame jis mokosi
Studiju_pakopa	Studento, atsakiusio į anketą, studijų pakopa
Modulio_kodas	Modulio, kuris buvo vertinamas, kodas
Dest_pavarde	Dėstytojo, kuris buvo vertinamas, pavardė
Kursas	Kuriame kurse mokėsi studentas, kuris vertino modulį
id	Lentelės pirminis raktas

Informacijai apie dalykus, kuriuos mokėsi mokykloje ir kurie yra abiturientų atestatuose, saugoti skirta lentelė Atestato\_dalykai. Jos laukų paskirtis aprašyta 3.3 lentelėje.

**3.3 lentelė.** Atestato\_dalykai lentelė

Lauko pavadinimas	Paskirtis
Kodas	Atestato dalyko kodas pirminis raktas
Pavadinimas	Atestato dalyko pavadinimas
Tipas	Nurodo, ar tas atestato dalykas yra egzaminas, ar jis vertinimas metiniu pažymiu
Lama_uzskaitymas	Nurodo kaip kokį modulį LAMABPO užskaito šį atestato dalyką

Visi anksčiau paminėti dalykai turi egzamino arba metinius įvertinimus. Šie įvertinimai saugomi lentelėje Stojimo\_balai, kurios laukai aprašyti 3.4 lentelėje.

**3.4 lentelė.** Stojimo\_balai lentelė

Lauko pavadinimas	Paskirtis
Vidko	Studento, kurio yra šis įvertinimas, vidinis kodas universitete
Kodas	Atestato dalyko kodas, kuris parodo koks tai dalykas
Pazymys	Atestato dalyko įvertinimas, kurį gavo studentas
Lygis	Kokiu lygiu buvo besimokamas šis dalykas

Studentas besimokydamas universitete gauna įvairių studijuojamų dalykų įvertinimus. Šie studijų dalykai saugojami lentelėje Dalykas, kuri aprašyta 3.5 lentelėje.

**3.5 lentelė.** Dalykas lentelė

Lauko pavadinimas	Paskirtis
Modulis	Dalyko kodas universitete
Mod_pavad	Modulio pavadinimas
Tab_nr	Dėstytojo, kuris atsakingas už modulį, tabelio numeris
Pad_kod	Padalinio, kuris atsakingas už modulį, kodas
Mod_apavad	Modulio angliškas pavadinimas

Visus anksčiau paminėtus studijų dalykus kuruoja vienas iš universiteto darbuotojų. Šio darbuotojo informacija saugojama lentelėje Destytojas, kurios laukai aprašyti 3.6 lentelėje.

**3.6 lentelė.** Destytojas lentelė

Lauko pavadinimas	Paskirtis
Tab_nr	Dėstytojo tabelio numeris universitete
Pavarde	Dėstytojo vardas ir pavardė
Lytis	Dėstytojo lytis
Pad_kod	Padalinio, kuriam priklauso dėstytojas, kodas
Par_kod	Dėstytojo pareigybės kodas
M_L	Dėstytojo mokslinio laipsnio kodas
Pare_poz	Dėstytojo pareigų požymis, parodantis, ar tos pareigos pagrindinės
Etat_koef	Etato, kuriuo dirba dėstytojas, koeficientas

Kiekvienas studentas turi sudarytą savo individualų planą, kuris susideda iš dalykų. Kiekvieno dalyko, esančio individualiame plane, informacija saugoma lentelėje Individualus\_planas. Jos struktūra aprašyta 3.7 lentelėje.

3.7 lentelė. Individualus\_planas lentelė

Lauko pavadinimas	Paskirtis
Vidko	Studento, kurio yra informacija apie individualų planą, vidinis kodas universitete
Grupe	Studento grupė, kurioje mokėsi studentas studijuodamas modulį
Skyrpt	Studento studijų forma, kuria buvo studijuotas modulis
Fakul	Studento fakultetas, kuriame studijavo besimokydamas studentas
Status	Nurodo, kokios pakopos studijos buvo studijuojant šį modulį
Semes	Kuriame semestre buvo studijuojamas šis modulis
Modulis	Kokio modulio šis individualus planas
Forma	Kokia forma dėstomas studentui modulis
Kalba	Kokia kalba dėstomas modulis studentui
Mod_poz	Kokio tipo tai buvo modulis studentui (privalomas, pasirenkamas ir pan.)
Kred_sk	Kiek kreditų turėjo šis modulis studentui
Pazsemes	Kaip buvo įvertintas studento semestro darbas
Pazymys	Kokį įvertinimą gavo studentas šiame modulyje
I_data	Kada buvo įrašytas paskutinis įrašas apie šį modulį
Tab_nr	Koks dėstytojas įrašinėjo informaciją apie šį modulį
Pad_kod	Kokio padalinio dėstytojas įrašinėjo informaciją apie šį modulį
Mok_mod	Ar studentas turi mokėti už šį modulį
Ind_id	Individualaus plano unikalus pirminis numeris

Studentas, kuris mokosi universitete, turi daugybę informacijos apie save. Visa ši informacija surašoma į lentelę, kurios aprašymas matomas 3.8 lentelėje.

3.8 lentelė. Studentas lentelė

Lauko pavadinimas	Paskirtis
Metai	Metai, kuriais sukurtas šis studento duomenų egzempliorius
Asm_kodas	Studento asmens kodas
Vidko	Studento vidinis kodas universitete
Pavarde	Studento pavardė
Vardas	Studento vardas
Kursas	Kursas, kuriame mokosi studentas
Busena	Ar studentas po šių metų liko universitete, ar iškrito
Ist_met	Kuriais metais studentas įstojo į universitetą
Fakul	Kuriame fakultete studentas mokosi
Vidurk	Kokį vidurkį turi studentas
Viso_kr	Kiek kreditų jau turi studentas
Bendr	Ar studentas gyvena bendrabutyje
Lytis	Studento lytis
Gmetai	Studento gimimo metai
Sesijar	Koks pažangumas buvo paskutinės rudens sesijos
Sesijap	Koks pažangumas buvo paskutinės pavasario sesijos
Vidurkr	Koks buvo vidurkis paskutinio rudens semestro
Vidurkp	Koks buvo vidurkis paskutinio pavasario semestro
Finans	Ar studentas yra finansuojamas valstybės
Kambarys	Bendrabučio kambarys, kuriame gyvena studentas
Pilietybe	Studento pilietybė
Vieta	Kokioje vietoje studijuoja studentas (klausytojo, valstybinė ir pan.)
Valstybe	Valstybė, kurioje gyvena studentas
Miestas	Miestas, kuriame gyvena studentas
Gyvenv	Gyvenvietė, kurioje gyvena studentas

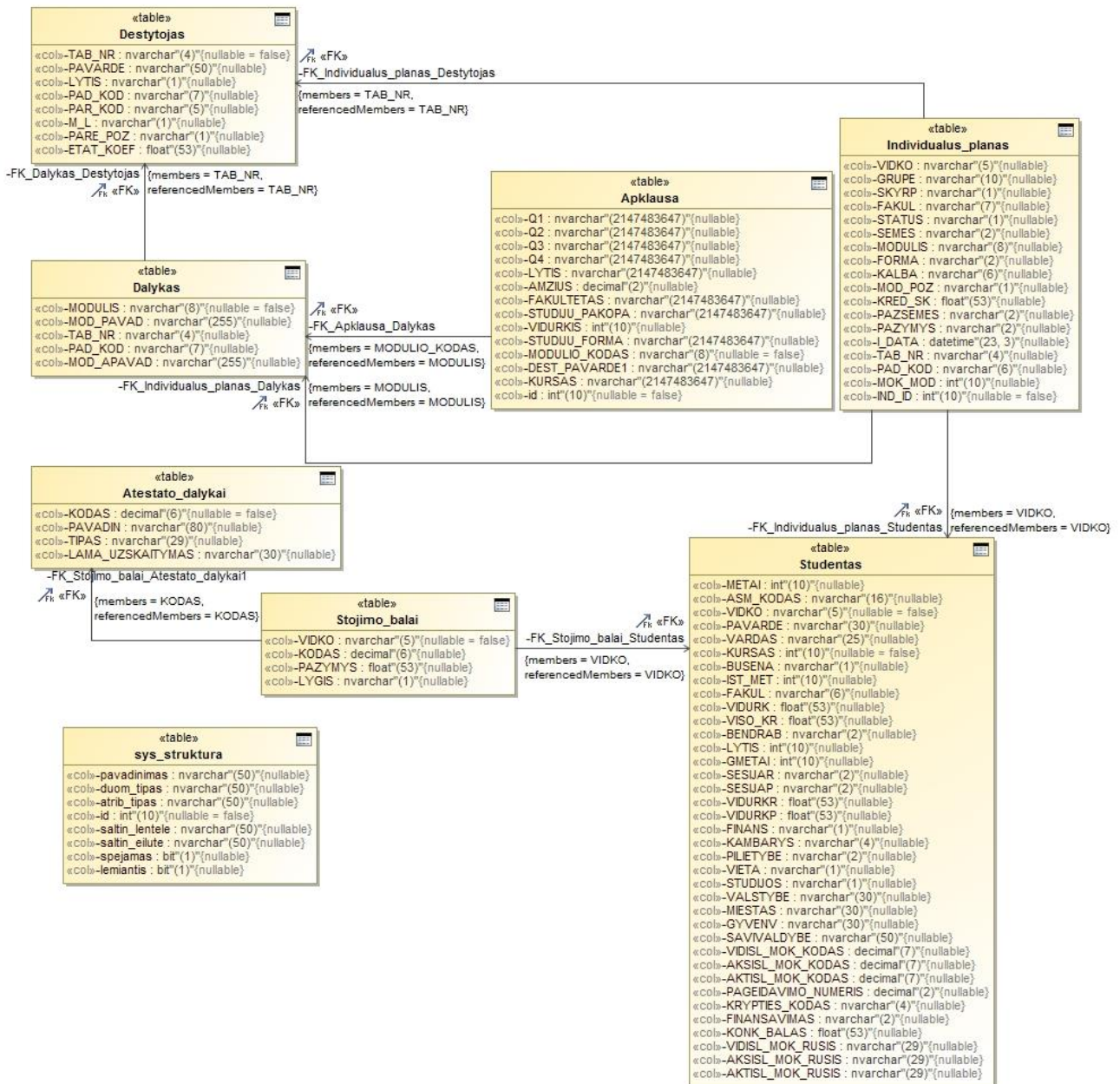
<b>Lauko pavadinimas</b>	<b>Paskirtis</b>
Savivaldybe	Savivaldybė, kurioje gyveno studentas
Vidisl_mok_kodas	Vidurinės mokyklos, kurią baigė studentas kodas
Aksisl_mok_kodas	Aukštesniosios mokyklos, kurią baigė studentas kodas
Aktisl_mok_kodas	Aukštosios mokyklos, kurią baigė studentas kodas
Pageidavimo_numeris	Pageidavimo numeris, kuriuo įstojo studentas
Krypties_kodas	Krypties kodas, kurios studijų programą mokosi studentas
Finansavimas	Į kokio finansavimo vietą įstojo studentas
Konk_balas	Kokį konkursinį balą turėjo studomas studentas
Vidisl_mok_rusis	Kokios rūšies mokyklą baigė studentas gavęs vidurinį išsilavinimą
Aksisl_mok_rusis	Kokios rūšies mokyklą baigė studentas gavęs aukštesnįjį išsilavinimą
Aktisl_mok_rusis	Kokios rūšies mokyklą baigė studentas gavęs aukštąjį išsilavinimą

