



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

LUKAS GIEDRAITIS

VEIKLOS INTELEKTIKOS METODŲ TYRIMAS IR TAIKYMAS
PARDAVIMAMS PROGNOZUOTI

Baigiamasis magistro projektas

Vadovė
doc. dr. L. Čeponienė

KAUNAS, 2016

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

VEIKLOS INTELEKTIKOS METODŲ TYRIMAS IR TAIKYMAS
PARDAVIMAMS PROGNOZUOTI

Baigiamasis magistro projektas
Informacinių sistemų inžinerijos studijų programa (kodas 621E15001)

Vadovė

doc. dr. L. Čeponienė
2016-05-18

Recenzentas

doc. dr. V. Šakys
2016-05-18

Projektą atliko

Lukas Giedraitis
2016-05-18



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

(Fakultetas)

Lukas Giedraitis

(Studento vardas, pavardė)

Informacinių sistemų inžinerijos studijų programa, 621E15001

(Studijų programos pavadinimas, kodas)

Baigiamojo projekto „Veiklos intelektikos metodų tyrimas ir taikymas pardavimams prognozuoti“

AKADEMINIO SAŽININGUMO DEKLARACIJA

20 16 m. gegužės 18 d.
Kaunas

Patvirtinu, kad mano, **Luko Giedraičio**, baigiamasis projektas tema „Veiklos intelektikos metodų tyrimas ir taikymas pardavimams prognozuoti“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Giedraitis, Lukas. Veiklos intelektikos metodų tyrimas ir taikymas pardavimams prognozuoti. Magistro baigiamasis projektas / vadovė doc. dr. Lina Čeponienė; Kauno technologijos universitetas, Informatikos fakultetas.

Mokslo kryptis ir sritis: Informatikos inžinerija, technologijos mokslai

Reikšminiai žodžiai: *prognozavimas, veiklos intelektika, pardavimai, prognozavimo metodai, metamodelis.*

Kaunas, 2016. 76 p.

SANTRAUKA

Pagrindinis šio darbo tikslas – pagerinti ir palengvinti pardavimų prognozavimą, taikant veiklos intelektikos priemones. Palengvinimą galima suprasti kaip prognozavimo modelio sudarymo paprastumą, neturint specialių matematinių ir prognozavimo žinių. Supaprastinti modelio sudarymo procesą padeda veiklos intelektikos gerosios praktikos, duomenų saugyklos kūrimo metodikos ir jų pritaikymas. Proceso rezultatų tikslumo gerinimą lemia esamų prognozavimo modelių sudarymo analizė, jų eksperimentas su realiais duomenimis ir atrastų tendencijų bei išvadų panaudojimas dalykinėje srityje. Tyrimo rezultatai orientuoti į įmonių analitikus ir sprendimų priėmėjus, jiems suteikiant galimybę numatyti ateinančio laikotarpio rinkos pokyčius ir atitinkamai pagal tai pakreipti savo organizacijos plėtrą teisinga kryptimi. Darbo metu buvo sukurtas duomenų saugyklos metamodelis, leidžiantis generuoti dinamines SQL užklausas, esant vieno lygio žvaigždės struktūros duomenų schemai. Eksperimento atlikimas su vaistų pardavimų duomenimis parodė, jog pardavimų prognozavime tiksliausius rezultatus parodė laiko eilučių modeliai, nuo jų, vertinant prognozių paklaidą, atsiliko dirbtinių neuroninių tinklų ir tiesinės regresijos modeliai.

Giedraitis, Lukas. Applying Business Intelligence Techniques For Sales Forecasting. Master's thesis in Information Systems Engineering / supervisor assoc. prof. Lina Čeponienė. The Faculty of Informatics, Kaunas University of Technology.

Research area and field: Informatics Engineering, Technology Science

Key words: *forecasting, business intelligence, sales, forecasting methods, metamodel*

Kaunas, 2016. 76 p.

SUMMARY

The main goal of this work – to improve sales forecasting, applying business intelligence techniques. Improvement can be understood as facilitation of forecasting process, so to make it available to business people without deep mathematical analysis knowledge. The accuracy of forecasting results depends on completeness of analysis of current forecasting methods and their features while using them for sales forecasting. Research results are focused on data analysts, to help them make decisions about upcoming future and shift company's processes in correct way. Research helped to create data warehouse metamodel, which allows dynamic SQL queries generation, while using star data warehouse schema. Experiment results showed that time series models provided the most accurate results while forecasting pharmaceutical drug's sales, artificial neural network and linear regression models forecasted with larger errors and could not reach accuracy of time series models.

TURINYS

Lentelių sąrašas.....	8
Paveikslų sąrašas.....	9
Įvadas.....	11
1. Probleminės srities analizė.....	13
1.1. Analizės tikslas.....	13
1.2. Tyrimo objektas, sritis ir problema.....	13
1.3. Tyrimo objekto (veiklos proceso, modelio, metodo, darinio, sistemos ar jos dalies ir pan.) analizė.....	13
1.4. Prognozavimo metodų analizė.....	17
1.5. Tyrimo objekto naudotojų analizė.....	20
1.5.1. Naudotojų tipai.....	20
1.5.2. Naudotojų atsakomybės.....	20
1.5.3. Naudotojų kvalifikacija.....	20
1.5.4. Naudotojų kiekis.....	20
1.5.5. Naudotojų tikslai.....	20
1.5.6. Naudotojų problemos.....	20
1.6. Esamų problemos sprendimo metodų analizė (Lietuvos ir tarptautiniu mastu).....	21
1.7. Darbo tikslas, uždaviniai ir siejami privalumai.....	25
1.8. Siekiamo sprendimo apibrėžimas.....	26
1.9. Analizės išvados.....	26
2. Pardavimų prognozavimo įrankio sprendimo reikalavimų specifikacija ir projektas, formalus aprašas.....	27
2.1. Reikalavimų specifikacija.....	27
2.2. Nefunkciniai reikalavimai.....	28
2.3. Dalykinės srities modelis.....	28
2.4. Naudotojų sąsajos modelis.....	33
2.5. Formalus sprendimo aprašas.....	35
3. Pardavimų prognozavimo įrankio sprendimo arba eksperimentinės realizacijos projektas.....	41
3.1. Sistemos loginė architektūra.....	41
3.2. Reikalavimų analizės modeliai.....	42
3.3. Sistemos elgsenos modeliai.....	45
3.4. Duomenų bazės modelis.....	50
3.5. Sistemos klasių modelis.....	51
3.6. Sistemos realizacijos modelis.....	52
3.7. Sistemos diegimo modelis.....	53
4. Sprendimo realizacija ir sprendimas.....	54
4.1. Sprendimo realizacijos ir veikimo aprašas.....	54
4.2. Testavimo modelis, duomenys, rezultatai.....	58

5. Eksperimentinis sprendimo tyrimas.....	61
5.1. Eksperimento dalykinė sritis.....	61
5.2. Eksperimento planas	62
5.2.1. Tiesinės regresijos ir jos parametru įvertinimas	62
5.2.2. Laiko eilučių metodų paklaidų palyginimas ir įvertinimas	62
5.2.3. Neuroninių tinklų įvertinimas.....	63
5.3. Eksperimento rezultatai	63
5.3.1. Tiesinės regresijos eksperimento rezultatai	64
5.3.2. Laiko eilučių eksperimento rezultatai.....	65
5.3.3. Neuroninių tinklų eksperimento rezultatai	68
5.3.4. Bendri eksperimento rezultatai	71
5.4. Sprendimo taikymo rekomendacijos	72
6. Rezultatų apibendrinimas ir išvados	74

LENTELIŲ SĄRAŠAS

1.1. Žmogaus rasės kintamųjų apibrėžimas dvejomis būsenomis	17
1.2. Ištirtų įrankių lyginamoji analizė	24
2.1. Esysbės „Kompanijos“ specifikacija	29
2.2. Esysbės „Vartotojai“ specifikacija	29
2.3. Esysbės „Filtrais“ specifikacija	30
2.4. Esysbės „Ataskaitos“ specifikacija	30
2.5. Esysbės „Dimensijos“ specifikacija	30
2.6. Esysbės „Parametrai“ specifikacija	31
2.7. Esysbės „Parametrų reikšmės“ specifikacija	32
4.1 Nefunkcinių reikalavimų testavimo aprašas	59
4.2 Funkcinių reikalavimų testavimo aprašas	60

PAVEIKSLŲ SĄRAŠAS

Pav. 1 Siūlomas prognozavimo metodo pasirinkimo procesas [7]	15
Pav. 2 Laiko eilutės dedamųjų išskyrimas	19
Pav. 3 „RStudio“ pagrindinio lango pavyzdys	22
Pav. 4 Testinių duomenų atvaizdavimas ir krypties komponentės išskyrimas Excel programoje	23
Pav. 5 „Base CRM“ švieslentės langas testinėje aplinkoje.....	24
Pav. 6 Prognozavimo įrankio panaudojimo atvejų diagrama	27
Pav. 7 Prognozavimo įrankio esybių klasių diagrama	29
Pav. 8 Pavyzdinės saugyklos esybių klasių modelis.....	32
Pav. 9 Metaduomenų panaudojimo duomenų aprašyme pavyzdys	33
Pav. 10 Prognozavimo sistemos vartotojo sąsajos navigavimo planas.....	34
Pav. 11 Duomenų pasirinkimo lango prototipas.....	34
Pav. 12 Parametrų pasirinkimo lango prototipas	35
Pav. 13 Prognozavimo vykdymo veiklos diagrama dabartinėje aplinkoje	36
Pav. 14 Kompiuterizuotos parametrų valdymo sistemos veiklos diagrama	37
Pav. 15 Duomenų importavimo ir vitrinės kūrimo proceso veiklos diagrama.....	38
Pav. 16 Prognozavimo proceso veiklos diagrama	40
Pav. 17 Sistemos loginės architektūros diagrama.....	41
Pav. 18 „Ištrinti parametą“ PA projekto klasių diagrama	42
Pav. 19 „Nustatyti parametro keitimo teises“ PA projekto klasių diagrama	42
Pav. 20 „Priskirti parametą vartotojams“ PA projekto klasių diagrama	43
Pav. 21 „Sukurti naują parametą“ PA projekto klasių diagrama.....	43
Pav. 22 „Sukurti prognozavimo ataskaitą“ PA projekto klasių diagrama	44
Pav. 23 „Suvesti parametro reikšmes“ PA projekto klasių diagrama.....	44
Pav. 24 „Ištrinti parametą“ PA sekų diagrama	45
Pav. 25 „Nustatyti parametro keitimo teises“ PA sekų diagrama.....	45
Pav. 26 „Pasirinkti prognozavimo duomenų turinį“ PA sekų diagrama.....	46
Pav. 27 „Pasirinkti prognozavimo metodą“ PA sekų diagrama	46
Pav. 28 „Pasirinkti prognozavimo parametrus“ PA sekų diagrama	47
Pav. 29 „Priskirti parametą vartotojams“ PA sekų diagrama.....	47
Pav. 30 „Sukurti naują parametą“ PA sekų diagrama	48
Pav. 31 „Sukurti prognozavimo ataskaitą“ PA sekų diagrama.....	48
Pav. 32 „Suvesti parametro reikšmes“ PA sekų diagrama	49
Pav. 33 „Paimti parametrinius duomenis“ PA sekų diagrama.....	49
Pav. 34 „Importuoti parametrinius duomenis“ PA sekų diagrama.....	50
Pav. 35 Prognozavimo sistemos duomenų bazės diagrama.....	50
Pav. 36 Ataskaitos generavimo posistemės klasių diagrama.....	51
Pav. 37 Saugyklų posistemės klasių diagrama	52
Pav. 38 Prognozavimo sistemos realizacijos diagrama	53
Pav. 39 Prognozavimo sistemos diegimo diagrama	53
Pav. 40 Patekimas į prognozavimo langą	54
Pav. 41 Pagrindinis prognozavimo langas	55
Pav. 42 Prognozavimo antro žingsnio langas	56
Pav. 43 Prognozavimo antras žingsnis pasirinkus prognozavimą su parametrais	56
Pav. 44 Prognozavimo trečio žingsnio langas	57
Pav. 45 Prognozavimo rezultato grafikas	58
Pav. 46 Prognozavimo rezultato eksportavimo langas	58
Pav. 47 Eksperimentinės duomenų saugyklos dalis	61
Pav. 48 Metaduomenų lentelės turinys	62
Pav. 49 Tiksliniai vaistų rinkos duomenys 2015 metais.....	63

Pav. 50 Tiesinės regresijos rezultatų apžvalga	64
Pav. 51 Tiesinės regresijos paklaidų įverčiai	64
Pav. 52 Tiesinės regresijos prognozės ir realių pardavimų palyginimas	65
Pav. 53 Laiko eilučių prognozės ir realių pardavimų palyginimas. <i>Auto.Arima()</i> modelis.....	66
Pav. 54 Laiko eilučių paklaidų įverčiai. <i>Auto.Arima()</i> modelis.....	66
Pav. 55 Laiko eilučių prognozės ir realių pardavimų palyginimas. <i>ETS</i> modelis	66
Pav. 56 Laiko eilučių paklaidų įverčiai. <i>ETS</i> modelis	67
Pav. 57 Laiko eilučių prognozės ir realių pardavimų palyginimas. <i>BATS</i> modelis.....	67
Pav. 58 Laiko eilučių paklaidų įverčiai. <i>BATS</i> modelis.....	67
Pav. 59 Laiko eilučių prognozės ir realių pardavimų palyginimas. <i>HoltWinters</i> modelis	68
Pav. 60 Laiko eilučių paklaidų įverčiai. <i>HoltWinters</i> modelis	68
Pav. 61 Neuroninio tinklo prognozės ir realių padavimų palyginimas (1).....	69
Pav. 62 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (2)	70
Pav. 63 Neuroninio tinklo paklaidų įverčiai (2).....	70
Pav. 64 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (3)	70
Pav. 65 Neuroninio tinklo paklaidų įverčiai (3).....	71
Pav. 66 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (4)	71
Pav. 67 Neuroninio tinklo paklaidų įverčiai (4).....	71
Pav. 68 Paklaidų įverčių suvestinė.....	72

IVADAS

Darbas kuriamas „Informacinių sistemų inžinerijos“ programai. Šis darbas aptaria ir panaudoja veiklos intelektikos priemones ir gerąsias praktikas pardavimų prognozavimo srityje. Tyrimo rezultatai orientuoti į įmonių analitikus ir sprendimų priėmėjus, jiems suteikiant galimybę numatyti ateinančio laikotarpio rinkos pokyčius ir atitinkamai pagal tai pakreipti savo organizacijos plėtrą teisinga kryptimi.