Duomenų bazėje taip pat egzistuoja nemažai lentelių, kurios yra klasifikatoriai. Šie klasifikatoriai paprastai nustatomi universiteto valdžios arba kitų valstybinių institutų. Dėl reto šios informacijos pasikeitimo labai patogu ją atskirti ir naudoti tik indeksus į ją. Žemiau esančioje 3.9 lentelėje pateiktas visų šių klasifikatorių lentelių sąrašas ir nurodoma, kam šios lentelės skirtos.

**3.9 lentelė.** Duomenų bazės klasifikatorių sąrašas

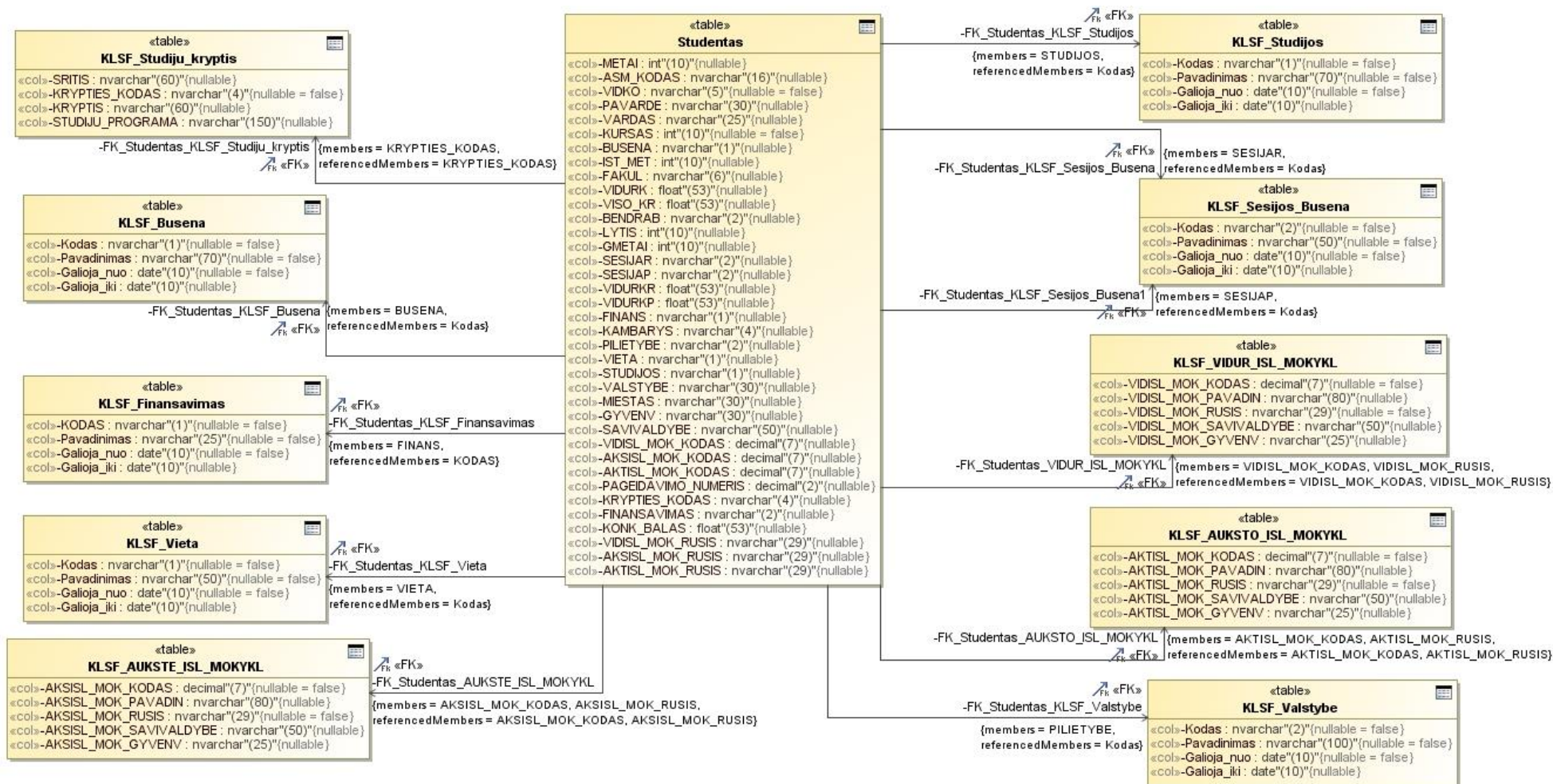
<b>Lentelė</b>	<b>Paskirtis</b>
KLSF_Aukste_isl_mokykl	Saugomas aukštesniųjų mokyklų sąrašas
KLSF_Auksto_isl_mokykl	Saugomas aukštųjų mokyklų sąrašas
KLSF_Busena	Saugomos visos būsenos, kurias gali turėti studentas universitete
KLSF_Finansavimas	Saugojami visi įmanomi finansavimai, kuriuos studentas gali gauti
KLSF_Ivertinimai	Saugojami visi universitete egzistuojantys modulio įvertinimai
KLSF_Modulio_mokescio_pozymis	Saugomi visi įmanomi mokėjimo už modulį variantai
KLSF_Modulio_pozymis	Saugomi visi universiteto modulių požymiai (privalomas, laisvai pasirinktas ir pan.)
KLSF_Mokslinis_laipsnis	Saugomi moksliniai laipsniai (habilituotas daktaras, daktaras)
KLSF_Padaliny	Saugomi visi universiteto padaliniai
KLSF_Pareigu_kodas	Saugojamos dėstytojų pareigybės (lektorius, docentas, profesorius)
KLSF_Pareigu_pozymis	Saugojami pareigų požymiai (pagrindinės, papildomos)
KLSF_Sesijos_busena	Saugojamos visos galimos sesijos būsenos
KLSF_Studijos	Saugojamos visos studijų pakopos
KLSF_Studiju_forma	Saugojamos visos galimos studijų formos
KLSF_Studiju_kryptis	Saugojamos studijų kryptys ir studijų programos
KLSF_Valstybe	Saugojamas valstybių sąrašas
KLSF_Vidur_isl_mokyl	Saugojamas vidurinį išsilavinimą suteikiančių mokyklų sąrašas
KLSF_Vieta	Saugojami studento vietų tipai

Duomenų bazės schema taip pat pateikta trijuose modeliuose. Pirmame modelyje 3.30 paveiksle parodytos tik lentelės, kurios nėra klasifikatoriai. Sekančiame 3.31 paveiksle parodyta lentelė Studentas su klasifikatoriais priklausančiais jai. Trečiame 3.32 paveiksle parodytos likusios lentelės su joms priklausančiais klasifikatoriais.



3.30 pav. Duomenų bazės schema be klasifikatorių





3.31 pav. Duomenų bazės lentelė Studentas su savo klasifikatoriais



## **4. SUKURTO DUOMENŲ GAVYBOS SPRENDIMO ĮVERTINIMAS**

### **4.1. Eksperimento planas**

Visas sukurtas sprendimas susideda iš dviejų dalių. Pirmoji dalis – prognozės atlikimas, pasinaudojant analizės paslaugose esančiais duomenų gavybos modeliais. Antroji dalis – teksto gavybos atlikimas, pasinaudojant integravimo paslaugose esančiomis transformacijomis. Kadangi abi šios dalys naudoja skirtingas duomenų bazių valymo sistemos SQL Server dalis, abi jas reikia patikrinti.

Tikrinant dalį, kuri naudoja analizės paslaugų duomenų gavybos algoritmus, reikia patikrinti, kaip tiksliai jie prognozuoja. Taip pat, kadangi modeliai su struktūromis kuriami internetinės sąsajos pagalba pagal joje nurodytą informaciją, reikia patikrinti, ar modeliai tikrai susikuria tokie, kokie formoje nurodyti.

Tuo tarpu, tikrinant teksto gavybos veikimą, reikia patikrinti, ar surandamos svarbios frazės pateiktame tekste. Tai atliekama vienos iš sistemos funkcijų pagalba „Rodyti frazių vietas tekste“, kurios veikimas buvo pateiktas anksčiau. Kadangi SQL Serverio integravimo paslaugų transformacijos neteisingai veikia su lietuvių kalba, studijų dalykai buvo įvertinti, naudojantis kitokiu principu. Visi šie dalykai toliau pateikti sekančiuose skyreliuose.

### **4.2. Duomenų gavybos algoritmų kūrimo ir tikslumo patikrinimas**

Pirmiausia bus patikrinta, ar duomenų gavybos struktūros ir modeliai sukuriami korektiškai. Tam bus palyginta struktūra formoje ir modelyje, kuris susikurs analizės paslaugų serveryje. Sukūrus modelį bus skaičiuojama, kaip tiksliai jis prognozuoja prognozuojamą atributą.

#### **4.2.1. Duomenų gavybos sukurto modelio atitikimas nurodytam formoje**

Kadangi kuriama duomenų gavybos struktūra ir modelis kuriamas ne su Business Intelligence Development studija, o internetinės sąsajos pagalba, reikia patikrinti, ar modelis tikrai susikuria toks, koks nurodytas. Ši problema gali įvykti, nes visos užklausos yra sudarinėjamos dinamiškai, ir yra nemaža klaidos tikimybė.

Tarkime, kad norime sukurti modelį, kuriame matomi universitete besimokančių studentų įvertinimai. Modelyje reikia modelyje turėti tokią studento informaciją: kokiame kurse mokosi studentas, studento lytis, studento vidurkis ir kokioje būsenoje yra studento studijos. Iš šio duomenų rinkinio atributas, kurį reikia prognozuoti, yra studento būseną. Tokios struktūros sukūrimui formoje į struktūrą turi būti sukelti tokie laukai, kokie matomi 4.1 paveiksle.

Lentele	Laukas	LaukoTipas	Pavadinimas modelyje	Diskretizavimas	Spejamas	Lemiantis	TikSpejamas
Studentas	VIDKO	text	VIDKO	Key	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Studentas	BUSENA	nvarchar	busena	Discrete	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Studentas	KURSAS	int	kursas	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Studentas	LYTIS	int	lytis	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Studentas	VIDURK	float	vidurkis	Discrete	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

**4.1 pav.** Pasirinkti laukai, kurie įtraukiami į duomenų gavybos modelį, matomi formoje

Studento vidinis identifikacinis kodas (VIDKO) yra visada automatiškai įtraukiamas į struktūrą ir naudojamas kaip pirminis raktas. Į duomenų gavybą įtraukiami ir dalykų, kuriuos mokėsi studentas universitete, įvertinimų atributai. Dėl to, kad neįtraukti tokių atributų, kurie turi beveik vien tuščias reikšmes, yra įvestas apribojimas, kiek mažiausiai reikšmių turi turėti laukas su duomenimis. Šis pasirinktas skaičius yra dvidešimt penki – turi būti daugiau nei dvidešimt penki studentai, mokėsi studijų modulį. Jeigu studentas pradėjo studijuoti tam tikrą dalyką, visada tam tikras įvertinimas atsiranda duomenyse – pažymys arba skaičius nulis, kai modulis braukiamas iš studijų planų ar įvyksta kita situacija, kai studentas negauna įvertinimo.

Kai duomenų gavybos modelis ir struktūra yra sukurtas, galima Microsoft SQL Server Business Intelligence Development studijos pagalba, atsidaryti analizės paslaugų duomenų bazę ir peržiūrėti, kokie atributai yra panaudoti modelyje. Ankščiau aprašytu atveju gauname vaizdą, kuris parodytas 4.2 paveiksle.

Structure ▲	dalykumodelis
	Microsoft_Clustering
	PredictOnly
busena	Input
H120B031	Input
H570B101	Input
H570B104	Input
H570B107	Input
H592B011	Input
kursas	Input
lytis	Input
P000B105	Input
P130B001	Input
P130B002	Input
P160B003	Input
P170B008	Input
P175B100	Input
P175B304	Input
P175B312	Input
P175B501	Input
P175B502	Input
P175B505	Input
P175B602	Input
P190B101	Input
P230B202	Input
S180B302	Input
S265B010	Input
S273B001	Input
S273B002	Input
T121B101	Input
T150B016	Input
T190B393	Input
VIDKO	Key
vidurkis	Input

4.2 pav. Duomenų gavybos modelis analizės duomenų bazėje

Kaip matome, palyginus abudu paveikslėlius, modelio struktūra yra visiškai tokia pati. Skiriasi tik tuo, kad interneto sąsajoje nerodomi universiteto studijų dalykų kodai. Toliau reikia patikrinti, kaip tiksliai modelis spėja studijų būklės atributo reikšmes.

#### 4.2.2. Duomenų gavybos modelio tikslumo įvertinimas

Patikrinti duomenų gavybos modelio tikslumą galima pagal tai, kokios prognozuojamo atributo reikšmės yra gaunamos palyginti su reikšmėmis, kurios yra realiuose duomenyse. Pavyzdyje, kuris buvo aprašytas praeitame skyrelyje, toks atributas buvo studento studijų būklė.

Šis atributas dažniausiai turi dvi reikšmes: nulį, kai studentas iškrenta iš universiteto, ir vienetą, kai mokosi universitete.

Tam, kad suskaičiuoti būtų lengviau duomenų gavybos modelio prognozavimo tikslumą, paimta atsitiktiniai 100 studentų iš duomenų rinkinio. Suskaičiuota, kiek studentų turi būseną vienas ir kiek nulis. Taip pat suskaičiuota, kiek kartų apsirikta prognozuojant. Gautas rezultatas, kuris parodytas sekančioje 4.1 lentelė

**4.1 lentelė.** Būsenų pasiskirstymas, kai prognozuojantys duomenys yra universiteto modulių pažymiai

	<b>Prognozuojamos reikšmės</b>	
<b>Tikros reikšmės</b>	<b>0</b>	<b>1</b>
<b>0</b>	27	5
<b>1</b>	1	67

Matome, kad iš šimto studentų 5 studentai pateko į neiškritusių sąrašą, nors jie realiai iškrito. Tuo tarpu vienas studentas tapo iškritusiu, kai toks iš tikrųjų nebuvo. Kaip pastebime, iš viso visame duomenų rinkinyje neteisingas būsenas gavo 6 studentai. Todėl modelio tikslumas yra 94 procentai.

Antrasis modelis, kurį galima sudaryti yra, kai prognozės atliekamos pagal dėstytojų sąrašą. Tokiu atveju yra laikoma, kad įvertinimas gautas ne nuo modulio, bet nuo dėstytojo. Vėl paimama atsitiktiniai 100 studentų duomenys. Tokiu būdu gaunamas rezultatas toks, kaip matoma 4.2 lentelėje.

**4.2 lentelė.** Būsenų pasiskirstymas, kai prognozuojantys duomenys yra universiteto dėstytojų parašyti įvertinimai

	<b>Prognozuojamos reikšmės</b>	
<b>Tikros reikšmės</b>	<b>0</b>	<b>1</b>
<b>0</b>	28	4
<b>1</b>	3	65

Galima pastebėti, kad iš 100 parinktų studentų iš viso 7 studentai pateko ne į tą studijų būseną. Į besimokančiųjų būseną, nors iš tikrųjų iškrito, buvo priskaičiuoti 3 studentai. Tuo tarpu studentų, kurių tikra būseną yra „Mokosi“, o parinkta „Iškrito“ yra 3. Apibendrinus gauname, kad modelio tikslumas yra 93 procentai.

Trečiasis modelis, kurį galima tirti, yra, kai prognozėje yra naudojama mokyklos brandos egzaminų ir mokytojų dalykų įvertinimai. Šiuose duomenyse valstybinis brandos egzaminas, mokyklinis brandos egzaminas ir metinis įvertinimas yra, kaip trys atskiri dalykai, kuriuos mokėsi stojantysis. Atlikti tikslumo įvertinimą vėl paimama 100 studentų duomenys. Tokiu atveju gaunamas rezultatas matomas 4.3 lentelėje.

**4.3 lentelė.** Būsenų pasiskirstymas, kai prognozuojantys duomenys yra mokykloje gautų dalykų įvertinimai

<b>Prognozuojamos reikšmės</b>	<b>0</b>	<b>1</b>
<b>Tikros reikšmės</b>		
<b>0</b>	7	19
<b>1</b>	4	70

Pagal gautą rezultatą matome, kad didelis netikslumas susidaro su būseną išskirto. Net 19 studentų pateko klaidingai iš šios būsenos į besimokančiųjų būseną. Tuo tarpu atvirkštinis variantas – kai tikra būseną yra mokosi, o priskiriama į būseną išskirto - 4 studentai. Modelio tikslumas yra 77 procentai. Toks tikslumas gaunamas todėl, kad stojančiųjų dalykai per eilę metų labai stipriai keitėsi dėl studijų reformų, taip pat skirtingose mokyklose mokoma kai kurių dalykų skirtingos versijos. Ypatingai tokia situacija pasireiškia su kalbų dalykais. Seniau egzaminai buvo laikomi žodžiu, dėl to jie taip pat turi atskirą eilutę stojimo duomenyse.

Galiausia galima sukurti modelį, kuriame naudojama informacija apie dalykų įvertinimus universitete ir mokykloje. Tokiu atveju kuriamas bendras sukinys. Atlikus prognozę gaunamas būsenų pasiskirstymas toks, kaip matoma 4.4 lentelėje.

**4.4 lentelė.** Būsenų pasiskirstymas, kai naudojami apjungti duomenys

<b>Prognozuojamos reikšmės</b>	<b>0</b>	<b>1</b>
<b>Tikros reikšmės</b>		
<b>0</b>	28	1
<b>1</b>	13	58

Gautas tikslumas yra geresnis, nei buvo naudojant tik mokyklos egzaminų ir metinių įvertinimų informaciją, tačiau blogesnis, nei naudojant tik universiteto duomenis. Iš viso teisingai prognozuotų būsenų gauta 86 iš 100 atvejų.