Darbo problematika ir aktualumas

Norint atlikti tinkamą prognozavimo procesą visų pirma reikia pasirūpinti tikslinga istorinių duomenų struktūra. Istoriniai duomenys reikalingi norint iš jų išrinkti esamas tendencijas ir toliau jas panaudoti ateities prognozavimams. Pirmoji problema iškyla dėl duomenų saugyklos sudėtingumo ir jos nuolatinio plėtimosi. Kiekvienos įmonės saugyklos architektūra gali būti savitai unikali ir dėl to yra sudėtinga sukurti ataskaitų generavimo įrankį veikiantį korektiškai ir nepriklausantį nuo konkrečios duomenų saugyklos specifikos.

Tačiau, norint vykdyti prognozavimo procesą su duomenų saugykla, išspręsti duomenų saugojimo ir kaupimo problemą neužtenka. Šiuo metu galima rasti nemažai matematinių modelių, kurie naudojami prognozuoti ateities informaciją, tačiau jų veikimas ir panaudojimas yra suprantamas tik matematinės analizės specialistams. Dėl šios priežasties pardavimų prognozavimas vyksta neefektyviai arba jo įmonės tiesiog nevykdo. Taip pat, turint didelį kiekį istorinių duomenų, prognozavimo proceso atsisakoma dėl ilgai užtrunkančio parengimo, kuris netenkina verslo atstovų. Ypač daug laiko sąnaudų reikalauja parametrinių duomenų administravimas, jų kaupimas ir atnaujinimas.

Darbo tikslas ir uždaviniai

Norint išspręsti nurodytas problemas, yra užsibrėžtas darbo pagrindinis tikslas – pagerinti ir palengvinti pardavimų prognozavimą, taikant veiklos intelektikos priemones. Palengvinimą galima suprasti kaip prognozavimo modelio sudarymo paprastumą, neturint specialių matematinių ir prognozavimo žinių. Supaprastinti modelio sudarymo procesą padeda veiklos intelektikos gerosios praktikos, duomenų saugyklos kūrimo metodikos ir jų pritaikymas. Proceso rezultatų tikslumo gerinimą lemia esamų prognozavimo modelių sudarymo analizė, jų eksperimentas su realiais duomenimis ir atrastų tendencijų bei išvadų panaudojimas dalykinėje srityje.

Norint įgyvendinti darbo tikslą išskiriami šie darbo uždaviniai:

1. Išanalizuoti svarbiausias duomenų saugojimo metodikas.
2. Išanalizuoti esamus prognozavimo metodus.
3. Aprašyti esamas prognozavimo aplinkas ir esminius įrankių paketus.
4. Specifikuoti ir sukurti prognozavimo modulį.
5. Sukurti duomenų saugyklos metamodelį, kuris leistų automatizuoti dalį prognozavimo proceso.
6. Eksperimentiškai ištirti sukurtą įrankį su farmacijos įmonių pardavimų duomenimis.
7. Apibendrinti tyrimo rezultatus, pateikti rekomendacijas ir išvadas.

Darbo rezultatai ir jų svarba

Galutinis problemos sprendimas – efektyvus pardavimų prognozavimo įrankis. Įrankis pateiktas farmacijos įmonių analitikams ir sulaukti pirmieji teigiami atsiliepimai. Ateityje darbo išvadas ir aprašytas rekomendacijas bus galima pateikti ir kitų dalykinių sričių įmonėms, kurioms reikalingas pardavimų prognozavimas. Galima išskirti tokius sukurto sprendimo privalumus, lyginant su kitais prognozavimo paketais:

- Sukurtas ir panaudotas duomenų saugyklos metamodelis leidžia pilną ataskaitos generavimo ir parametrinių duomenų surišimo automatizavimą.
- Prognozavimo proceso paprastumas, t. y. norint gauti ataskaitą nereikia turėti specialių matematinės analizės ar programavimo žinių. Tai didelis privalumas vartotojams, kurie šių žinių neturi ir dėl to negali pasinaudoti matematinės analizės įrankiais, tokiais kaip „Excel“ ar „R“.
- Automatizuotas bendrinių parametrinių duomenų surinkimas ir atnaujinimas. Vartotojams belieka surinkti ir suvesti jų asmeninius parametrus, pvz.: įmonės biudžeto pokyčius, reklamos išlaidas ir pan.
- Automatizuotas reikšminių parametrų atrinkimas priklausomai nuo pasirinkto ataskaitos turinio, panaudojant atgalinę parametrų atranką (angl. *backward selection*).
- Esant poreikiui, labiau pažengusiems vartotojams su gilesnėmis žiniomis suteikiama galimybė konfigūruoti prognozavimo nustatymus ir generavimo procesą pritaikyti pagal savo reikmes. Tai didelis privalumas lyginant su verslo valdymo paketais, kurie dažniausiai pilnai veikia juodosios dėžės principu ir vartotojas neturi galimybių daryti įtaką prognozavimo procesui.

Darbo struktūra

Visas darbas suskirstytas į šešis skyrius. Pirmajame skyriuje aprašoma probleminės srities analizė, aptariamos duomenų saugojimo metodikos, prognozavimo modeliai ir esamų sprendimų analizė. Antras skyrius skirtas kuriamo įrankio reikalavimų specifikacijai ir projektui, jame pateikti funkciniai ir nefunkciniai reikalavimai, keletas specifikacijos modelių ir formalus sprendimo aprašas. Trečiajame skyriuje aprašomas įrankio realizacijos projektas. Ketvirtas skyrius skirtas sprendimo realizacijos aprašui – sprendimo veikimo instrukcijai, testavimo planui ir įgyvendinimui aprašyti. Penktajame skyriuje pateiktas eksperimentinis sprendimo tyrimas, skyrius suskirstytas į eksperimento dalykinės srities aprašą, eksperimento planą, rezultatus ir iš jų gautas sprendimo taikymo rekomendacijas. Paskutiniame šeštajame skyriuje įvardintos viso darbo išvados.

1. PROBLEMINĖS SRITIES ANALIZĖ

1.1. Analizės tikslas

Norint giliau pažvelgti į tiriamojo darbo temą ir efektyviau išspręsti pardavimų duomenų prognozavimo problemas, šiame skyriuje atliekama probleminės srities analizė. Analizės tikslas – palengvinti reikalingo sprendimo apibrėžimą ir suvokimą. Probleminės srities analizės skyriuje nagrinėjami jau esami technologiniai sprendimai, šiai dienai žinomos srities teorinės žinios ir kitų autorių literatūra.

1.2. Tyrimo objektas, sritis ir problema

Tyrimo objektas – pardavimų prognozavimas.

Tyrimo sritis – prognozavimo modeliai ir jų taikymas pardavimų srityje.

Tyrimo problema – yra daug modelių, nepritaikytų ir neištirtų specifiniai sričiai, neaišku kada kurį iš jų rekomenduojama naudoti ir kuris modelis gali suteikti didžiausią tikslumą naudojamoje srityje. Taip pat reikia pasirinkti efektyviu duomenų laikymu, tiek istorinių duomenų, tiek parametrinių duomenų, kurie naudojami ateičiai prognozuoti. Duomenų laikymo architektūra ir duomenų metamodelis turi leisti automatizuoti ataskaitų generavimo procesą ir užtikrinti lengvą duomenų migraciją, esant duomenų saugyklos pokyčiams.

1.3. Tyrimo objekto (veiklos proceso, modelio, metodo, darinio, sistemos ar jos dalies ir pan.) analizė

Pardavimų prognozavimas jau nuo seno yra sritis, kuri domina verslo kompanijas. Tikimasi, jog turint galimybę nuspėti ateities pardavimų kitimą, bus galima pasiruošti būsimoms krizėms, ar iš anksto nukreipti organizacijos veiklą kryptimi, kuri užtikrintų didžiausią pelną. Kaip teigiama šaltiniuose [1] ir [2], organizacijos lygmenyje pardavimų prognozavimas yra svarbus parametras dideliame kiekiui sprendimų priėmimų tokiose srityse kaip gamyba, vadyba, finansai.

Prognozavimas atliekamas panaudojant skirtingus prognozavimo modelius. Jų šiai dienai yra daugybė, didelė problema tampa sprendimas, kurį iš jų naudoti ir kuris modelis užtikrintų didžiausią tikslumą.

Prognozavimo modeliai paprastai skirstomi į dvi grupes: modeliai mokomi su priežiūra ir be priežiūros [3]. Mokymas su priežiūra pasižymi modelio sudarymu panaudojant istorinius duomenis: surišami aiškinamieji kintamieji su prognozuojamais kintamaisiais [4]. Tokiu būdu modelis atranda esamas tendencijas ir jas panaudodamas gali nustatyti prognozuojamus kintamuosius. Pavyzdžiui, jeigu turime patikimus duomenis apie oro užterštumo kitimą bei gyventojų sergamumą tiksliais laiko momentais, galime bandyti ieškoti tendencijos kaip užterštumas lemia žmonių sergamumą. Antroji grupė yra vadinama mokymu be priežiūros. Šis metodas bando ištraukti išvadas iš duomenų, kuriuos sudaro tik įėjimo duomenys be susietų rezultatų [4].

Mokymas su priežiūra

Pagrindinis duomenų išrinkimo būdų skirtumas yra mokymo su priežiūra ir mokymo be priežiūros metodų. Mokymo su priežiūra algoritmai naudojami klasifikavimo ir prognozavimo srityje. Norint naudoti šį algoritmą, mes turime turėti duomenis, kuriuose norimas sužinoti rezultatas (pvz., prekė buvo parduota ar ne) yra žinomas [5]. Tokie duomenys yra vadinami mokymo duomenimis (angl. *Training data*). Pasirinktas duomenis naudojantis mokymo modelis yra

išmokomas, tai dar kitaip vadinama modelio „treniravimu“. Tai atlikus modelis sužino apie esančias tendencijas tarp aiškinamųjų kintamųjų ir prognozuojamo kintamojo. Kai jau turimas ištreniruotas modelis, jam galima paduoti testinius duomenis ir gauti prognozuojamų kintamųjų reikšmes. Vienas iš paprasčiausių ir dažniausiai naudojamų mokymo su priežiūra modelių yra linijinė regresija. Linijinė regresija bando rasti ryšį tarp dviejų kintamųjų ir jų ryšį aprašyti linijine formule [5].

Mokymas be priežiūros

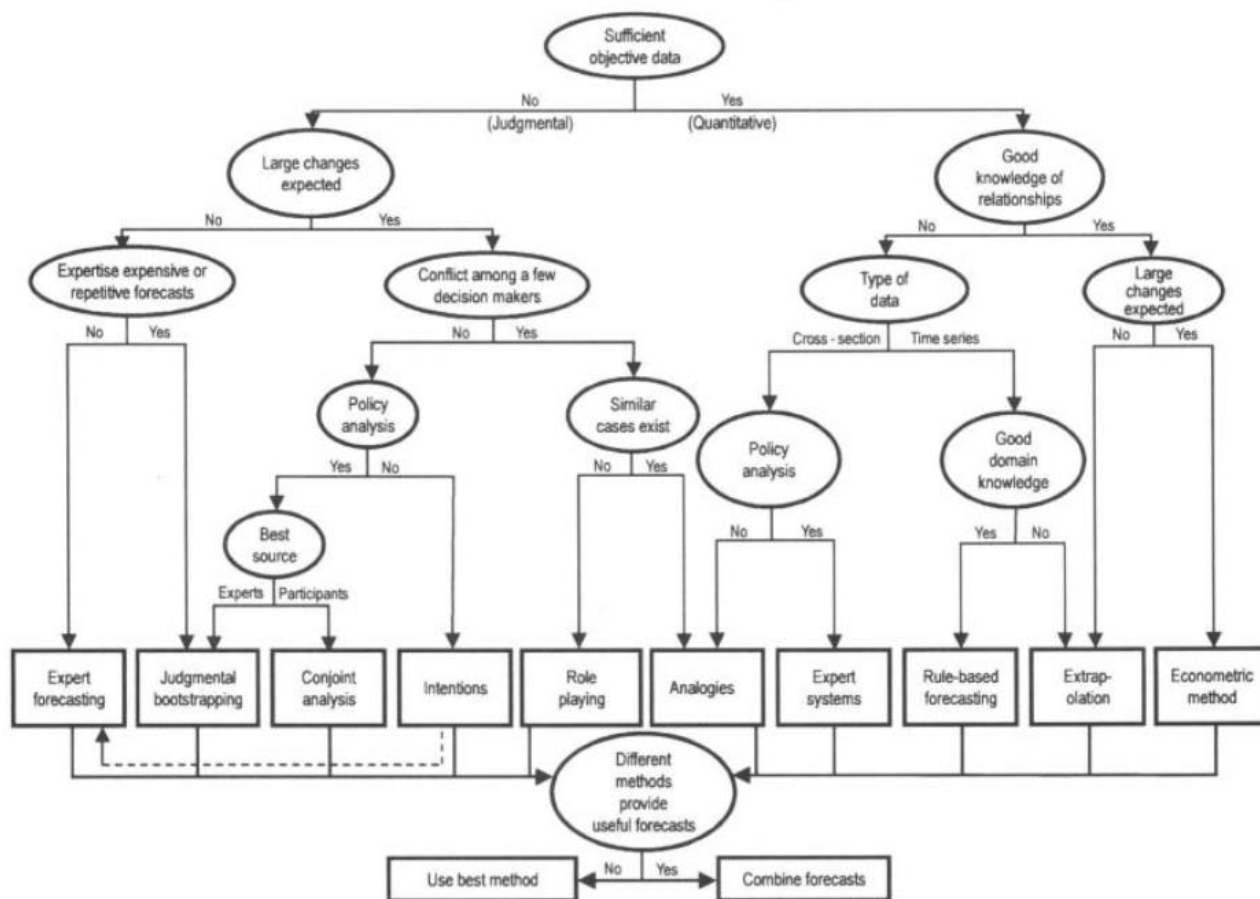
Lyginant su mokymu su priežiūra, mokymas be priežiūros bando rasti paslėptą struktūrą nesužymėtuose duomenimis. T. y. įeigos duomenų rinkinys yra surenkamas, analizuojamas ir traktuojamas kaip rinkinys atsitiktinių kintamųjų [6]. Kadangi įeigos duomenys nėra sužymėti, nėra galimybės gauti klaidos ar teisingo veikimo signalo, kurie įvertintų potencialų sprendimą.