Apibendrinant galima teigti, kad prognozės su duomenimis, kurie buvo sukurti studentui besimokant universitete, yra vykdomos gerai. Tuo tarpu tada, kai naudojami duomenys iš stojimo duomenų, yra gana didelė paklaida, ypač su išskirto būseną.

### **4.3. Teksto gavybos įvertinimas**

#### **4.3.1. Naudojama metodika vertinti teksto gavybą**

Norint atlikti teksto gavybos veikimo įvertinimą, reikia turėti galimybę matyti visą tekstą bei tame tekste pažymėtas frazes, kurios lemia studijų modulio įvertinimą. Ši galimybė yra realizuota sistemos funkcijoje „Rodyti frazių buvimo vietas tekste“. Visame tekste skirtingomis spalvomis yra nuspalvinama kiekviena frazė, kuri turi skirtingą įvertinimą. Pavyzdžiui frazė, kuri turi įvertinimą -1, visada yra spalvinama raudona spalva. Visi galimi įvertinimai anksčiau buvo pateikti 3.1 lentelėje.

Sukurtoje formoje interneto puslapyje reikia pasirinkti, kurio universiteto studijų modulio ir kurio anketos klausimo atsakymus norime matyti. Tai atlikus reikia pasirinkti, koku būdu gautas frazes norime atvaizduoti. Šios formos vaizdas su tekstais matomas 4.3 paveiksle.

**DUOMENŲ GAVYBOS TAIKYMAS** Sveiki svečias [Prisijungti](#)

**STUDIJŲ PROCESŲ**

Šios formos pagalba galima pamatyti kaip kokios frazės lemia studijų dalyko įvertinimą. Pasirinkite norima dalyką ir vieną iš trijų klausimų. Tada žemiau esančioje lentelėje bus parodyti studentų atsakymai į klausimą bei nuspalvintos frazės lemiančios įvertį.

Pasirinkite ką rodyti:

Modulis:  Klausimas:

Spalvų paaiškinimo lentelė

Labai neigiamai įtakoja (-1) Neigiamai įtakoja (-0.5) Truputėli neigiamai įtakoja (-0.2)  
 Jaučiama nežymi neigiama įtaka (-0.1) Nėra įtakos (0) Jaučiama nežymi teigiama įtaka (0.1)  
 Truputėli teigiamai įtakoja (0.2) Teigiamai įtakoja (0.5) Labai teigiamai įtakoja (1)

Eil. Nr.	Sakinys	Įvertinimas
1	Paskatinti studentus rinktis papildomus balus.	0
2	-	0
3	viskas puiku	1

Žodžiai: [viskas puiku](#)

**4.3 pav.** Frazių spalvinimo forma internetinėje sąsajoje

Rezultate pateikiamas atsakymo eilės numeris, pats atsakymo tekstas, kuriame nuspalvintos frazės, lemiančios kokybę, bei viso to teksto įvertinimas. Žemiau po šia informacija yra pateiktas frazių sąrašas, kuris buvo surastas tekste.

#### 4.3.2. Eksperto darbo su frazių žodynu įvertinimas

Norint atlikti bet kurią teksto gavybą, pirmiausiai reikia turėti frazių žodyną, kuriame turi būti surašyti kiekvienos iš frazių įvertinimai. Šiuos įvertinimus turi surašyti dalykinės srities ekspertas, pasinaudodamas savo asmenine patirtimi. Frazių įvertinimai rašomi taip, kaip buvo apžvelgta anksčiau skyrelyje Frazių įvertinimo rekomendacijos.

Surašant šiuos įverčius yra sugaištamas tam tikras laiko tarpas. Įvertinant frazes, kurias surado integravimo paslaugų transformacija, buvo suskaičiuota, kiek laiko ekspertas užtruko įvertinti žodžius. Buvo skaičiuojama kiekvieno iš trijų klausimų įvertinimo laikas. Buvo gautos tokios trukmės, kokios pateiktos 4.5 lentelėje.

**4.5 lentelė.** Frazių žodyno įvertinimo trukmės

Klausimas	Frazių skaičius	Užtrukta laiko	Vienai frazei skirta laiko
Dalyko teigiami aspektai	386	Apie 15min 30sek.	2,4 sek.
Dalyko neigiami aspektai	352	Apie 18min	3 sek
Pasiūlymai, ką pakeisti	242	Apie 13min 30sek	3,3 sek

Matoma tendencija, kad laikas skirtas vienai frazei ilgėja. Tai galima paaiškinti tuo, kad ekspertas pradeda jausti nuovargį, atlikdamas tą patį monotonišką darbą.



### 4.3.3. Integravimo paslaugų projekto pritaikymo tinkamumo įvertinimas

Pasinaudojant prieš tai aprašyta priemone, pirmiausia reikia peržiūrėti, kiek žodžių, kurie yra svarbūs, panaudojama skaičiuojant įvertinimą. Dar prieš pradėdant eksperimentą, buvo pastebėta, kad SQL Server integravimo paslaugų transformacijos, kurios naudojamos teksto gavybai, nepalaiko kitokių raidžių nei lotyniškos. Todėl teko paruošti tekstus be lietuviškų simbolių. Pavyzdys, kaip atrodo teksto spalvinimas, kai pasirinktas Profesinės kalbos kultūros modulis ir Nurodykite dalyko teigiamus aspektus klausimas, pateiktas 4.4 paveiksle.

Eil. Nr.	Sakinys	Įvertinimas
1	Labai sauni destytoja, labai malonu su ja bendrauti paskaitu metu <i>Žodžiai:</i>	0
2	Pravarcios ir <b>idomios paskaitos</b> , praktikos paskaitos tikrai pravars ateityj. <i>Žodžiai:</i> <u>idomios paskaitos</u>	0,5
3	Viskas idealu. <i>Žodžiai:</i>	0
4	<b>Maloni atmosfera</b> <i>Žodžiai:</i> <u>Maloni atmosfera</u>	0,5
5	Gera destytoja <i>Žodžiai:</i>	0
6	Apzvelgiamos gyvenimiskos klaidos, skatinamas lietuviu kalbos gerbimas. <i>Žodžiai:</i>	0
7	susizinojau daug svarbiu dalyku tiesiogiai nesusijusiu bet svarbiu mano specialybei. <i>Žodžiai:</i>	0

**4.4 pav.** Profesinės kalbos kultūros dalyko teigiamų aspektų įverčiai gauti integravimo paslaugų projektu

Peržiūrint gautą nuspalvintą tekstą, matoma tendencija, kad labai daug svarbių frazių yra nerasta. Kai kurių universiteto studijų modulių anketos klausimų atsakymuose niekas nenuspalvinta. Tai reiškia, kad algoritmas nesugeba surasti jokių frazių pagal žodyną. Taip nutinka dėl sekančių priežasčių:

- SQL Server neturi integruoto lietuvių kalbos gramatinių taisyklių palaikymo.
- Nepalaikomos nelotyniškos raidės.
- Studentai daro nemažai gramatinių klaidų, atsakinėdami į klausimus.
- Labai mažas procentas studentų atsakinėja anketas.
- Netgi užpildę anketą studentai neretai atvirus klausimus palieka tuščius.
- Kai kurie atsakymai į klausimus yra labai nerišlūs su klaidinga sintakse.

Dėl šių problemų, labiausiai susijusių dėl lietuvių kalbos nepalaikymo, algoritmas neatpažįsta kalbos dalių, todėl negali sudaryti žodyno. Surandamos visiškai nesvarbios frazės, tokios, kaip žodis „Dėstytojas“. Vien šis žodis nepasako nieko, kas lemia studijų kokybę. Tai galima paaiškinti tuo, kad žodis studentų naudojamas ypatingai dažnai atsakymuose. Pirmajame klausime apie privalumus šis žodis panaudotas 434 atsakymuose.

Norint gauti tikslesnį modulio įvertinimą, galima naudoti žodyną, nesinaudojant SQL Server integravimo paslaugų projektu tam, kad surasti naujai paduotų tekstų svarbių frazių. Apie šį analizės būdą yra aprašyta kitame skyrelyje.

#### 4.3.4. Teksto gavybos taikymo, tiesiogiai ieškant frazių tekste, įvertinimas

Antrasis būdas analizuoti tekstą yra tiesiogiai ieškoti žodyno frazių universiteto studijų modulių kokybės klausimų atsakymuose. Tokiu būdu nesinaudojama integravimo paslaugomis. Yra tik naudojamas žodynas, kurio frazės surastos Frazių išgavimo (*angl. Term Extraction*) transformacijos pagalba iš visų universiteto fakultetų studentų atsakymų į anketos klausimus. Vienintelis trūkumas yra tas, kad sudarytame frazių žodyne frazes reikia paversti į tekstą su lietuviškais simboliais. Taip pat tekstas, iš kurio gauname žodyną, turi būti tik su lotyniškėmis raidėmis. Dirbant šiuo būdu yra paimama po vieną frazę iš žodyno ir ieškoma jos tekste. Jei ji randama – įtraukiama į sąrašą: frazė tampa nuspalvinta, ir parašoma jos įtaka kokybei.

Atvejais, kaip atrodo teksto nuspalvinimas, kai tiriamas Profesinės kalbos kultūra studijų modulis, ir pasirenkamas klausimas „Nurodykite dalyko teigiamus aspektus“, matomas 4.5 paveiksle. Jame pateikta dalis visų atsakymų.

Eil. Nr.	Sakinys	Įvertinimas
1	<b>Labai šauni dėstytoja</b> , labai malonu su ja bendrauti paskaitu metu Žodžiai: <u>šauni dėstytoja</u> , <u>Labai</u>	1,1
2	Pravarčios ir <b>įdomios paskaitos</b> , praktikos paskaitos tikrai pravers ateityj. Žodžiai: <u>paskaitos</u> , <u>įdomios</u>	0,4
3	<b>Viskas idealu</b> . Žodžiai: <u>Viskas idealu</u>	1
4	<b>Maloni atmosfera</b> Žodžiai: <u>Maloni atmosfera</u>	0,5
5	<b>Gera dėstytoja</b> Žodžiai: <u>Gera dėstytoja</u>	1
6	Apžvelgiamos gyvenimiškos klaidos, skatinamas lietuvių kalbos gerbimas. Žodžiai:	0
7	susižinojau daug svarbių dalykų tiesiogiai nesusijusių <b>bet</b> svarbių mano specialybei. Žodžiai: <u>bet</u>	0

#### 4.5 pav. Profesinės kalbos kultūros teigiamų aspektų įverčiai, tiesiogiai ieškant frazių žodyne

Palyginti su anksčiau aprašytu būdu, matome, kad nuspalvinto teksto yra daugiau. Kai kuriuose atsakymuose vis dar nėra nei vienos frazės, kuri lemia kokybę. Iš surastų frazių matome frazes, kurios, be jokios abejonės, tikrai modulio kokybę lemia teigiamai. Šios frazės viso universiteto mastu dažnai pasitaiko atsakymuose.

Tačiau, kaip matoma anksčiau pateiktame paveikslėlyje, nevisos svarbios frazės yra pastebimos. Taip yra dėl to, kad sudaromas žodynas turi trūkumų, kurie buvo ankstesniame skyrelyje. Pagerinti rezultatą galima pasinaudojant eksperto, kuris gerai supranta studijų kokybės procesą, pagalba. Jis gali į žodynus įtraukti frazes, kurios yra nepastebimos algoritmu, tačiau yra svarbios studijų kokybės įvertinimui.

#### 4.3.5. Teksto gavybos taikymo, kai žodynas yra papildytas eksperto įvertinimas

Norint pagerinti studijų modulio įvertinimo skaičiavimą, ekspertas gali pridėti naujas frazes. Kiekvienam iš trijų klausimų buvo atsitiktinai parinkti 100 atsakymų į klausimus. Ekspertas perskaito atsakymą ir iš jo išrenka svarbias frazes bei jas įvertina. Po to buvo dar kartą peržiūrėti atsakymų tekstai. Vienas iš pavyzdžių, kai buvo surasta svarbių frazių, yra Matematika 1 modulis ir jo klausimas Nurodykite dalyko neigiamus aspektus. Šio klausimo keletas atsakymų, iki pasinaudojant eksperto papildytu žodynu, parodyta 4.6 paveiksle.

3	Prieš <b>uždavinių</b> sprendimą nepasako uždavinio sąlygos, <b>ir</b> kartais <b>kai</b> eina po tris studentus prie lentos, nespėja kiti padaryti <b>uždavinių</b> tiek Teorinės paskaitos metu galėtų būti daugiau uždavinių ir aiškiau aiškinti	-0,6
	Žodžiai: <u>kai, ir, uždavinių, uždavinių</u>	
5	Informacijos perteklius, kurį sunku įsisavinti.	0
	Žodžiai:	
17	<b>Gal</b> paskaitos metu teorija <b>ir</b> pavyzdžiai dėstomi šiek tiek per greitai. Kartais sunku suspėti nusirašyti.	0
	Žodžiai: <u>ir, Gal</u>	

#### 4.6 pav. Matematika 1 modulio neigiamų aspektų tekstai kai naudojamas ne eksperto sudarytas žodynas

Kaip matome, atsakymai 5 ir 17 neturi visiškai jokios svarbios frazės. Todėl ekspertas, perskaitęs tokius atsakymus, gali nuspręsti, kad frazes iš šių tekstų reikia įtraukti į žodyną. Tai atlikus, kai kitą kartą sistema pastebės tokias frazes, jos bus nuspalvinamos. Šie tekstai po eksperto peržiūrėjimo atrodo taip, kaip matoma 4.7 paveiksle.

3	Prieš <b>uždavinių</b> sprendimą nepasako uždavinio sąlygos, <b>ir</b> kartais <b>kai</b> eina po tris studentus prie lentos, nespėja kiti padaryti <b>uždavinių</b> tiek Teorinės paskaitos metu galėtų būti daugiau uždavinių ir aiškiau <b>aiškinti</b>	-0,7
5	Informacijos perteklius, kurį <b>sunku įsisavinti</b> .	-1
	Žodžiai: <u>sunku įsisavinti</u>	
17	<b>Gal</b> paskaitos metu teorija <b>ir</b> pavyzdžiai dėstomi šiek tiek per greitai. Kartais <b>sunku suspėti</b> nusirašyti.	-0,5
	Žodžiai: <u>sunku suspėti, ir, Gal</u>	

#### 4.7 pav. Matematika 1 modulio neigiamų aspektų tekstai kai žodyną yra papildęs ekspertas

Kaip matome, įtrauktos frazės, kurios tikrai neigiamai įtakoja studijų modulio kokybės įvertimą. Per greitas dėstymo tempas ir sunkiai įsisavinama medžiaga, kurios yra per daug, tampa studijų dalyko problemomis.

Vertinant ekspertui, dažniausiai pastebimos frazės, kurios susideda iš daugiau nei vieno žodžio. Labiau tikėtina, kad ekspertas įtrauks frazę „Sunku įsisavinti“ negu frazę „Sunku“. Taip labiau matoma tokios frazės prasmė. Įtraukiant frazes išskyla sunkumų. Pirmiausia, kai kurie žodžiai vienais atvejais gali reikšti teigiamą įvertinimą, o kitu atveju - neigiamą įvertinimą arba būti neutralūs. Taip dažnai nutinka su fraze „Nėra“. Jei klausimas yra apie teigiamus aspektus, tai dažnai gali reikšti, kad nėra jokių teigiamų aspektų. Tačiau, kai šis žodis eina kartu su kitu žodžiu, gali reikšti teigiamą įvertinimą.

Dar viena problema yra ta, kad fiziškai neįmanoma į žodyną įtraukti visų galimų žodžių formų. Lietuvių kalboje žodžiai asmenuojami, linksniuojami ir kitaip kaitomi, todėl norint turėti idealų žodyną, reikėtų visas galimas kombinacijas įrašyti į žodyną.

#### 4.3.6. Morfologiškai anototo teksto panaudojimo galimybė

Norint išspręsti praeitame skyrelyje aprašytą problemą, reikia pasinaudoti morfologiniu teksto anotatoriumi. Tam tinka analizės dalyje paminėtas Vytauto Didžiojo universiteto sukurtas sprendimas. Kadangi šio darbo kūrimo šis anotatorius metu buvo pertvarkomas, nepavyko gauti tinkamai anototo teksto. Tik prieš pat darbo pabaigą, apie gegužės pradžią vėl buvo galima korektiškai naudoti šį įrankį. Todėl toliau aprašomą metodiką galima panaudoti ateityje kuriamoms informacinėms sistemoms.

Naudojantis tokiu analizės būdu, pirmiausia reikia gauti iš duomenų bazės atsakytus anketų tekstus su atsakymais į klausimus. Kadangi anototas tekstas bus perkeltas į duomenų bazę pagal identifikacinį lauką, reikia turėti ir anketos lentelės pirminį raktą, kuris yra skaičius. Kadangi studentai atsakymuose naudoja ir skaitmenų simbolius, juos reikia paversti į žodžius. Tai atliekama tam, kad nesusimaišytų atsakymo tekstas su identifikaciniu numeriu.

Tai atlikus antrasis žingsnis yra turimą tekstą nukopijuoti į morfologinį anotatorių. Paspaudus pateiktoje sąsajoje mygtuką, galima parsisiųsti tekstinę bylą. Informacinėje sistemoje parašyto scenarijaus pagalba galima nuskaityti šią bylą ir perkelti anototą tekstą į duomenų bazę. Tokį tekstą galima naudoti ir pasinaudojant integravimo paslaugomis, ir atliekant tiesioginę paiešką. Pirmuoju atveju yra būtina lietuviškus simbolius pakeisti į lotyniškus dėl suderinamumo problemų.

Kadangi kai kurie žodžiai gaunami su papildomai nurodytomis galūnėmis, pavyzdžiui žodžio rašyti(-o,-ė) atveju, matome, kad papildomai pateiktos esamojo ir būtojo kartinio laiko galūnės. Šią informaciją reikia pašalinti iš žodžio, kad gautume morfologiškai anototo teksto žodį. Turint tokį tekstą galima vykdyti vieną iš analizių, kuri buvo paminėta ankstesniuose skyreliuose.

Šį algoritmą galima parodyti veiklos diagrama, kuri atvaizduota žemiau esančiame 4.8 paveiksle.