1.5 skyriuje bus nagrinėjami atskiri prognozavimo metodai. Didesnis dėmesys bus skiriamas metodams priskiriamiems mokymui be priežiūros grupei. Pagrindinė to priežastis yra tai, kad naudojant mokymą be priežiūros negalima įvertinti ar gautas prognozavimo rezultatas yra teisingas. Didžiausias mokymo be priežiūros privalumas yra tai, kad nereikia turėti nuspėjamojo kintamojo reikšmių mokant modelį, tačiau šio darbo rezultato naudotojai naudos sužymėtus istorinius duomenis (pardavimų duomenys), todėl nebus galima išnaudoti mokymo be priežiūros privalumų.

Svarbus klausimas lieka, kaip teisingiau pasirinkti prognozavimo metodą: parametrizuotą ar ne. Šaltinyje [7] išskirti svarbiausi punktai:

- Jeigu turima pakankamai duomenų, kiekybinis prognozavimas tinka geriau už kokybinį prognozavimą.
- Kai numatomi dideli pokyčiai parametrizuoti metodai tinka geriau už neparimetrizuotus.
- Neradus svarių argumentų modelio pasirinkime, teikti pirmenybę paprastesniam ir pigesniam metodui (t. y. neparimetrizuotam).
- Dažniausiai brangiausias ir tiksliausias kelias yra panaudoti abu metodus ir naudoti jų vidurkį.

Taip pat šaltinyje [7] pateikiamas sprendimų medis leidžiantis nuspręsti kurį prognozavimo modelį rinktis (Pav. 1). Nors nemaža dalis galimų modelių šiame darbe neanalizuojami, visgi galima susidaryti nuomonę, kokius klausimus reikia iškelti ir apsvarstyti renkantis prognozavimo kelią.



Pav. 1 Siūlomas prognozavimo metodo pasirinkimo procesas [7]

Duomenų saugykla (angl. *Data warehouse*)

Duomenų saugykla yra saugykla iš subjektyviai atrinktos ir pritaikytos informacijos, kuri sėkmingai gali atsakyti į sudėtingas statistines ir analitines užklausas [8]. Duomenų saugykla paprastai yra visos organizacijos sprendimų priėmimo ašis, jis integruoja visai organizacijai bendrus istorinius duomenis. Vis didėjantis duomenų kiekis priverčia ieškoti sprendimų kaip organizuotai ir efektyviai juos kaupti, klasifikuoti, apdoroti ir šią informaciją panaudoti žinioms apie organizacijos padėį gerinimui.

Duomenys gali būti skirstomi į tris kategorijas, kaip rašoma šaltinyje [8]: istoriniai duomenys, išvestiniai duomenys ir metaduomenys. Duomenų saugykla paprastai laiko kelių praėjusių metų duomenis, tačiau tai priklauso nuo poreikių ir techninių galimybių, pavyzdžiui, prieinamos disko talpos. Istoriniai duomenys gali būti gaunami iš duomenų bazių archyvo, tačiau labai dažnai jie gaunami iš išorinių šaltinių, kurių duomenų saugyklos projektuotojai negali numatyti ar koreguoti. Tokiu atveju iškyla įvairios duomenų suderinamumo ir integracijos problemos.

Išvestiniai duomenys yra generuojami pagal atitinkamą algoritmą iš istorinių duomenų. Šie duomenys nebūna gaunami iš šaltinio, jie yra kaip dalis duomenų palaikymo operacijų arba būna generuojami dinamiškai užklauso vykdymo metu [9]. Išvestinių duomenų pavyzdys galėtų būti toks: turime lentelę, kurioje saugome produkto pirkimo kiekį ir produkto kainą, išvestiniai duomenys tokioje situacijoje būtų visa pirkimo vertė, gaunama sudauginant kiekį su vieno vieneto kaina. Visada kyla sunkus klausimas ką pasirinkti: išvestinių duomenų formavimą užklauso metu ar jų laikymą duomenų saugykloje. Pagrindinis išvestinių duomenų laikymo saugykloje privalumas yra greitesnis užklauso vykdymas ir užtikrinimas, kad išvestiniai duomenys bus skaičiuojami vieną kartą, t. y. bus sumažinama tikimybė jų blogam suskaičiavimui. Taip pat verta paminėti, kad

išvestinių duomenų saugojimas saugykloje gali privesti prie didžiulio ir sunkiai kontroliuojamo duomenų bazės išsiplėtimo [9]. Išvestinių duomenų generavimas užklauskos metu leidžia neapkrauti duomenų bazės, toks būdas geriausiai tinka, kai vieno įrašo išvestiniai duomenys gali kisti atsiradus naujiems įrašams [9], pvz.: turime lentelę su testo dalyvių identifikatoriais ir jų užimama pozicija turnyrinėje lentelėje, jų pozicija gali kisti atsiradus naujiems dalyviams, todėl būtų tikslinga šią informaciją formuoti dinamiškai užklauskos vykdymo metu.

Metaduomenys yra duomenys, kurie apibūdina kitus duomenis. Jie yra labai svarbūs užtikrinant teisingą ir produktyvų duomenų saugyklos naudojimą visą jo gyvavimo ciklą [10]. Metaduomenis galima suskirstyti į dvi grupes [11]: vidiniai (angl. *back-room*) ir išoriniai (angl. *front-room*). Vidiniai metaduomenys yra svarbesni ir jie aprašo duomenų valymo, užkrovimo, gavybos procesus. Tuo tarpu išoriniai metaduomenys yra labiau skirti paprastam vartotojui, jie gali aprašyti vartotojų privilegijas, duomenų saugyklos struktūrą, ataskaitų gavimo instrukcijas [11] ir pan.

Duomenų vitrinos

Duomenų vitrina yra atskira duomenų bazė ar duomenų bazių aibė, kiekviena iš jų turinti specifinį kontekstą [12]. Duomenų vitrina yra kuriama kai duomenų saugykla negali pateikti informacijos reikiamu būdu informacijos naudotojams.

Duomenų vitrinos būna fiziškai ar logiškai atskirtos nuo duomenų saugyklos, iš kurio yra gaunami duomenys. Duomenų vitrina gali būti įgyvendinama keliais būdais [12] :

- Vitrina kuriama fiziškai atskiriant dalį duomenų saugyklos duomenų bazių aibės ir patalpinama skirtingoje platformoje. Toks metodas užtikrina didžiausią galimą vitrinos izoliaciją. Taip pat įgalinama galimybė tvarkingai ir struktūrizuotai keisti duomenis, priklausomai nuo vitrinos naudotojų poreikio, jų perkėlimo metu. Didžiausias šio būdo trūkumas yra įgyvendinimo kaina. Reikia pasirūpinti atskira platforma, kuri reikalinga vitrinai patalpinti, duomenų perkėlimo programine įranga ir platformos palaikymu.
- Antras metodas yra sukurti aibę vaizdinių, kurie paima duomenis iš duomenų saugyklos. Tokiu atveju nereikalinga atskira platforma, tačiau yra paaukojama galimybė turėti nepriklausomą vitriną ir tenka pasirūpinti resursų paskirstymu tarp vitrinos ir saugyklos. Šis metodas taip pat turi galimybę keisti duomenis aprašant vaizdinį SQL kalba, tačiau galimybė yra apribota palyginus su vitrinos laikymu skirtingoje platformoje.

Duomenų saugyklos kūrimo architektūra

Prieš pradėdant kurti duomenų saugyklą, reikia apsispręsti kokią architektūrą naudoti. Šiuo metu laiko patvirtintos kūrimo architektūros yra dviejų autorių Ralph'o Kimballo (kūrimas nuo apačios į viršų) ir Billo Inmono (kūrimas nuo viršaus į apačią) [13].

Kūrimas nuo apačios į viršų pasižymi tuo, jog pirma sukuriama duomenų vitrinos leidžiančios analizavimo ir ataskaitų darymo galimybes specializuotiems įmonės procesams. Vėliau šios vitrinos sujungiamos sukuriant išsamią duomenų saugyklą.

Kūrimas nuo viršaus į apačią visų pirma apibrėžia duomenų saugyklos kūrimą panaudojant tradicines duomenų bazės technikas. Verta paminėti, jog šis būdas naudojamas kartu su trečios normalizuotos formos duomenų struktūra [13]. Skirtingai nuo anksčiau aprašytos architektūros, ši architektūra teigia, jog duomenų vitrinos turėtų būti kuriamos vėlesniuose etapuose, o duomenų saugykla laiko informaciją sujungiančią visas organizacijos sritis ir jį galima laikyti visos organizacijos duomenų šerdimi.

Nors jau ilgą laiką vyksta diskusijos, kuris iš dviejų kūrimo būdų yra patikimesnis, dažniausiai organizacijos pasirenka vieną iš architektūrų remdamosi darbuotojų patirtimi, daromomis prielaidomis, konsultantų ar pardavėjų rekomendacijomis [13].

1.4. Prognozavimo metodų analizė

Logistinė regresija

Logistinė regresija naudojama, kai spėjamas kintamasis gali įgyti tik dvi reikšmes. Dažniausiai tai būna atsakymas „taip“ arba „ne“, arba matematikoje dažniau naudojamas 1 arba 0. Tai pagrindinis skirtumas tarp logistinės regresijos ir linijinės regresijos metodų. Logistinės regresijos naudojimo pavyzdys galėtų būti toks: žinome žmogaus amžių, turimų vaikų kiekį ir pajamas ir reikia rasti ryšį tarp šių duomenų ir žmogaus patikimumo suteikiant kreditą. Modelio nuspėjamas kintamasis šiuo atveju yra žmogaus patikimumas ir jį galima išreikšti dvejomis būsenomis: 1 – t. y. žmogus patikimas ir jam galima suteikti kreditą, 0 – žmogus nepatikimas ir jam suteikti kreditą nerekomenduojama.

Bandant naudoti logistinę regresiją gali iškilti situacija, kai kintamojo nepavyksta įterpti į dviejų reikšmių (1 arba 0) rėmus. Tokiu atveju reikia naudoti dizaino arba dar kitaip vadinamus fiktyvius (angl. *dummy*) kintamuosius [14]. Pavyzdys galėtų būti duomenys apie žmogaus rasę, jeigu tarkim galimi variantai yra baltaodis, juodaodis ir „kita“ rasė. Tokiu atveju reikia galimas reikšmes pakeisti į matematinę išraišką ir tai galima padaryti naudojant fiktyvius kintamuosius k_1 ir k_2 . Kintamasis k_1 rodytų, ar rasė yra juodaodis, k_2 rodytų ar rasė yra „kita“. Visas pavyzdyje esančias rases galima aprašyti lentelė 1.1 [14]. Tai atlikus jau galima bandyti kurti matematinę modelį ir gauti matematinę lygtį.

1.1. Žmogaus rasės kintamųjų apibrėžimas dvejomis būsenomis

Rasė	K1	K2
Baltaodis	0	0
Juodaodis	1	0
Kita	0	1

Dirbtiniai neuroniniai tinklai

Dirbtiniai neuroniniai tinklai (angl. *Artificial Neural Network*) yra informacijos apdorojimo metodika įkvėpta žinių apie biologinę nervų sistemą. Tinklas yra sudaromas iš atskirų ir stipriai sujungtų elementų (neuronų), kurie dirba kartu siųsdami vieni kitiems signalus ir tokiu būdu bendrai sprenddami iškilusią problemą. Pagrindiniai neuroninių tinklų panaudojimo privalumai: [15]

- Automatinis mokymas iš pavyzdžių.
- Naudotojui nereikia daryti prielaidų apie sąryšį tarp įvesties ir įvesties duomenų
- Greitas mokymas (jeigu tinklas turi mažiau nei 50 neuronų)

Pagrindiniai neuroninių tinklų trūkumai:

- Neuroninis tinklas veikia kaip „juodoji dėžė“, t. y. sunku interpretuoti neuroninio tinklo sprendimą ir veikimą.
- Norint sudaryti gerą neuroninį tinklą vartotojui reikia apibrėžti pradinius tinklo duomenis [16]: tinklo mokymo gyli, paslėptų neuronų kiekį kiekviename tinklo

sluoksnyje ir pan. Šiuos parametrus gali būti sunku nustatyti ir tikslesnės jų reikšmės gali paaiškėti tik po kelių modelio sudarymo bandymų.