4.8 pav. Morfologiškai anototo teksto panaudojimas teksto analizei

## 5. IŠVADOS

1. Atlikus Kauno technologijos universiteto dokumentų, susijusių su studijų kokybe ir stebėseną, analizę, buvo surasta, kad ši veikla vykdoma pagal Vidinę studijų kokybės užtikrinimo sistemą. Taip pat visa universiteto veikla reorganizuojama pagal Mokslo ir studijų atsinaujinimo programą. Dėl to universitete turi pasikeisti universiteto organizacinė struktūra, tapti efektyviau išnaudojama infrastruktūra, ir studijų programos turėtų atrodyti patrauklesnės visuomenei.
2. Pastebėta, kad svarbiausias universiteto padalinys, susijęs su studijų kokybe, yra Strateginio planavimo ir kokybės departamento Kokybės skyrius. Todėl sprendimas yra suprojektuotas pagal šio padalinio poreikius ir stengiamasi išspręsti problemas, su kuriomis susiduria jo darbuotojai.
3. Visas sprendimas suprojektuotas ir realizuotas taip, kad nereikėtų naudotis specialiais įrankiais, esančiais SQL Server duomenų bazių valdymo sistemoje, kuriuos įvaldyti nėra paprasta. Visos sprendimo funkcijos pasiekiamos interneto naršyklės pagalba, todėl sumažėja tikimybė padaryti klaidą, kuri gali sugadinti bet kurią iš duomenų gavybos sprendimo objektų.
4. Kurti objektus analizės paslaugų serveryje panaudota analizės objektų valdymo (*angl. Analysis Management Objects (AMO)*) biblioteka, esanti ASP.NET karkase, o gauti duomenis iš jau sukurtų ir apmokytų modelių duomenų gavybos praplėtimo - (*angl. Data mining extensions (DMX)*) kalba. Kūrimas pirmosios pagalba yra labai patogus, nes jis vykdomas panašiai, kaip ir kitų komponentų, naudojamų objektyviame programavime. Tuo tarpu užklausų kalba patogi, nes turi panašią sintaksę kaip standartinės SQL užklausos.
5. Kuriant modelius sistemos pagalba, pastebėta, kad prognozavimas vyksta žymiai geriau, kai naudojami universitete gauti įvertinimai nei stojimo į universitetą balai. Tai galima paaiškinti tuo, kad mokyklos dalykų yra daugiau, egzistuoja ne viena to paties dalyko versija, valstybinių egzaminų įvertinimai skaičiuojami procentais, o mokyklinių balais iki 10.
6. Teksto gavybai atlikti lietuvių kalba SQL Server integravimo paslaugų transformacijos nelabai tinka, nes jos nepalaiko lietuviškų raidžių, taip pat nėra jokio lietuvių kalbos gramatikos palaikymo. Todėl nesurandamos svarbios frazės tekstuose ir negalima korektiškai vykdyti teksto gavybą.

7. Transformacijų naudojimo alternatyva vykdant teksto gavybą, yra tiesioginė frazių paieška tekste. Tokiu būdu galima skaičiuoti įvertinimą pagal tai, ar tokia frazė yra tekste ar ne. Tai atliekama, peržiūrint kiekvieną frazę žodyne.
8. Žodynas yra sudaromas dalykinės srities eksperto arba pasinaudojant SQL Server integravimo paslaugų transformacijų pagalba. Abiem atvejais ekspertas turi įvertinti kiekvieną iš frazių. Vienai frazei įvertinti vidutiniškai užtrunkama apie 3 sekundes.
9. Ateityje galima teksto gavybai panaudoti morfologiškai anotuotą tekstą, kas anksčiau buvo neįmanoma dėl vykdomų Vytauto Didžiojo universiteto morfologinio anotatoriaus tvarkymo darbų.

## LITERATŪRA

1. Kauno technologijos universiteto vidinės studijų kokybės užtikrinimo sistemos (VSKUS) aprašas: Kauno technologijos universiteto senato 2010 m. liepos 2 d. nutarimas Nr 49.
2. Kauno technologijos universiteto statutas: Lietuvos Respublikos Seimo 2012 m. birželio 28 d. nutarimo Nr. XI-2149 1 priedėlis.
3. Dėl darbo grupės fakultetuose vedamai studijų proceso dokumentacijai ir atliekamiems duomenų įrašams patikrinti sudarymo: Kauno technologijos universiteto studijų prorektorius 2008 m. rugsėjo 24 d. potvarkis Nr. PP-66
4. Kauno technologijos universitetas neakademiniai padaliniai, administracijos ir aptarnavimo padaliniai: Kauno technologijos universiteto rektoriaus 2013 m. vasario 28 d. įsakymu Nr. A-72
5. Kauno technologijos universiteto mokslo ir studijų atsinaujinimo programa ir jos įgyvendinimo planas: Priimta Kauno technologijos universiteto senato 2013 m. kovo 27 d. nutarimu Nr. V3-S-16
6. Dėl darbo grupės universiteto mokslo ir studijų atsinaujinimo programai ir jos įgyvendinimo priemonių planui parengti sudarymo: Kauno technologijos universiteto Rektoriaus 2013 m. vasario 25 d. įsakymas Nr. A-65
7. Kauno technologijos universiteto strategijos įgyvendinimo ir kokybės užtikrinimo tarnybos nuostatai: Kauno technologijos universiteto Rektoriaus 2012 sausio 31 d. įsakymu Nr. A-97
8. Kauno technologijos universiteto strategijos įgyvendinimo ir kokybės užtikrinimo tarnybos kokybės skyriaus nuostatai: Kauno technologijos universiteto Rektoriaus 2012 vasario 13 d. įsakymu Nr. A-124
9. Studentų apklausos apie studijuojamus modulius ir jų dėstymo kokybę 2011 m. pavasario semestro ataskaita. Kauno technologijos universitetas Studijų tarnyba Studijų kokybės ir stebėsenos skyrius
10. Kauno technologijos universiteto studentų apklausų sistema. [žiūrėta: 2013-02-09]. Prieiga per internetą:< <https://apklausos.ktu.lt/pagrindinis/dispatch> >
11. Liu, X. and Ruiz, M. E. (2008), Using data mining to predict K–12 students performance on large-scale assessment items related to energy. *J. Res. Sci. Teach.*, 45: 554–573.
12. Mori, H. and Kosemura, N. (2002), A data mining method for short-term load forecasting in power systems. *Elect. Eng. Jpn.*, 139: 12–22. doi: 10.1002/eej.1150

13. Suresh A, Shanmughanathan K. Important Bussiness factor analysis using datamining approach in financial sector// International Journal of Engineering Science and Technology. – 2012, Vol. 4, Issue 2, p. 606-610
14. Do Not Pay Too Much for Your Database Licenses. Microsoft. [žiūrėta: 2011-11-13]. Prieiga per internetą:< <http://download.microsoft.com/download/1/5/3/153F450F-3CC0-4137-B4CB-E0E665F6AF7C/Do%20Not%20Pay%20Too%20Much%20For%20Your%20Database%20Licenses.docx> >.
15. Jamie MacLennan, ZhaoHui Tang, Bogdan Crivat. Data Mining with Microsoft SQL Server 2008, 2009. – 636 p
16. Oracle Data Mining Concepts [interaktyvus]. [žiūrėta 2011-10-17]. Prieiga per Internetą :< [http://download.oracle.com/docs/cd/E11882\\_01/datamine.112/e16808/toc.htm](http://download.oracle.com/docs/cd/E11882_01/datamine.112/e16808/toc.htm) >
17. Dynamic Warehousing: Data Mining Made Easy [interaktyvus]. New York, 2007 [žiūrėta 2011-10-18]. Prieiga per Internetą: <http://www.redbooks.ibm.com/redbooks/pdfs/sg247418.pdf>
18. Apie Tildę. [interaktyvus]. [žiūrėta 2013-04-19]. Prieiga per internetą: <http://www.tilde.lt/apie-tilde>
19. Tildės Biuro funkcionalumas. [interaktyvus]. [žiūrėta 2013-04-19]. Prieiga per internetą: <http://www.tilde.lt/tildes-biuras/funkcionalumas/rasybos-ir-gramatikos-tikrintuvus>
20. Morfologinis anotatorius internete. Vytauto Didžiojo universitetas Kompiuterinės lingvistikos centras. [interaktyvus]. [žiūrėta 2013-04-27]. Prieiga per internetą: [http://donelaitis.vdu.lt/main.php?id=4&nr=7\\_2](http://donelaitis.vdu.lt/main.php?id=4&nr=7_2)
21. Chua, B. B. Dyson, L. E. Applying the ISO9126 model to the evaluation of an e-learning system.// In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds), Beyond the comfort zone: Proceedings of the 21<sup>st</sup> ASCILITE Conference [Sidnėjus, 2004 m. gruodžio 5-8 d.]. Sidnėjus, 2004, p. 184-190.
22. AMO Concepts and Object Model [interaktyvus]. [žiūrėta 2012-11-15]. Prieiga per internetą: <http://msdn.microsoft.com/en-us/library/bb522603.aspx>
23. Irina Gorbach, Alexander Berger, Edward Melomed. Microsoft SQL Server 2008 Analysis Services Unleashed, 2009. – 865 p
24. SQL Server 2005 Text Mining Tutorial. [interaktyvus]. [žiūrėta 2012-12-24]. Prieiga per internetą: <http://www.sqlserverdatamining.com/ssdm/Default.aspx?tabid=94&Id=164>
25. Term Extraction Transformation. [interaktyvus]. [žiūrėta 2012-12-24]. Prieiga per internetą: <http://msdn.microsoft.com/en-us/library/ms141809.aspx>



## PRIEDAI

### 1 priedas. Straipsnis paruoštas konferencijai

# DUOMENŲ GAVYBOS PRIEMONIŲ TAIKYMAS STUDENTŲ PRIĖMIMO Į UNIVERSITETĄ TYRIMUI

Vigintas Šakys<sup>1</sup>, Tomas Skroblas<sup>2</sup>

Kauno technologijos universitetas, Informacijos sistemų katedra, Studentų g. 50, LT-51368  
Kaunas, vigintas.sakys@ktu.lt<sup>1</sup>, tomas.skroblas@stud.ktu.lt<sup>2</sup>

Straipsnyje pristatoma studentų priėmimo į universitetą tyrimo metodika, parengta naudojant informacinių technologijų veiklos analitikos priemones. Iš universiteto akademinės informacinės sistemos duomenų sukuriamas duomenų saugykla ir atliekama įstojusiu į universitetą abiturientų analizė su įvairiais *Microsoft* duomenų gavybos algoritmais. Ją atlikus galima sužinoti prognozuojamus būsimų studentų pažymių vidurkius universitete, kurie iš jų turi didžiausią iškritimo iš universiteto riziką, bei įvairias kitas svarbias priėmimui tendencijas.

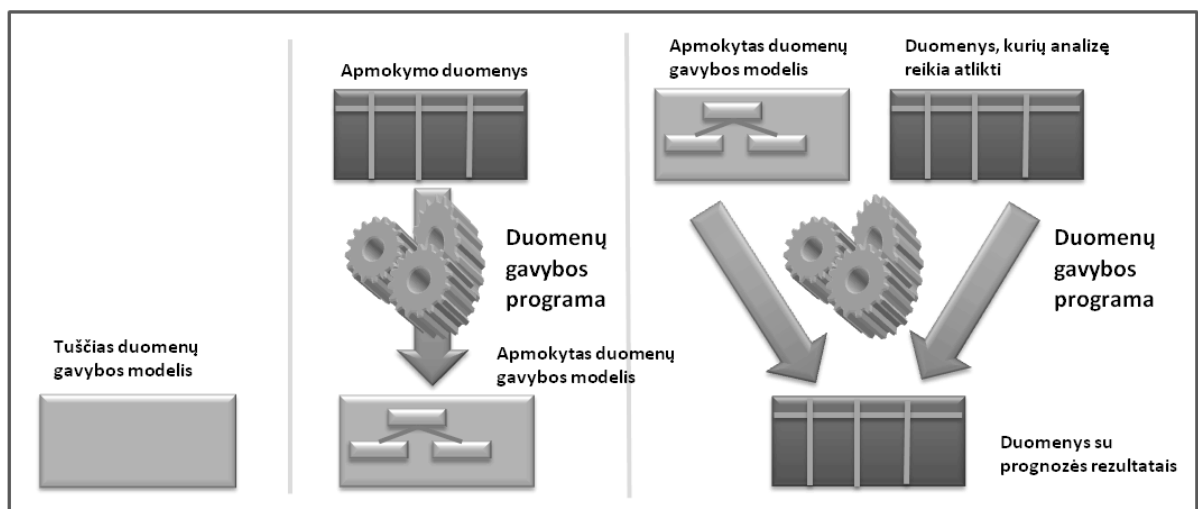
**Raktiniai žodžiai:** duomenų gavyba, duomenų saugykla, veiklos analitika

## 1 Įžanga

Šiame žinių amžiuje visoms įmonėms ir įstaigoms yra labai aktualu efektyviai panaudoti informaciją, su kuria jos susiduria kasdieninėje veikloje. Daugiausia laimi tos organizacijos, kurios sugeba analizuoti duomenis, operatyviai nustatyti ir vaizdžiai pateikti esminius rodiklius sprendimų priėmimui. Kaip ir bet kuriai kitai organizacijai Kauno technologijos universitetui svarbu žinoti studijų kokybei svarbią informaciją apie studentus ir abiturientus: jų skaičiaus kitimą, pažymių vidurkius, priimamų abiturientų mokymosi mokykloje rezultatų įtaką studijoms universitete ir pan. Gali būti naudingos šių rodiklių prognozės, kurios rodo galimas situacijas ateityje, o žinant jas galima geriau organizuoti universiteto veiklą [1]. Šiam tikslui pasiekti straipsnyje aptariama duomenų gavybos (angl. *data mining*) priemonių taikymo metodika studentų priėmimo į universitetą tyrimui. Duomenų gavyba yra viena iš modernių veiklos analitikos (angl. *business intelligence*) priemonių.

## 2 Tyrimo metodika

Duomenų gavyba yra apibrėžiama, kaip procesas, kuris, naudodamas statistinius, matematinius, dirbtinio intelekto ir sistemos mokymosi metodus, iš didelių duomenų bazių siekia rasti naudingą informaciją, o vėliau žinias, ir jas panaudoti organizacijos veiklos gerinimui [2]. Straipsnyje pateikiama studentų priėmimo į universitetą tyrimo metodika, naudojanti duomenų gavybos metodus ir *Microsoft* firmos programines priemones, kurios yra *Microsoft SQL* serverio analizės paslaugose (angl. *Microsoft SQL Server 2008 Analysis Services*). Duomenų gavybos procesas (žr. 1 pav.) susideda iš kelių žingsnių. Pirmiausiai parinktam *Microsoft* duomenų gavybos algoritmui (jį iš viso yra devyni [3]) sukuriamas tuščias gavybos modelis, po to jis apmokomas faktiniais jau seniau įstojusiu abiturientų duomenimis ir jų mokymosi universitete rezultatais. Tada algoritmui pateikiami naujų abiturientų duomenys, kurių prognozę reikia atlikti, ir randami prognozuojami vidurkia, iškritimo iš universiteto rizikos įvertinimai ir kita naudinga informacija.



1 pav. Duomenų gavybos procesas

Duomenų gavybos išsamūs rezultatai pateikiami lentelėje, kurioje matomos visos prognozuojamo atributo reikšmės ir pasirinkti pradiniai duomenų stulpeliai. Algoritmai skirstomi į keletą grupių, pagal tai kokiai dalykinei sričiai jie skirti tirti. Paprastai kiekvienas algoritmas dar pateikia papildomus apibendrintus rezultatus skirtingais grafiniais atvaizdais. Tai gali būti sprendimų medis, į grupes suskirstyti duomenys, ar įvairios lentelės ir diagramos su rastomis duomenų priklausomybėmis. Iš viso Microsoft SQL Server 2008 programoje [4] esančius algoritmus galima skirstyti į 6 kategorijas (1 lent.). Kai kurios dalykinės sritys turi daugiau nei vieną algoritmą. Tada algoritmas parenkamas pagal duomenų sudėtingumą ir kokių tikslumu bei sparta norime gauti rezultatus [5].

**Lentelė Nr. 1 Algoritmų palyginimas**

Dalykinė sritis	Microsoft duomenų gavybos algoritmai
Grupuoti objektų rinkinius pagal panašias savybes	Grupavimo (angl. <i>Clustering</i> ), Sekų grupavimo (angl. <i>Sequence Clustering</i> )
Priskirti objektus tam tikroms kategorijoms, pavyzdžiui, ar abiturientas bus išbrauktas iš universiteto	Sprendimo medžių (angl. <i>Decision Trees</i> ), Bajeso klasifikavimo (angl. <i>Naive Bayes</i> ), Neuroninių tinklų (angl. <i>Neural Network</i> ), Logistinės regresijos (angl. <i>Logistic Regression</i> )
Rasti, kuris objektas turi tendenciją būti kartu su kitu objektu	Asociacijos taisyklių (angl. <i>Association Rules</i> ), Sprendimo medžių
Rasti tam tikras reikšmes pagal kitas objekto reikšmes	Sprendimo medžių, Tiesinės regresijos (angl. <i>Linear Regression</i> ), Logistinės regresijos, Neuroninių tinklų
Prognozuoti, kaip tiriamas objektas atrodys ateityje, naudojant tam tikrus laiko intervalus, pavyzdžiui, kaip kis abiturientų skaičius ateinančiais metais	Laiko serijų (angl. <i>Time Series</i> )
Rasti, kuris objektas turi tendenciją būti kartu tam tikra tvarka su kitu objektu, pavyzdžiui, kokia seka produktai yra perkami arba kokius žiniatinklio tinklalapius vieną po kito naršo lankytojai	Sekų grupavimo

Studentų priėmimo į universitetą tyrimui buvo parinkti šie algoritmai:

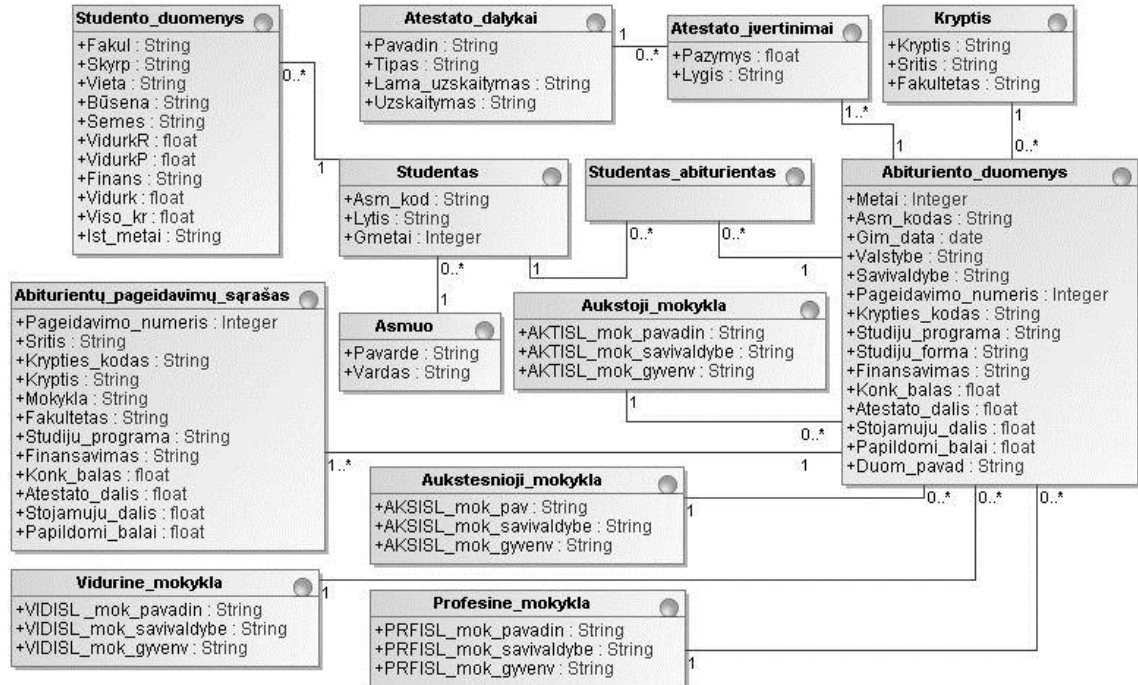
- Sprendimo medžio algoritmas panaudotas prognozuoti, kurie abiturientai turi didžiausią riziką būti išbrauti iš universiteto ir kokie bus jų pažymių vidurkiai universitete.
- Laiko serijų algoritmas panaudotas stojančių abiturientų į pasirinktą fakultetą prognozei po tam tikro metų skaičiaus.
- Grupavimo algoritmas panaudotas rasti abiturientų duomenyse slypintiems dėsningumams, pavyzdžiui, kokias savybes turintis abiturientas nestudijuos mokamose studijose.
- Asociacijos taisyklių algoritmas panaudotas rasti ryšiams tarp duomenų pagal tam tikras taisykles, pavyzdžiui, matematikos A lygio pažymys didesnis už 9,34 nurodo, kad universitete vidurkis bus didesnis už 7,6.