Iš šių privalumų ir trūkumų galima daryti išvadą, kad neuroninių tinklų metodas labiausiai tinka situacijose, kai turimas mokymo duomenų rinkinys, tačiau sunku nustatyti ryšį tarp įvesties ir išvesties duomenų [15].

Tiesinė regresija

Tiesinė regresija yra vienas labiausiai paplitusių ir giliausiai išnagrinėtų matematinės analizės metodų [17]. Tiesinė regresija modeliuoja ryšį tarp priklausomo kintamojo ir vieno ar kelių aiškinamųjų kintamųjų. Situacija [18], kai naudojamas vienas aiškinamasis kintamasis, vadinama paprastoji tiesinė regresija, daugiau nei vieno kintamojo naudojimo procesas vadinamas daugialype tiesine regresija. Šis metodas savo principu panašus į logistinę regresiją, pagrindinis skirtumas yra tai, kad priklausomas (mūsų atveju spėjamas) kintamasis gali įgyti tęstines skaitines reikšmes. Pagrindinė užduotis, norint tinkamai panaudoti tiesinės regresijos modelį pardavimų prognozavimui, yra surinkti parametrus lemiančius priklausomą kintamąjį ir įvertinti jų stiprumą. Po bandymo sudaryti modelį dažniausiai pamatoma, jog ne visi parametrai yra reikšmingi, kaip buvo manoma iš anksto, ir modelis vėl perdaromas jau atmetus nereikšmingus parametrus. Tai kartojama, kol gautas modelis turi tik reikšmingus parametrus, jeigu tokių parametrų nerandama, reikia iš naujo pergaltvoti modelio ir probleminės srities logiškumą. Modelio reikšmingų parametrų koeficientų radimas vyksta pasitelkiant istorinius duomenis ir ieškant tendencijų, šis procesas paprastai yra automatizuojamas naudojant matematinės analizės programinę įrangą, pvz.: *R*, *MatLab* ir pan.

Atlikę tiesinės regresijos modelio sudarymą, gauname matematinę lygtį, lygties pavyzdys pateiktas žemiau:

$$y = \beta_1x_1 + \beta_2x_2 + \dots + \varepsilon$$

Koeficientai β mums būna žinomi po modelio sudarymo ir jų dydis parodo kuria kryptimi atitinkamo parametro reikšmė veikia priklausomą kintamąjį y . Reikėtų atkreipti dėmesį ir į lygtyje esantį ε kintamąjį. Jis yra vadinamas atsitiktine paklaida arba triukšmu ir parodo kaip stipriai ar silpnai aiškinamieji kintamieji apibrėžia priklausomą kintamą [19].

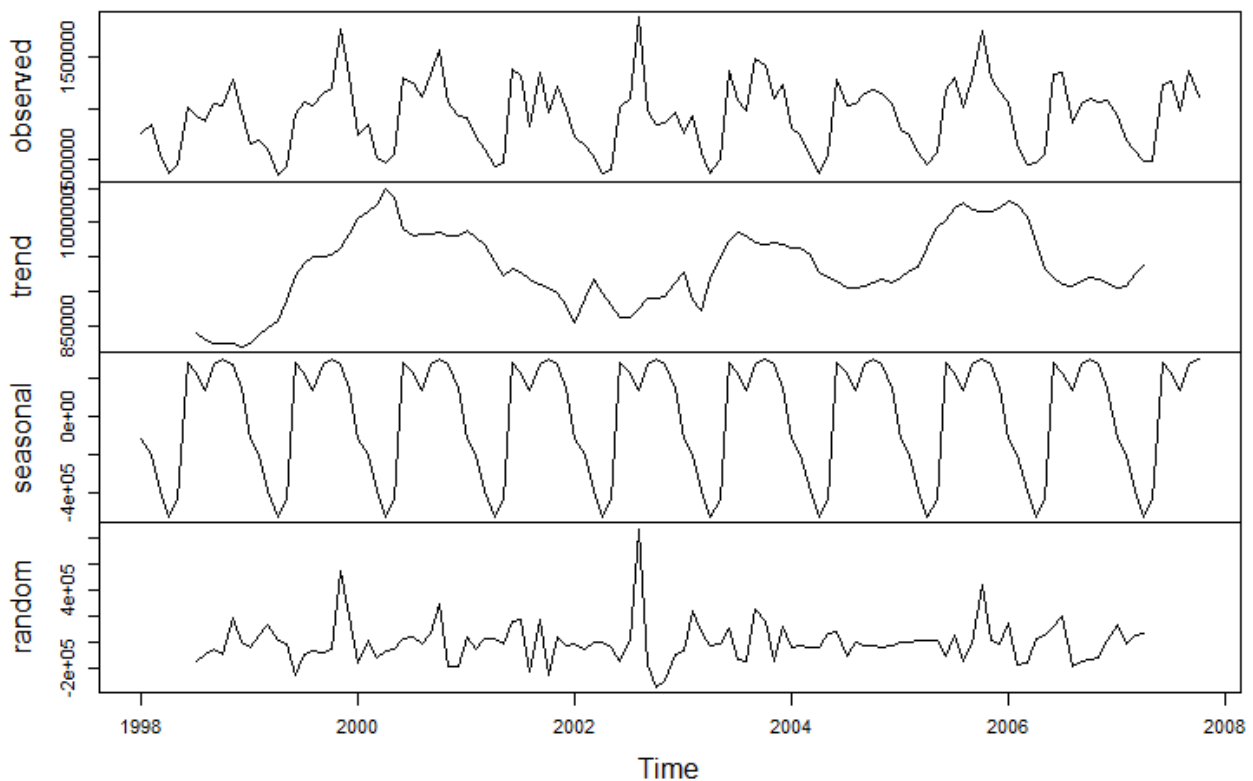
Laiko eilutės

Laiko eilutės yra periodiškai surinktas duomenų rinkinys. Paprastai naudojamas laikotarpis yra mėnuo, tačiau dažniau kintamiems duomenims galima naudoti ir dienos ar valandos duomenų surinkimo periodiškumą. Vienas geriausiai suprantamų laiko eilutės pavyzdžių yra akcijų rinkos kitimas ar pardavimų analizė pamėnesiui. Laiko eilučių analizė – tai ekonometrinis tyrimo metodas, kai naudojami tik vieno kintamojo įvairių laikotarpių duomenys [20]. Tokia analizė leidžia išsiaiškinti pagrindines nagrinėjamos laiko eilutės tendencijas, suprasti nagrinėjamo reiškinio praėjusio laikotarpio kitimo dėsningumus ir, žinoma, prognozuoti kitimą ateityje.

Tradiciniai laiko eilučių analizės metodai apima laiko eilutės išskirstymą į tris dedamąsias: krypties (angl. *trend*), sezoniškumo (ciklinius) ir atsitiktinius svyravimus [21]. Krypties dedamoji gali būti apibūdinama kaip „ilgalaikis vidurkio kitimas“. Pagrindinė problema yra nustatyti ką reiškia „ilgalaikis kitimas“: pavyzdžiui, turint dešimties metų duomenis apie klimato kaitą, krypties matymas gali būti klaidingas, nes tikslinga kryptis gali išryškėti tik per žymiai ilgesnį laikotarpį – apie 50 metų. Sezoniškumo dedamoji – tai reguliarūs cikliniai svyravimai, vykstantys per trumpesnę nei metų laikotarpį [20]. Atsitiktiniai svyravimai – tai nepriklausoma, atsitiktinai laike kintanti dedamoji, dar kitaip vadinama triukšmu. Iliustracijoje Pav. 2 parodytas pavyzdinis laiko eilutės duomenų išskirstymas į tris minėtas dedamąsias. Diagramoms braižyti panaudoti kvėpavimo sistemą veikiančių vaistų pardavimai. Dekomponavimas atliktas naudojant *R* matematinės analizės

paketą. Viršutinis blokas rodo turimus realius duomenis. Iš sezoniškumo juostos matyti ryškus ciklinis kitimas priklausomas nuo metų laiko. Kryptis (angl. *trend*) nėra tiesiška ir yra gan chaotiškai kintanti, norint detaliau iširti šią laiko eilutę reikėtų išanalizuoti papildomus duomenis, kurie galėjo daryti įtaką krypties kitimui tam tikrais metais.

Vienas didžiausių laiko eilučių privalumų, lyginant su kitais mokymo su priežiūra prognozavimo metodais, yra tai, kad prognozavimui nebūtini papildomi parametrai, t. y. prognozavimas gali vykti tik iš istorinių duomenų tendencijų. Dėl šios priežasties palengvėja ir pagreitėja visas prognozavimo procesas, jis yra labiau suprantamas paprastam vartotojui, neturinčiam didelių analizės ir prognozavimo žinių. Taip pat laiko eilučių prognozavimas tinka situacijose, kai turimi tikslūs istoriniai duomenys, tačiau srities logika nėra aiški ir papildomų parametru įvedimas gali būti tiesiog neįmanomas. Tiesa, visi šie išvardinti privalumai tam tikrose situacijose gali pavirsti ir didžiausiais laiko eilučių trūkumais: esant parametru, kurio pasikeitimas ateityje gali daryti didelę įtaką duomenis (pavyzdžiui, didelę rinkos dalį užimanti kompanija išleidžia naują produktų seriją), prognozė tik iš istorinių duomenų bus netiksli. Taip pat laiko eilutės netinka analizei, kaip ateityje aiškinamojo parametro pakeitimas veiks prognozuojamą parametru (pavyzdžiui, įmonė nori sužinoti kaip pasikeis jų užimama rinka, jeigu bus padidinta produkto kaina).



Pav. 2 Laiko eilutės dedamųjų išskyrimas

Šiame skyriuje buvo apžvelgti prognozavimo metodai: logistinė regresija, dirbtiniai neuroniniai tinklai, laiko eilutės ir tiesinė regresija. Iš atliktos analizės galima daryti išvadą, jog probleminiai sričiai labiau tiktų neuroninių tinklų ir tiesinės regresijos modeliai dėl galimybių rezultate gauti teštingas skaitines reikšmes. Logistinė regresija tinkamiausia, kai rezultata galima išreikšti dvejomis būsenomis, tačiau šio darbo tikslas yra pateikti kuo tikslesnes prognozavimo skaitine reikšmes ir dėl to logistinė regresija nėra patrauklus sprendimo būdas probleminiai sričiai. Nereikėtų atmesti ir laiko eilučių prognozavimo įgyvendinimo, tai tiktų situacijose, kai išoriniai parametrai neturimi arba nėra aktualūs.

1.5. Tyrimo objekto naudotojų analizė

Pardavimų prognozavimas gali būti naudingas beveik kiekvienai pelno siekiančiai organizacijai. Pagrindinis privalumas pagrinde slypi galimybėje organizacijai suteikti informaciją ko tikėtis per ateinančius mėnesius. Pardavimų prognozavimas identifikuoja planuojamus rinkos pakilimus ir nuopusius prieš kompanijai nustatant biudžetą ir marketingo kryptis, tai leidžia optimizuoti organizacijos plėtros planus.

1.5.1. Naudotojų tipai

Pardavimų prognozavimo rezultatai labiausiai aktualūs įmonių vadovams ir analitikams. Įmonių rinkinys gali būti labai platus, jis apima visas pelno siekiančias organizacijas, kurių viena iš veiklos sričių yra pardavimų valdymas.

1.5.2. Naudotojų atsakomybės

Naudotojų atsakomybės apima pagrindines įmonių vadovų ir analitikų veiklos sritis, t. y. priimti svarbius plėtros, biudžeto sprendimus įmonės mastu.

1.5.3. Naudotojų kvalifikacija

Kalbant apie įmonių vadovų kvalifikaciją, tikėtina, jog vartotojai turi geras kompiuterio naudojimo žinias, tačiau matematinės analizės ir duomenų prognozavimo žinios yra minimalios. Tai neleidžia jiems pasinaudoti matematinės analizės įrankiais, tokiais kaip „Excel“ ir „R“. Tuo tarpu, įmonių analitikų kvalifikaciją apibrėžti sunkiau, tai labai priklauso nuo kiekvienos organizacijos vidinių nuostatų. Vis dėl to, galima teigti, kad analitikai, turintys galias matematikos, programavimo ir prognozavimo žinias, ir sugebantys naudoti tokius plačius paketus kaip „R“, yra prieinami tik didžiausioms įmonėms.

1.5.4. Naudotojų kiekis

Naudotojų kiekis neturėtų būti didelis. Sprendimas nėra kuriamas masinei rinkai. Sprendimą naudotų tik įmonių aukštesnes pareigas užimantys atstovai. Taip pat naudojimas sprendimu vyktų retesniais intervalais, greičiausiai kas metų ketvirtį arba pusmetį, t. y. etapais, kai planuojamas įmonių sekančio laikotarpio biudžetas, išlaidos ir plėtros planai.

1.5.5. Naudotojų tikslai

Pagrindinis sprendimo naudotojų tikslas yra padidinti įmonės pelną, optimizuoti ir pagerinti įmonės ateities planų sudarymo procedūras.

1.5.6. Naudotojų problemos

Pagrindinė problema su kuria susiduria naudotojai – prognozavimo proceso sudėtingumas ir jo parengimui reikalingos žinios. Prognozavimo modelio sudarymas ir pritaikymas yra ilgas ir daug resursų reikalaujantis procesas. Dažniausiai ši prognozavimo informacija turi būti prieinama

kuo greičiau, nes kuo vėliau jiniai yra sužinoma, tuo mažesnė jos dabartinė vertė ir tuo sunkiau sudaryti organizacijos planus atsižvelgiant į ją.

1.6. Esamų problemos sprendimo metodų analizė (Lietuvos ir tarptautiniu mastu)

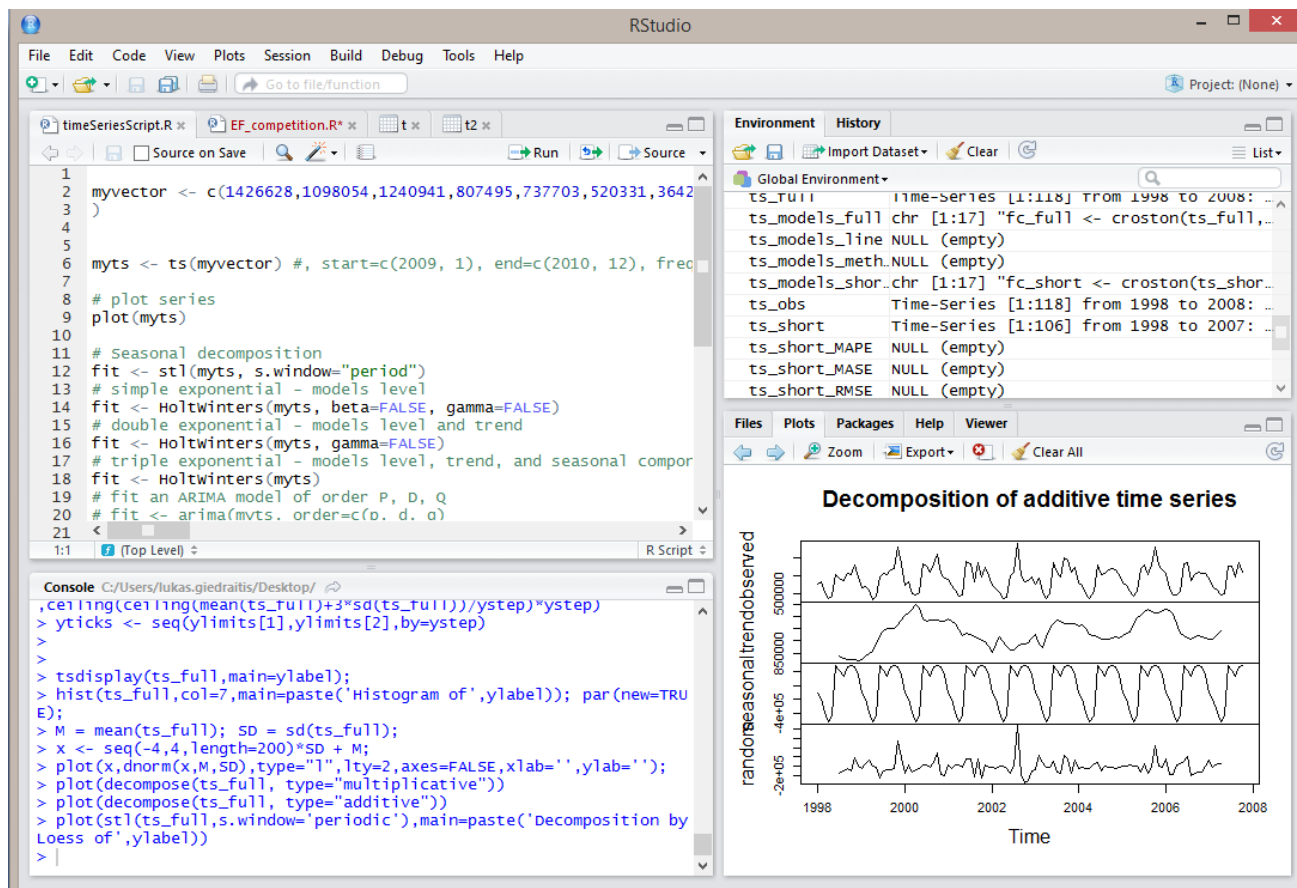
Toliau bus pateikta analizė įrankių, kurie galėtų atlikti prognozavimo procesą ir padėti įgyvendinti užsibrėžtus tikslus. Nors yra daugybė skirtingų įrankių, kai kurie iš jų yra labai panašūs tarpusavyje ir nėra reikalo bandyti juos atskirai kiekvieną išanalizuoti, atrasti privalumai ir trūkumai lengvai pritaikomi visiems tos srities įrankiams. Tam, kad palengvinti analizės procesą esami sprendimai suskirstyti į kelias grupes ir iš kiekvienos grupės atrinkti pagrindiniai lyderiai, kurie analizuojami plačiau.

Pasirinktos įrankių grupės:

- Integruotos kūrimo aplinkos (*RStudio*, *Mathematica*, *MatLab* ir pan.). Ši grupė pasižymi tuo, jog jos skirtos patyrusiems analitikams, turintiems programavimo pradmenis. Beveik visi įrankiai yra naudojami kartu su jiems pritaikyta programavimo kalba (*R*, *Wolfram Language*).
- Komerciniai verslo valdymo paketai (*Base CRM*, *SalesForce*, *Zoho CRM* ir pan.). Šios grupės produktai skirti visam įmonės valdymui, prognozavimo modulis sudaro tik mažąją dalį sistemos. Pagrindinė problema – norint panaudoti pardavimų prognozavimą, reikia įsidiegti ir naudoti visą paketą, apkraunant darbuotojus nereikalingu darbu ir skiriant didesnes lėšas negu būtų galima naudojant optimalų sprendimą.
- Duomenų analizės sistemos (*Excel*, *Tableau* ir pan.). Tai sistemos, skirtos įmonės duomenims valdyti ir veiklos analitikos procesams atlikti. Paprastai pasižymi plačiomis ataskaitų sudarymo ir atvaizdavimo galimybėmis, tačiau prognozavimo sritis būna paviršutiniška ir leidžianti pritaikyti tik pačius paprasčiausius prognozavimo metodus.

„RStudio“

„RStudio“ tai integruota kūrimo aplinka, palaikanti R programavimo kalbą, kuri skirta statistikos skaičiavimams, analizei ir grafiškiems atvaizdavimams. „RStudio“ naudojimas įmanomas keliomis versijomis: „RStudio Desktop“ versija paleidžiama lokaliai ir naudojama kaip įprastinė asmeninė programa, „RStudio Server“ leidžia priei prie programos naudojant interneto naršyklę, kai pati programa paleista nuotoliniame Linux serveryje. Tik pradėjus naudoti kūrimo aplinką, labiausiai į akis krenta funkcijų gausa, įvairios funkcijos pasiekiamos įdiegus papildomus paketus, kurie nuolat plečiami ir papildomi. Taip pat džiugina išsami dokumentacija, apie kiekvieną funkciją galima nesunkiai sužinoti pasinaudojant integruota pagalba arba nukeliavus į internete esančią dokumentacijos duomenų bazę. Esminis „RStudio“ trūkumas yra sudėtingumas: norint atlikti prognozavimo analizę reikalingos programavimo ir neblogos matematinės analizės žinios, galima drąsiai teigti, jog paprastas vartotojas, kuriam aktualus tik prognozavimo rezultatas, tikrai nesugebės įgyvendinti prognozavimo proceso. Sistema labiau skirta patyrusiems analitikams, kurie sugebėtų išnaudoti „RStudio“ galimybes, eiliniam vartotojui reikalingas „juodosios dėžės“ prognozavimo procesas, kurio „RStudio“ negali suteikti. Pagrindinis „RStudio“ langas pateiktas Pav. 3, matoma sritis padalinama į keturias dalis: užklauso langas, konsolės langas, aplinkos (kintamųjų) langas ir grafiko (galima perjungti į failų valdymo, pagalbos ir t.t.) sritis.



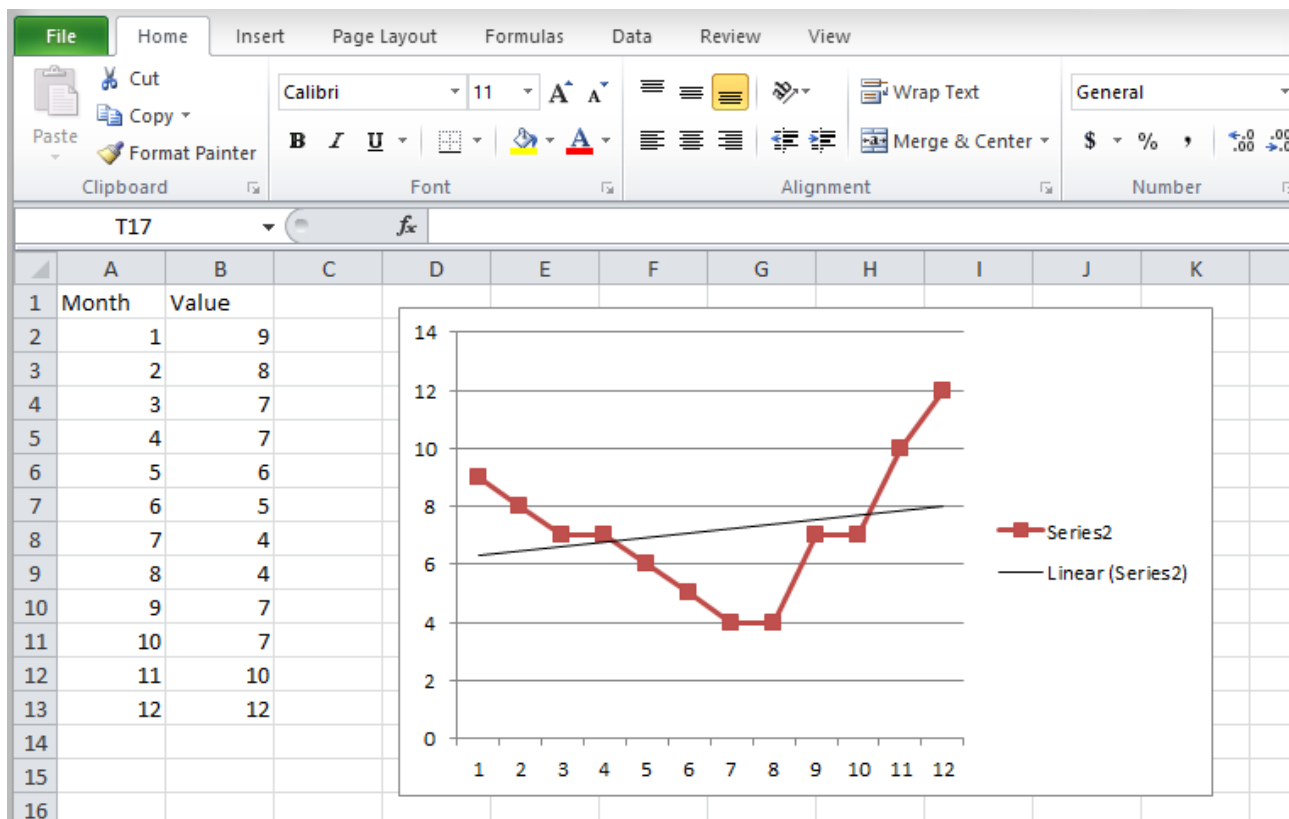
Pav. 3 „RStudio“ pagrindinio lango pavyzdys

“Microsoft Excel”

Pastaba: „Microsoft Excel“ programa analizuota nenaudojant papildomų įskiepių.

„Microsoft Excel“ programa labiausiai žinoma dėl savo skaičiavimo ir grafinio atvaizdavimo galimybių. Ji detalizuotų prognozavimo funkcijų neturi, tačiau prognozavimą gali imituoti pats vartotojas. Tai daroma pasinaudojus plačiomis formulių užrašymo ir skaičiavimo funkcijomis. Žinoma, tai įmanoma tik tuo atveju jeigu vartotojas turi geras prognozavimo matematinės žinias. Vartotojas, suprantantis prognozavimo procesą ir įvairių metodų aspektus, gali jį padalinti į smulkesnes užduotis ir, žingsnis po žingsnio šias užduotis apskaičiuodamas, gauti prognozavimo rezultatą. Pavyzdžiui, norint prognozuoti laiko eilutę, reikia pažingsniui rankiniu būdu atlikti duomenų glodinimą, sezoniškumo ir krypties komponentių išgavimus bei kitus žingsnius. Žinoma, teisingai atlikus visus etapus, gauta prognozė neturėtų skirtis nuo prognozavimo rezultato gautu specializuotomis programomis, tačiau laiko sąnaudos yra žymiai didesnės. Tai ir yra pagrindinis trūkumas, kai prognozavimas atliekamas bendresnės paskirties analizės programomis: automatizavimo galimybės yra per menkos, iš vartotojo reikalaujama per daug pastangų ir žinių, o skaičiavimo klaidos kaina yra kritinė.

Iliustracijoje Pav. 4 parodytas testinių duomenų suvedimas bei atvaizdavimas „Excel“ programoje. Taip pat krypties (angl. *trend*) komponentės išskyrimas automatinėmis „Excel“ priemonėmis.