**Lentelė Nr. 2 Duomenų saugyklos duomenų lentelės**

Lentelės pavadinimas	Paskirtis
Studento_duomenys	Studento duomenys, kurie aprašo jo studijas universitete
Studentas	Studento bendrieji duomenys (nesusiję su studijų rezultatais)
Asmuo	Visų asmenų, kurie tiriami sistemoje vardai ir pavardės
Atestato_dalykai	Atestato dalykų pavadinimai ir tipai
Atestato_įvertinimai	Abiturientų atestato įvertinimai
Abituriento_duomenys	Abituriento bendrieji duomenys
Abiturientų_pageidavimo_sąrašas	Abiturientų pageidavimo sąrašas
Aukštoji_mokykla	Aukštųjų mokyklų bendroji informacija (pavadinimas, vieta)
Aukštesnioji_mokykla	Aukštesniųjų mokyklų bendroji informacija (pavadinimas, vieta)
Vidurinė_mokykla	Vidurinių mokyklų bendroji informacija (pavadinimas, vieta)
Profesinė_mokykla	Profesinių mokyklų bendroji informacija (pavadinimas, vieta)
Kryptis	Fakultetų studijų kryptių informacija

### 3 Naudojami duomenys

Studentų priėmimo į universitetą tyrimas buvo atliktas naudojant Kauno technologijos universiteto Informatikos fakulteto 2008 ir 2009 metų stojimo ir studijų pirmame kurse realius duomenis iš KTU akademinės IS sukūrus duomenų saugyklą. Duomenų saugyklos esybių modelis pateiktas 2 paveiksle, o lentelių paskirtis aprašyta 2 lentelėje. Visi algoritmai yra apmokami naudojant duomenų saugyklos abiturientų duomenis, kuriuose jau yra žinomi jų mokymosi universitete rezultatai.

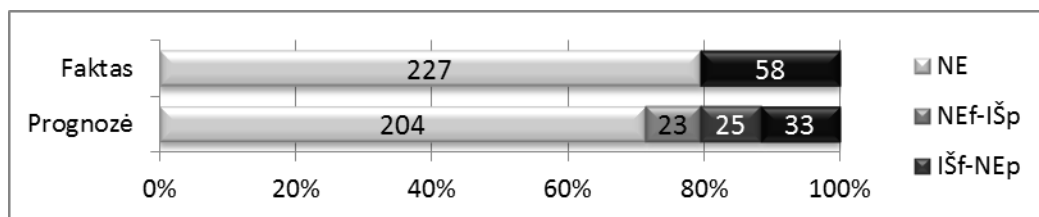


2 pav. Duomenų saugyklos loginė schema

### 4 Tyrimo rezultatai

Supažindinsime su keliais šiame duomenų gavybos projekte gautais tyrimo rezultatų pavyzdžiais.

**Sprendimų medžio** algoritmu (angl. *Microsoft Decision Trees*) atlikome 2009 metais įstojusių į Informatikos fakultetą 285 studentų iškritimo prognozę (3 pav.). Diagramoje matome, kad neiškrito (faktas), nors buvo prognozuoti, kad iškris (NEf-IŠp) 23 studentai ir iškrito (faktas), nors buvo prognozuoti, kad neiškris (IŠf-NEp) 25 studentai. Papildomai nustatyta, kad nors pirmakursiai, kurie neiškrenta sudaro 77,18% visų abiturientų, tuo tarpu paėmus tik tuos abiturientus, kurie moka už mokslą išlieka 38,30%. Svarbiausias dalykas, kuris lemia studento būseną yra finansavimas (mokama ar valstybės finansuojama vieta), po to seka Matematikos mokyklinio egzamino ir metinio A lygio įvertinimai ir kt.



3 pav. 2009 metais įstojusių į Informatikos fakultetą studentų iškritimo prognozė

Šiuo algoritmu taip pat suprognozavome būsimus studentų pažangumo vidurkius I kurse. Prognozės rezultatai parodė, kad svarbiausias dalykas lemiantis vidurkį yra matematikos A lygio metinis pažymys. Antrasis kriterijus yra finansavimo forma, o trečiasis pasirinkta studijų programa. Pavyzdžiui, jeigu abiturientas matematikos (A lygio) turėjo 10 balų – tada iš jų 73,68% studentų universitete turės didesnę vidurkį už 7,5. Tuo tarpu esant įvertinimui 5, net 53,13% studentų turės žemus pažymius (tarp 5,3 ir 6,24) ir kt.

**Asociacijų taisyklių** algoritmu (angl. *Microsoft Association Rules*) ištyrėme, kas labiausiai turi įtakos pažangumo vidurkiui I kurse. Pavyzdžiui, jei matematikos A lygio metinis pažymys didesnis už 9,34, tai pažangumo vidurkis turėtų būti didesnis už 7,6. Arba jei matematikos A lygio metinis pažymys didesnis už 9,34 ir pageidavimo numeris pirmas – tai vidurkis bus didesnis nei 7,6 ir kt.

**Grupavimo algoritmu** (angl. *Microsoft Clustering*) suskirstėme įstojusiuosius į grupes (angl. *cluster*) ir išanalizavome šį grupavimą lemiančius duomenis (4 pav.). Algoritmas rado 7 reikšmingas grupes su joms charakteringomis savybėmis. Pavyzdžiui, daugiausiai studentų su aukštais vidurkiais universitete yra 4 grupėje, daugiausiai iškritusių studentų yra 2 grupėje, tačiau joje net 75% studentų mokėjo už mokslą. Didžiausia grupė yra pirma, joje dominuoja studentai, kurie neiškrito, o visi studentai turi valstybės finansuojamas vietas.



4 pav. Grupavimo algoritmo grupių palyginimas pagal du rodiklius – BUSENA (0 – išbrauktas, 1 – studijuoja) ir FINANSAVIMAS (M – mokama vieta, VF – valstybės finansuojama vieta)

## 5 Išvados

1. Pasaulinė patirtis rodo, kad naudojant veiklos analitikos ir duomenų gavybos metodus bei priemones ieškant efektyvių sprendimų, žymiai pagerina įmonių ir organizacijų veiklos rezultatus, todėl tikėtina, kad ir universitetuose juos galima sėkmingai panaudoti tiriant sukauptus akademinėse duomenų bazėse duomenis.
2. Studentų priėmimo į universitetą tyrimams buvo panaudoti realūs KTU akademinės IS duomenys, todėl sukurti modeliai ir metodika bei gauti rezultatai gali būti vertingi tiek priėmimo procesui tobulinti, tiek studijų efektyvumui didinti.
3. Studentų išbraukimo iš universiteto ir jų pažangumo prognozė gali sudaryti prielaidas studijų kokybei gerinti, kadangi jau semestro pradžioje galima prognozuoti kiek abiturientų iškris, kokie bus jų mokymosi rodikliai. Taip pat rizikos grupėms priskirtiems studentams galima organizuoti įvairias papildomas konsultacijas, paskaitas, pratimus ir kt., taip žymiai sumažinant jų iškritimą iš universiteto.
4. Tyrimo metu gauti išsamūs ir įvairiapusiški rezultatai apie studentų priėmimo į universitetus procesą bei jų tolimesnes studijas gali būti naudingi įvairios funkcinės paskirties aukštojo mokslo administratoriams, kurie gali priimti sprendimus šio proceso argumentuotam gerinimui.

## Literatūros sąrašas

- [1] Šakys V., Butleris R. Business intelligence tools and technologies for the analysis of university studies management. Transformations in Business & Economics = Verslo ir ekonomikos transformacijos / Vilniaus universitetas, Brno technologijos universitetas, Latvijos universitetas. 2011, Vol. 10, no. 2 (23). p. 125-136.
- [2] Turban E., Sharda R., Delen D., King D., Aronson J.E. Business Intelligence. A Managerial approach. *Perentice Hall*, 2011. – 312 p.
- [3] Langit L., Goff K.S., Mauri D., Malik S., Welch J. Smart Business Intelligence Solutions with Microsoft SQL Server 2008. *Microsoft Press*, 2009. – 765 p.
- [4] Nielsen P., Microsoft MVP, White M., Parui U. Microsoft SQL Server 2008 Bible. *Wiley Publishing*, 2009. – 1642 p.
- [5] MacLennan J., Tang Z., Crivat B. Data Mining with Microsoft SQL Server 2008. *Wiley Publishing*, 2011. – 312 p.

### Research of Students Admission to the University Using Data Mining tools

This article introduces the university admission research methodology, which was developed using Microsoft data mining tools. The real data from the University Academic Information System is used to build the university data warehouse, which is then analyzed using different data mining algorithms provided by Microsoft. As a result, various tendencies that are important to the student admission processes can be revealed, including predicted grades of future students, probabilities of them dropping out, at all.