Pav. 4 Testinių duomenų atvaizdavimas ir krypties komponentės išskyrimas Excel programoje

Base CRM

„Base CRM“ paketas priklauso santykių su klientu valdymo (angl. *CRM*) programų grupei. CRM galima apibrėžti kaip verslo požiūrį, kuris vienija žmones, procesus ir technologijas siekiant pagerinti bendradarbiavimą su klientais [22]. Visų pirma reikėtų paminėti, jog „Base CRM“ skirtas procesų ir žmonių valdymui, o prognozavimo galimybės sudaro mažą dalį viso funkcionalumo. Deja, dėl šios pačios priežasties, prognozavimo galimybės yra labai menkos. Šiame darbe buvo minima, jog siekiamas sprendimas turėtų veikti juodosios dėžės principu. Šią užduotį „Base CRM“ sprendžia be priekaištų – vartotojas neapkraunamas matematiniais skaičiavimais, jam rodomas tik galutinis rezultatas, tačiau vartotojui net ir pageidaujant nesuteikiama informacija apie naudojamą modelį ar bendrąją prognozavimo logiką. Taip pat nėra ir prognozavimo konfigūracijos, visi reikalingi duomenys parenkami automatiškai. Vartotojui neturinčiam detalių srities žinių tai palengvina darbą, tačiau srities specialistas gali norėti savarankiškai optimizuoti prognozavimo procesą, ir galima drąsiai teigti, jog patys tiksliausi rezultatai gaunami tik tada, kai pats vartotojas dalyvauja prognozavimo koregavime, neatiduodant visų sprendimų priėmimų automatizuotai programai. Nors šioje vietoje apžvelgiamas būtent specifinis „Base CRM“ įrankis, minėti privalumai ir trūkumai atspindi visuose bandytuose įrankiuose, kurie šiame darbe priskirti komercinių verslo valdymo paketų grupei.

Iliustracijoje Pav. 5 parodytas „Base CRM“ programos langas testinėje aplinkoje. Kairėje pusėje matomas apskaičiuotas prognozės rezultatas.

Dashboard

TOTAL PIPELINE WORTH
€776,200
 €124,970 Forecasted

ACTIVE DEALS
12
 0 Hot

IN THE LAST 30 DAYS
 0 Won
 0 Lost
 0 Unqualified

TEAM DEAL LOAD
 6 Colin Welch
 6 rest

Displaying all activity ▾ by everyone ▾

- rest left a note for **Litigation Support Deal** 08/03/2015 15:52
 Introduced our products. Said there's a great fit and that they will start the process of getting us integrated. They suggested that I check back with them later in the month for a status report.
- rest left a note for **Business Structure / Entity Selection** 08/03/2015 15:52
 Not sure what their interest is. Seems like they are holding something back. Will send them a demo video and will leave it in their hands until next week.
- rest left a note for **Franchise Review** 08/03/2015 15:52
 Followed up on the status of us getting on the approved vendor list. They said it should be any day and encouraged me to get in touch with the department heads.
- rest left a note for **Due Diligence Management** 08/03/2015 15:52
 Decision-maker was running late to the meeting. They are going to call me back when everyone is in the same room together.

Pav. 5 „Base CRM“ švieslentės langas testinėje aplinkoje

Esamų problemos sprendimų analizės išvados

Šiame skyriuje buvo aptarti programiniai įrankiai, kurie turi galimybes atlikti pardavimų prognozavimą. Visi įrankiai suskirstyti į tris grupes: integruotos kūrimo aplinkos, komerciniai verslo valdymo paketai ir duomenų analizės sistemos. Iš kiekvienos grupės atrinkti pagrindiniai konkurentai ir jie įvertinti išsamiau.

Ištirtuose įrankiuose labiausiai pasigęsta galybės prognozuoti neturint gerų programavimo ir matematikos žinių, bet su minimalia prieiga keisti prognozavimo logiką jeigu to prisireiktų. Šis reikalavimas nukreiptas į potencialius sistemos vartotojus: įmonių vadovus ir analitikus. Patogiam įmonių vadovų darbiui reikalingas prognozavimas juodosios dėžės principu, tačiau įmonių analitikų tai pilnai netenkina – jiems reikalingos bent jau minimalios prognozavimo algoritmo konfigūracijos galimybės.

1.2 lentelėje pateikta glausta informacija apie ištirtas įrankių grupes skirtingais aspektais.

1.2. Ištirtų įrankių lyginamoji analizė.

	Integruotos kūrimo aplinkos („RStudio“)	Komerciniai verslo valdymo paketai („Base CRM“)	Duomenų analizės sistemos („Microsoft Excel“)
Juodosios dėžės veikimas	Nėra.	Yra. Vartotojas nieko nežino apie prognozavimo proceso vykdymo eigą.	Nėra.
Reikalingos matematinės analizės žinios	Vidutinės. Reikalinga žinoti skirtingų prognozavimo metodų principus.	Nereikalingos.	Stiprios. Prognozavimo žingsniai vykdomi pačio vartotojo rankomis.

Reikalingos programavimo žinios	Stiprios. Reikia mokėti specifinę naudojamo įrankio kalbą.	Nereikalingos	Nereikalingos
Įvesties duomenų pasirinkimo galimybės	Vidutinės. Pradinių duomenų nuskaitymo galimybės didelės, tačiau visus duomenis reikia konvertuoti į aplinkos palaikomus kintamuosius, pvz.: vektorius ar matricas.	Labai silpnos. Prognozavimas vyksta tik su sistemoje sukauptais duomenimis.	Didelės. Palaikoma didžioji dalis duomenų bazių ir įvairiai formatuotų tekstinių failų.
Prognozavimo konfigūracijos galimybės	Labai didelės.	Labai silpnos.	Priklauso nuo vartotojo prognozavimo ir matematikos žinių
Sistemos pritaikymo savom prognozavimo reikmėm galimybės	Stiprios. Įrankių galimybės apima didžiąją dalį verslo sričių.	Labai mažos. Reikia tenkintis jau iš anksto sukurtomis priemonėmis, nepriklausomai nuo naudojimo srities.	Vidutinės
Palaikomų prognozavimo metodų kiekis	Labai didelis.	Labai mažas. Naudojami tik bendri ir esminiai metodai.	Priklauso nuo vartotojo prognozavimo ir matematikos žinių.

1.7. Darbo tikslas, uždaviniai ir siejami privalumai

Norint išspręsti nurodytas problemas, yra užsibrėžtas darbo pagrindinis tikslas – pagerinti ir palengvinti pardavimų prognozavimą, taikant veiklos intelektikos priemones. Palengvinimą galima suprasti kaip prognozavimo modelio sudarymo paprastumą, neturint specialių matematinių ir prognozavimo žinių. Supaprastinti modelio sudarymo procesą padeda veiklos intelektikos gerosios praktikos, duomenų saugyklos kūrimo metodikos ir jų pritaikymas. Procesų rezultatų tikslumo gerinimą lemia esamų prognozavimo modelių sudarymo analizė, jų eksperimentas su realiais duomenimis ir atrastų tendencijų bei išvadų panaudojimas dalykinėje srityje.

Norint įgyvendinti darbo tikslą išskiriami šie darbo uždaviniai:

1. Išanalizuoti svarbiausias duomenų saugojimo metodikas.
2. Aprašyti duomenų metamodelį, kuris leistų automatizuoti dalį prognozavimo proceso.
3. Išanalizuoti esamus prognozavimo metodus.
4. Aprašyti esamas prognozavimo aplinkas ir esminius įrankių paketus.
5. Specifikuoti ir sukurti prognozavimo modulį.
6. Eksperimentiškai ištirti sukurtą įrankį su farmacijos įmonių pardavimų duomenimis.
7. Apibendrinti tyrimo rezultatus, pateikti rekomendacijas ir išvadas.

Magistro baigiamojo darbo planas pateiktas 3.2 priede.

1.8. Siekiamo sprendimo apibrėžimas

Siekiamas programinis sprendimas yra duomenų prognozavimo programinė įranga, kuri tenkintų anksčiau iškeltus ir aprašytus tikslus. Kuriamą programą neturi būti siejama su konkrečia dalykine sritimi ar duomenų saugykla, tam, kad ateityje ją būtų galima panaudoti kitose sistemose, jeigu atsirastų toks poreikis. Pagrindinė funkcija – suskaičiuoti prognozavimo rezultatą iš gautų parametrinių duomenų ir pasirinkto prognozavimo modelio. Prognozavimo modelių pasirinkimų kiekis turi būti įgyvendinamas toks: dirbtiniai neuroniniai tinklai, tiesinė regresija ir laiko eilutės. Programinė įranga turi atlikti prognozavimą su įvairių sričių (ne tik farmacijos įmonių pardavimų) duomenimis, tačiau įvesties duomenys privalo būti korektiški ir iš anksto apibrėžtame formate.

Taip pat, viena pagrindinių sistemos užduočių yra parametrinių duomenų surinkimas iš vartotojo, jų validavimas, pertvarkymas į formatą, kurį palaiko sukurta programa, ir jų perdavimas programai atliekančiai pagrindinius skaičiavimus. Gavus prognozavimo rezultatus, jie turi būti atvaizduoti vartotojui suprantama forma, taip pat suteikiama galimybė rezultatus išsaugoti ar perkelti į vartotojo pasirinktą formatą.

Viena iš sudėtingiausių reikalingų funkcijų yra parametrų surinkimas neapkraunant šiuo darbu sistemos vartotojo. Turi būti galimybė išsaugoti suvestų parametrų reikšmes ateities naudojimui. Taip pat automatizuoti labiausiai naudojamų parametrų (žmonių populiacija, BVP augimas ir pan.) surinkimą iš trečiųjų šalių šaltinių ir išsaugojimą, tuo tarpu galutiniam vartotojui palikti tik jam skirtų asmeninių parametrų valdymą (įmonės skiriamas biudžetas reklamai, planuojamas konkurentų kiekis ir pan.).

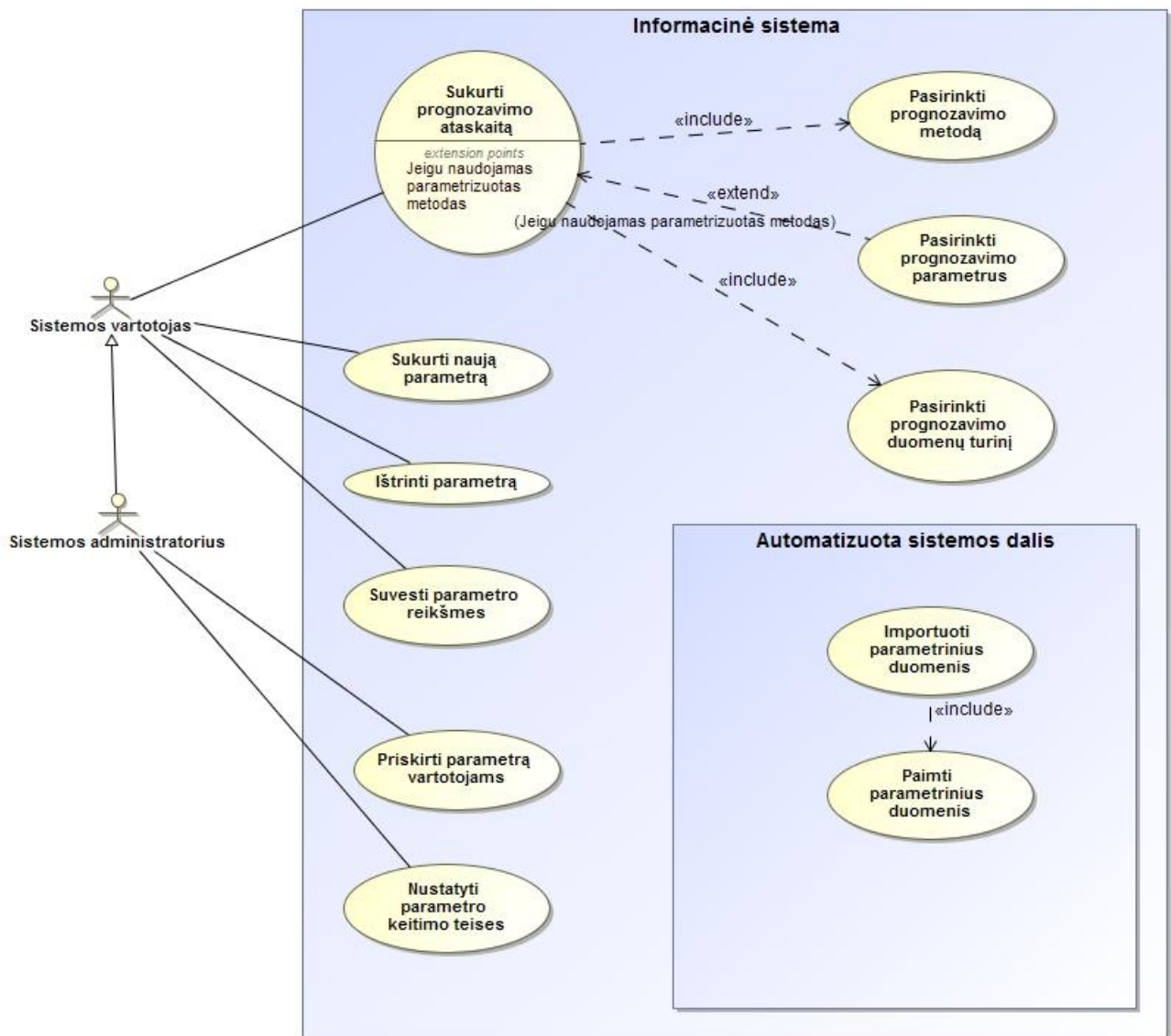
1.9. Analizės išvados

1. Darbe buvo išanalizuoti ir palyginti prognozavimo modeliai, mokomi su priežiūra ir be priežiūros, bei nuspręsta didesnę dėmesį skirti pirmajam tipui, dėl jo geresnio tinkamumo probleminei sričiai (galimybės gauti klaidos ar teisingo veikimo signalą, kuris įvertintų potencialų sprendimą).
2. Atlikus prognozavimo atskirų modelių analizę, nuspręsta įgyvendinti laiko eilučių, tiesinės regresijos ir neuroninių tinklų prognozavimo modelius, dėl jų automatizavimo galimybių ir tinkamumo taikymui pardavimų srityje. Tuo tarpu logistinės regresijos modelis atmestas dėl to, kad spėjamas kintamasis gali įgyti tik dvi reikšmes, kas nėra tinkama prognozuojant pardavimus.
3. Darbe išanalizuotos dabartinių prognozavimo įrankių kategorijos, nustatyti jų pagrindiniai trūkumai – veikimo sudėtingumas, specifinių žinių reikalingumas. Darbe sukurtas sprendimas, padengiantis aprašytus konkurentų trūkumus.

2. PARDAVIMŲ PROGNOZAVIMO ĮRANKIO SPRENDIMO REIKALAVIMŲ SPECIFIKACIJA IR PROJEKTAS, FORMALUS APRAŠAS

2.1. Reikalavimų specifikacija

Atlikus analizę ir nustatius įgyvendinamo įrankio preliminarius rėmus, buvo nuspręsta suprojektuoti ir realizuoti sistemą, kuri leistų atlikti pardavimų prognozavimą. Iliustracijoje Pav. 6 pateikta pardavimų prognozavimo sistemos panaudojimo atvejų diagrama. Sistema naudotųsi du vartotojų tipai: sistemos vartotojas ir sistemos administratorius. Administratorius turi daugiau teisių susijusių su parametru organizavimu ir valdymu. Diagramoje parodyta ir dalis sistemos veikimo, kurio nevaldo vartotojai – tai automatiškai ir periodiškai vykstantys saugyklos pildymo darbai.



Pav. 6 Prognozavimo įrankio panaudojimo atvejų diagrama

2.2. Nefunkciniai reikalavimai

Pavadinimas. Elementų paaiškinimas.

Apibrėžimas. Kiekvienas netrivialus grafinės sąsajos prognozavimo elementas turi būti paaiškintas 1–3 sakiniais. Jeigu trumpai paaiškinti nėra galimybės, reikia pateikti raktinius žodžius, kuriais remiantis būtų įmanoma sužinoti tikslesnį veikimą.

Patikrinimas. Prie kiekvieno testuotojui neaiškios paskirties elemento yra matomas elemento paaiškinimas ar nuoroda į jį.

Papildoma informacija. Nefunkcinio reikalavimo testavimą atlieka žmogus turintis vidutines kompiuterines žinias, bet gali būti neturinti jokių matematinių ar prognozavimo žinių.

Pavadinimas. Generavimo laikas.

Apibrėžimas. Ataskaitos generavimo laikas negali būti ilgesnis už 6 minutes. Jeigu laiko neužtenka – generavimas turi būti nutrauktas ir parodytas informacinis pranešimas.

Patikrinimas. Bandoma generuoti įvairaus sudėtingo ataskaitas ir stebėti ar sistema praneša apie per ilgą ataskaitos generavimą.

Pavadinimas. Informacijos konfidencialumas.

Apibrėžimas. Skirtingų kompanijų vartotojai negali matyti vienas kito sukurtų parametrų ir jų duomenų, norint užtikrinti informacijos konfidencialumą.

Patikrinimas. Sukurti parametrą ir užpildyti duomenis, prisijungti su kitos kompanijos vartotoju ir pažiūrėti ar sukurtas parametras nėra matomas ir nedaro įtakos kito vartotojo darbui.

Papildoma informacija. Parametras gali būti viešai prieinamas jeigu tokį pageidavimą išreiškia parametras sukūręs vartotojas (savininkas).

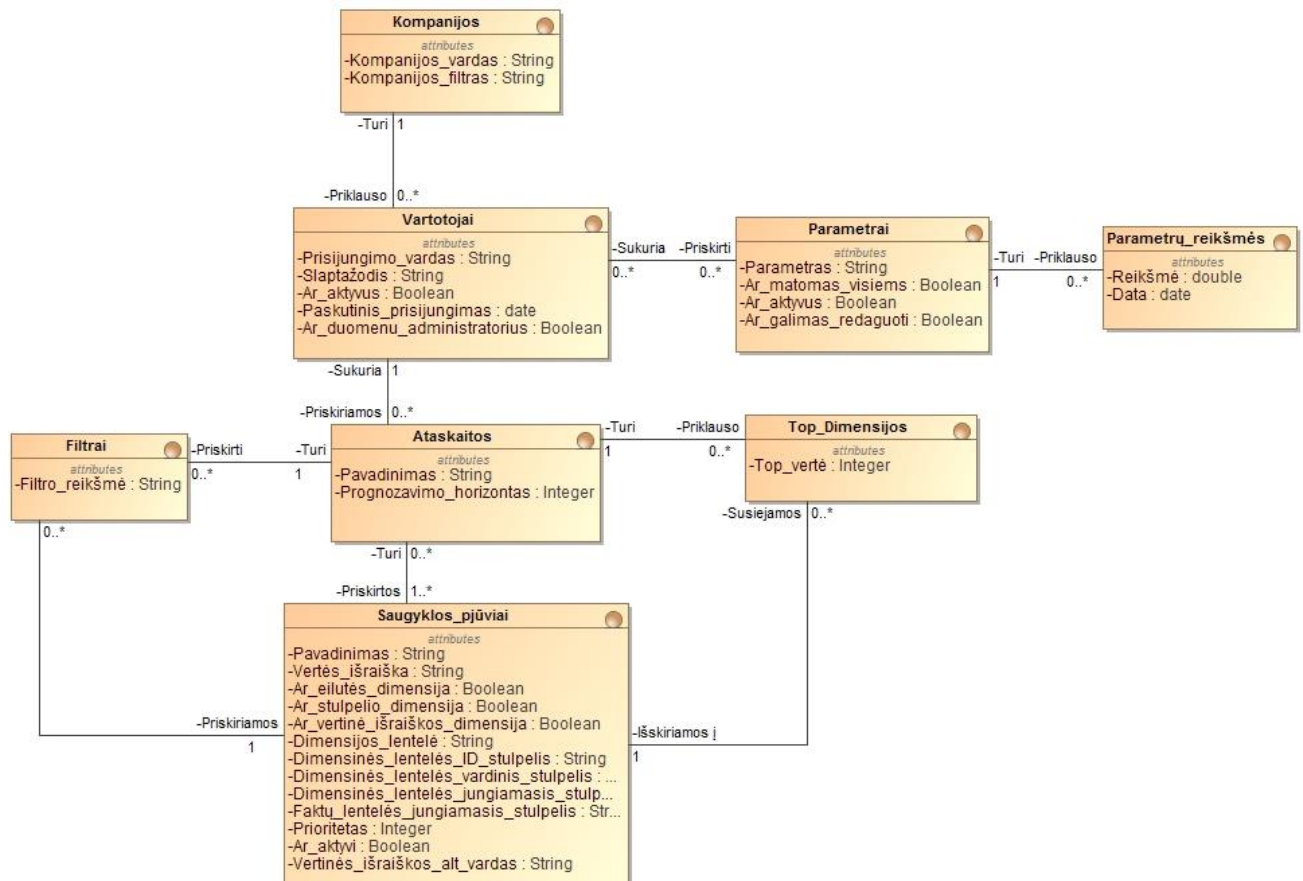
Pavadinimas. Generavimo lygiagretumas.

Apibrėžimas. Vienu metu turi būti palaikomas bent trijų vartotojų ataskaitų generavimas.

Patikrinimas. Prisijungti su trimis skirtingais vartotojais per skirtingus įrenginius ir vienu metu paleisti ataskaitos generavimą. Stebėti ar sistema atlaiko krūvį ir pateikia rezultatus.

2.3. Dalykinės srities modelis

Paveiksle Pav. 7 parodyta prognozavimo sistemos dalykinės srities esybių klasių diagrama. Esybės atsiranda iš realaus pasaulio objektų, sekančiame žingsnyje aprašomi ryšiai ir kardinalumai tarp jų. Esybių klasių diagrama gali būti naudojama kaip šaltinis kuriant sistemos duomenų bazės struktūrą. Toliau šiame skyriuje pateikta siauresnė informacija apie kiekvieną esybę.



Pav. 7 Prognozavimo įrankio esybių klasių diagrama

„Kompanijos“. Įmonės bei organizacijos, kurioms atstovauja sistemos vartotojai. Pagal vartotojo organizaciją sistema parinks reikiamus nustatymus, bendrus visiems tos organizacijos vartotojams (pvz., ar organizacija gali matyti specifinį prognozavimo parametą?).

2.1. Esybės „Kompanijos“ specifikacija

Atributas	Informacijos apibūdinimas
Kompanijos_vardas	Kompanijos pavadinimas.
Kompanijos_filtras	Filtras parašytas <i>SQL</i> kalba. Padeda kontroliuoti kiekvienos įmonės priėjimą tik prie jai skirtų duomenų. Filtras naudojamas, kai generuojama ataskaita iš duomenų vitrinės.

„Vartotojai“. Sistemą naudojantys vartotojai ir duomenys apie juos.

2.2. Esybės „Vartotojai“ specifikacija

Atributas	Informacijos apibūdinimas
Prisijungimo_vardas	Vardas, kurį vartotojas suveda norėdamas prisijungti prie sistemos
Slaptažodis	Vartotojo slaptažodis. Slaptažodis saugomas užšifruotas, kad apsaugoti duomenis įsilaužimo į duomenų bazę atveju.
Ar_aktyvus	Vartotojo būseną sistemoje

Paskutinis_prisijungimas	Paskutinio prisijungimo prie sistemos data
Ar_duomenų_administratorius	Būsena parodanti ar vartotojas gali redaguoti bendrą parametrų bazę.

„**Filtrai**“. Išsaugotų ataskaitų filtrai, kuriuos pasirinko vartotojas. Jeigu vartotojas dažnai generuoja sudėtingos struktūros ataskaitą, jam galima ataskaitos šabloną išsaugoti, kad kiekvieną kartą nereikėtų žymėti norimų filtrų iš naujo.

2.3. Esysbės „Filtrai“ specifikacija

Atributas	Informacijos apibūdinimas
Filtro_reikšmė	Naudojamo filtro reikšmė tekstiniu formatu. Reikšmė gali parodyti atitinkamos dimensijos reikšmės identifikatorių arba pačią reikšmę. Pvz., jeigu dimensija yra „Products“, o vartotojas išfiltruoja vaistą „Ibumex“, tai filtro reikšmė galėtų būti „Ibumex“ arba identifikatorius „4512“. Rašyti identifikatorių ar pačią reikšmę sprendžiama atsižvelgiant į duomenų bazės struktūrą ir specifinę dimensiją.

„**Ataskaitos**“. Vartotojo išsaugoti ataskaitų šablonai.

2.4. Esysbės „Ataskaitos“ specifikacija

Atributas	Informacijos apibūdinimas
Pavadinimas	Ataskaitos pavadinimas, kurį sugalvoja pats vartotojas išsaugodamas ataskaitą.
Prognozavimo_horizontas	Prognozuojamų į ateitį mėnesių kiekis

„**Saugyklos pjūviai**“. Pagrindinė sistemos esybė, kurioje saugomi metaduomenys apie duomenų saugyklos struktūrą. Naudodama šią esybę ir vartotojo pasirinktas dimensijas, sistema sukuria dinaminę *SQL* užklausą, kurią įvykdžius gaunami ataskaitos duomenys.

2.5. Esysbės „Dimensijos“ specifikacija

Atributas	Informacijos apibūdinimas
Pavadinimas	Dimensijos pavadinimas. Pvz., „Products“
Vertės išraiška	Stulpelio pavadinimas, kuriame saugoma dimensijos reikšmė. Jeigu dimensija yra vertinė išraiška, tada atributo reikšmė gali būti parašyta

	<i>SQL kalba, pvz., „Quantity*Price AS wp“</i>
Ar_eilutės_dimensija	Parodo ar dimensiją galima laikyti eilutės dimensija.
Ar_stulpelio_dimensija	Parodo ar dimensiją galima laikyti stulpelio dimensija.
Ar_vertinė_išraiškos_dimensija	Parodo ar dimensiją galima laikyti vertinės išraiškos dimensija.
Dimensijos_lentelė	Lentelės pavadinimas, kurioje saugomi duomenys apie dimensiją. Ši reikšmė dažniausiai sutampa su dimensijos pavadinimu, pvz., „Products“
Dimensinės_lentelės_ID_stulpelis	Stulpelio pavadinimas, kuriame saugojamas įrašo identifikatorius, pvz., „ProductID“
Dimensinės_lentelės_vardinis_stulpelis	Stulpelio pavadinimas, kuriame saugojamas įrašo vardas, pvz., „ProductName“
Dimensinės_lentelės_jungiamasis_stulpelis	Stulpelio pavadinimas, kurį panaudojus, tėvinę lentelę galima sujungti (angl. <i>join</i>) su duomenų vitrinės lentele. Dažniausiai ši reikšmė sutampa su „Tėvinės_lentelės_ID_stulpelis“ atributu.
Faktų_lentelės_jungiamasis_stulpelis	Duomenų vitrinės lentelės stulpelio pavadinimas, kurį panaudojus, tėvinę lentelę galima sujungti (angl. <i>join</i>) su duomenų vitrina. Dažniausiai ši reikšmė sutampa su „Tėvinės_lentelės_jungiamasis_stulpelis“ atributu.
Ar_galima_prognuoti	Parodo ar atitinkamos dimensijos reikšmės galima nuprognuoti.
Ar_aktyvi	Parodo ar dimensiją galima naudoti sistemoje.
Vertinės_išraiškos_alt_vardas	Ši reikšmė naudojama tik prie vertinės išraiškos dimensijų. Ji parodo koks yra vertinės išraiškos alternatyvinis vardas (angl. <i>alias</i>) naudojamas kuriant <i>SQL</i> užklausą.

„Parametrai“. Tai sistemoje egzistuojantys prognozavimo parametrai. Juos gali kurti visi vartotojai ir tokius parametrus naudoti asmeniškai, asmeninių parametrų pavyzdžiai: įmonės lėšos reklamos paslaugoms, planuojamos ateinančių mėnesių produktų kainos ir t.t. Jeigu prie sistemos prisijungęs duomenų administratorius, jam leidžiama kurti bendrinius parametrus, administruoti jų matomumą specifinių vartotojų atžvilgiu. Bendrinių parametrų pavyzdžiai: vidutinė mėnesinė temperatūra, Lietuvos BVP augimas ir t.t.

2.6. Esysbės „Parametrai“ specifikacija

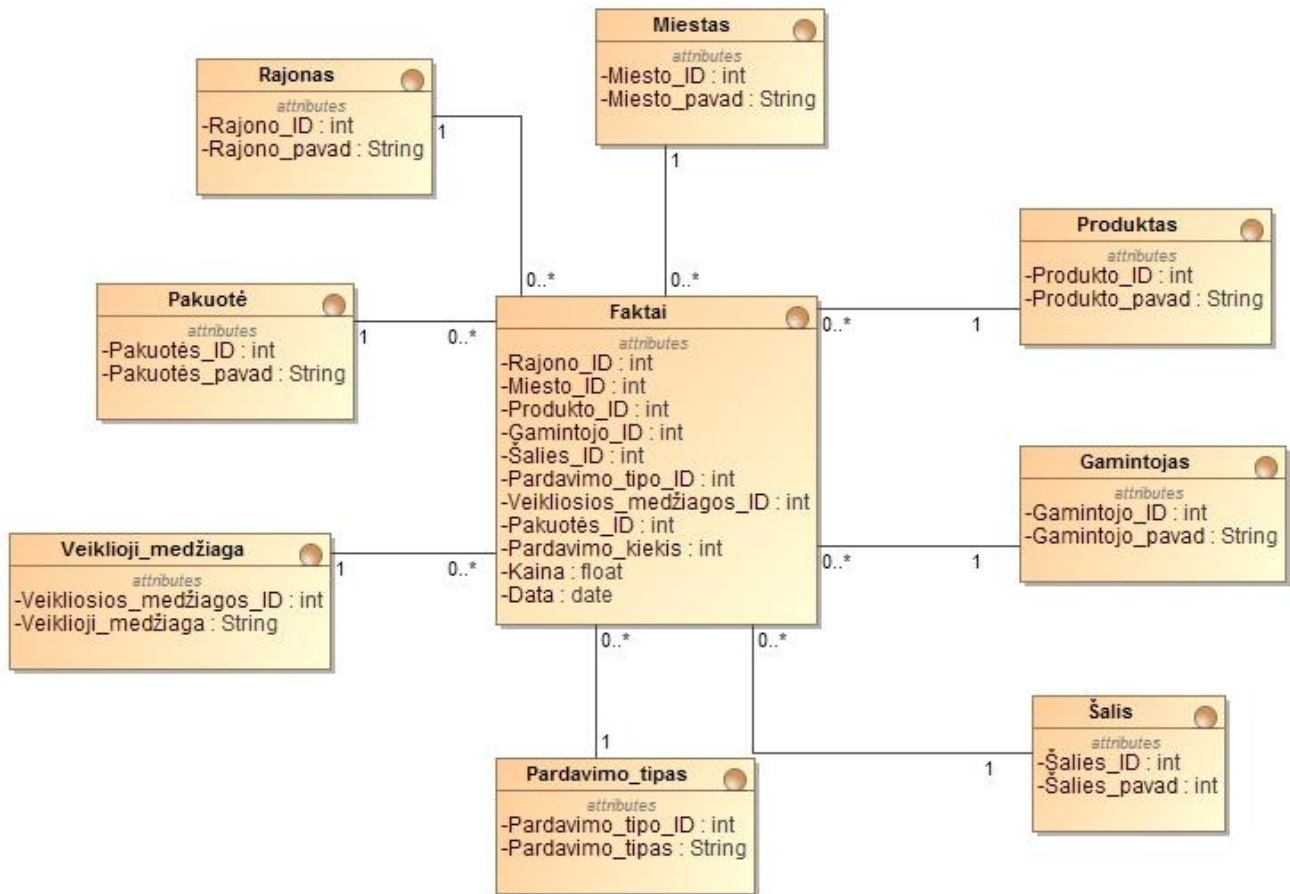
Atributas	Informacijos apibūdinimas
Parametras	Parametro pavadinimas
Ar_matomas_visiems	Būsena rodanti ar parametras rodomas visiems sistemos vartotojams
Ar_aktyvus	Būsena rodanti ar parametras matomas
Ar_galimas_redaguoti	Būsena rodanti ar priskirti parametru matytojai gali keisti parametro reikšmes

„Parametrų reikšmės“. Tai parametrų reikšmės tam tikru laikotarpiu. Šioje vietoje saugomos istorinės parametrų reikšmės bei ateities numatomos reikšmės.

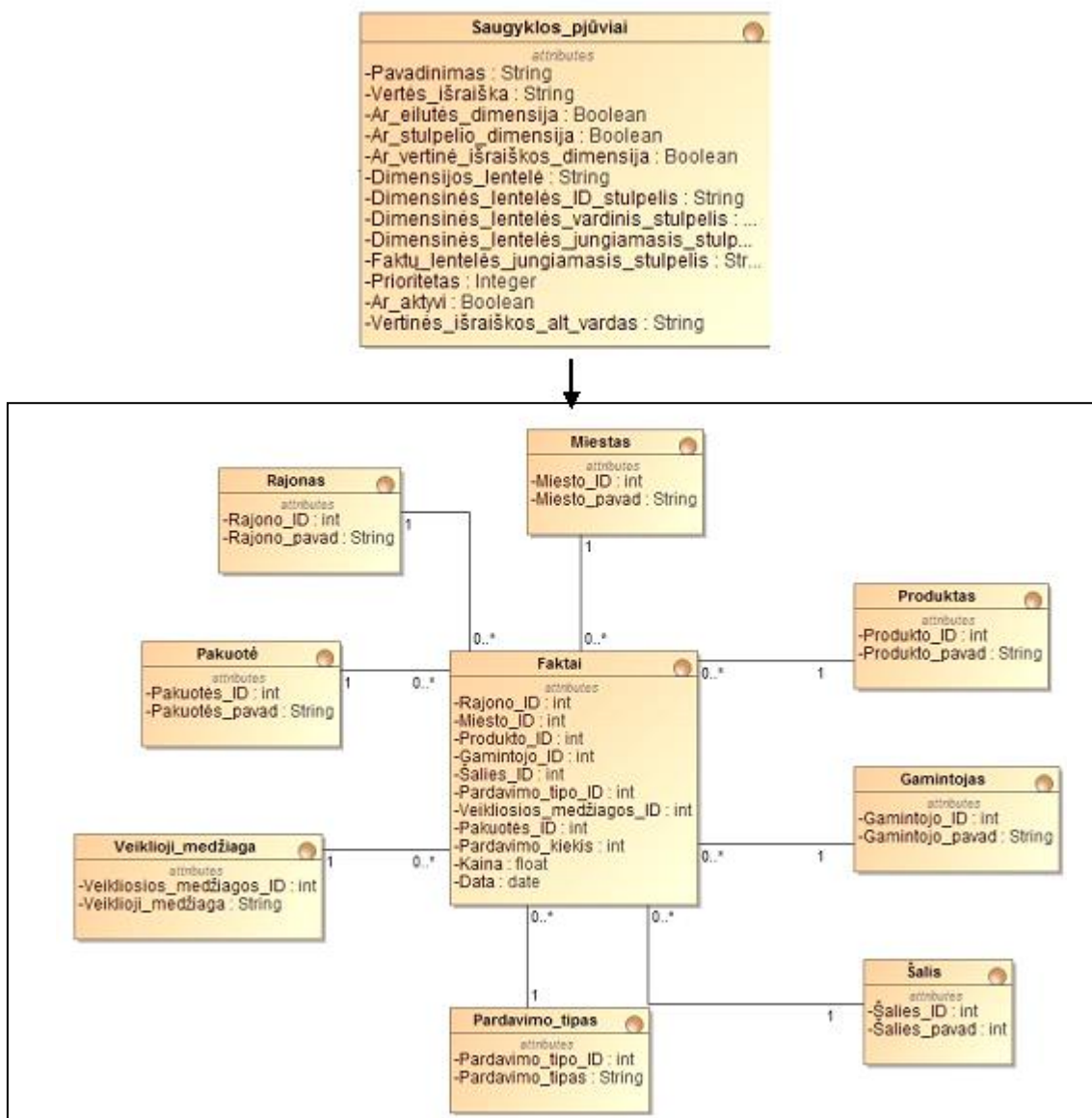
2.7. Esysbės „Parametrų reikšmės“ specifikacija

Atributas	Informacijos apibūdinimas
Reikšmė	Parametro skaitinė reikšmė
Data	Data, kada galiojo atitinkamo parametro reikšmė

Iliustracijoje Pav. 8 pateikta diagrama vaizduojanti pavyzdinę saugyklos esybių schemą. Naudojama schema yra žvaigždės tipo – viduryje viena faktų lentelė, surišta su neribotu kiekiu dimensijų. Pav. 9 diagramoje esanti „Saugyklos pjūviai“ esybė būtent ir saugo šios žvaigždės schemas struktūrą: kaip rišami faktai su dimensijomis, per kokius atributus vykdomas rišimas ir pan. Metaduomenų panaudojimas schemas žinioms aprašyti parodytas Pav. 9. Visa ši meta informacija padeda saugyklą naudojančioms sistemoms kurti ir vykdyti dinamines SQL užklausas, jų pareikalavimo metu.



Pav. 8 Pavyzdinės saugyklos esybių klasių modelis



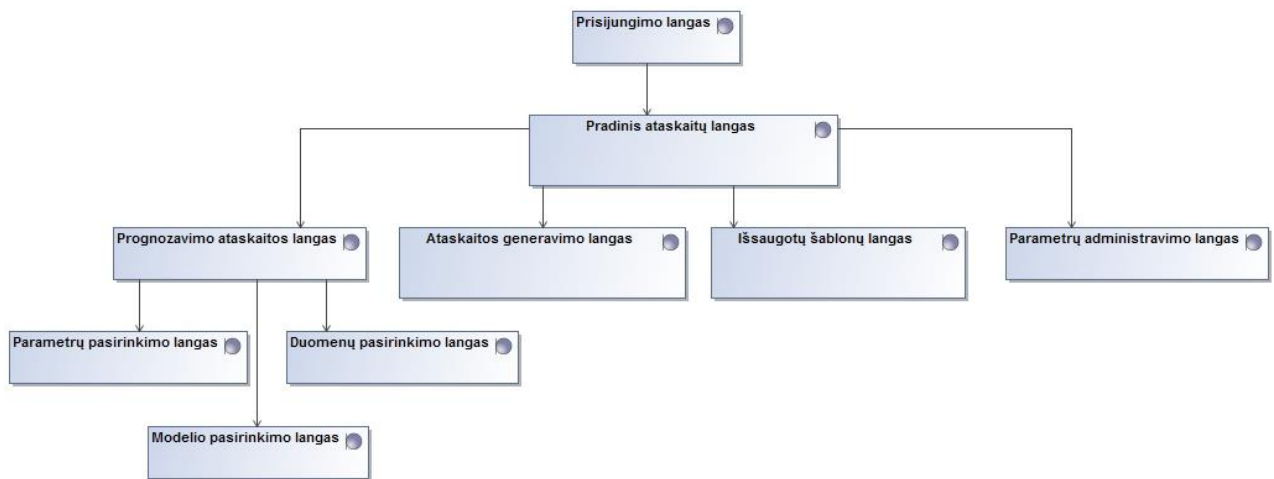
Pav. 9 Metaduomenų panaudojimo duomenų aprašyme pavyzdys

2.4. Naudotojų sąsajos modelis

Iliustracijoje Pav. 10 pateiktas prognozavimo sistemos navigavimo planas. Esminis langas yra „Prognozavimo ataskaitos langas“, kuriame galima nurodyti reikalingus parametrus (jeigu pasirinktas regresinis ar neuroninių tinklų modelis), pasirinkti prognozavimo modelį ir pasirinkti dimensijas, kurios bus prognozuojamos. Toliau pateikti atskiri svarbiausių langų grafinės sąsajos modeliai.

Iliustracijoje Pav. 11 pateiktas prognozavimo duomenų pasirinkimo langas. Jame reikia pasirinkti eilutės dimensiją, stulpelio dimensiją ir vertinę išraišką. Filtro skiltyje reikia pasirinkti vieną įrašą iš pasirinktos eilutės dimensijos, nes prognozė bus atliekama tik vienam egzemplioriui. T. y. jeigu pasirinkta eilutės dimensija yra produktai, tai filtro skiltyje reikia pasirinkti specifinį produktą, kurio pardavimai bus prognozuojami.

Iliustracijoje Pav. 12 parodytas parametrų pasirinkimo lango prototipas. Šis langas matomas tik tada, kai vartotojas pasirenka parametrizuotą prognozavimo metodą. Vartotojas gali pasirinkti vieną ar daugiau parametrų, jeigu parametrų kiekis netenkina, galima susikurti naują asmeninį parametą matomą tik jo savininkui.



Pav. 10 Prognozavimo sistemos vartotojo sąsajos navigavimo planas

Filters

Product A
 Product B
 Product C

Row Dimensions

Product
 Producer

Column Dimensions

Month
 Year

Value Dimensions

Quantity
 Euros

OK

Pav. 11 Duomenų pasirinkimo lango prototipas



Pav. 12 Parametrų pasirinkimo lango prototipas

2.5. Formalus sprendimo aprašas

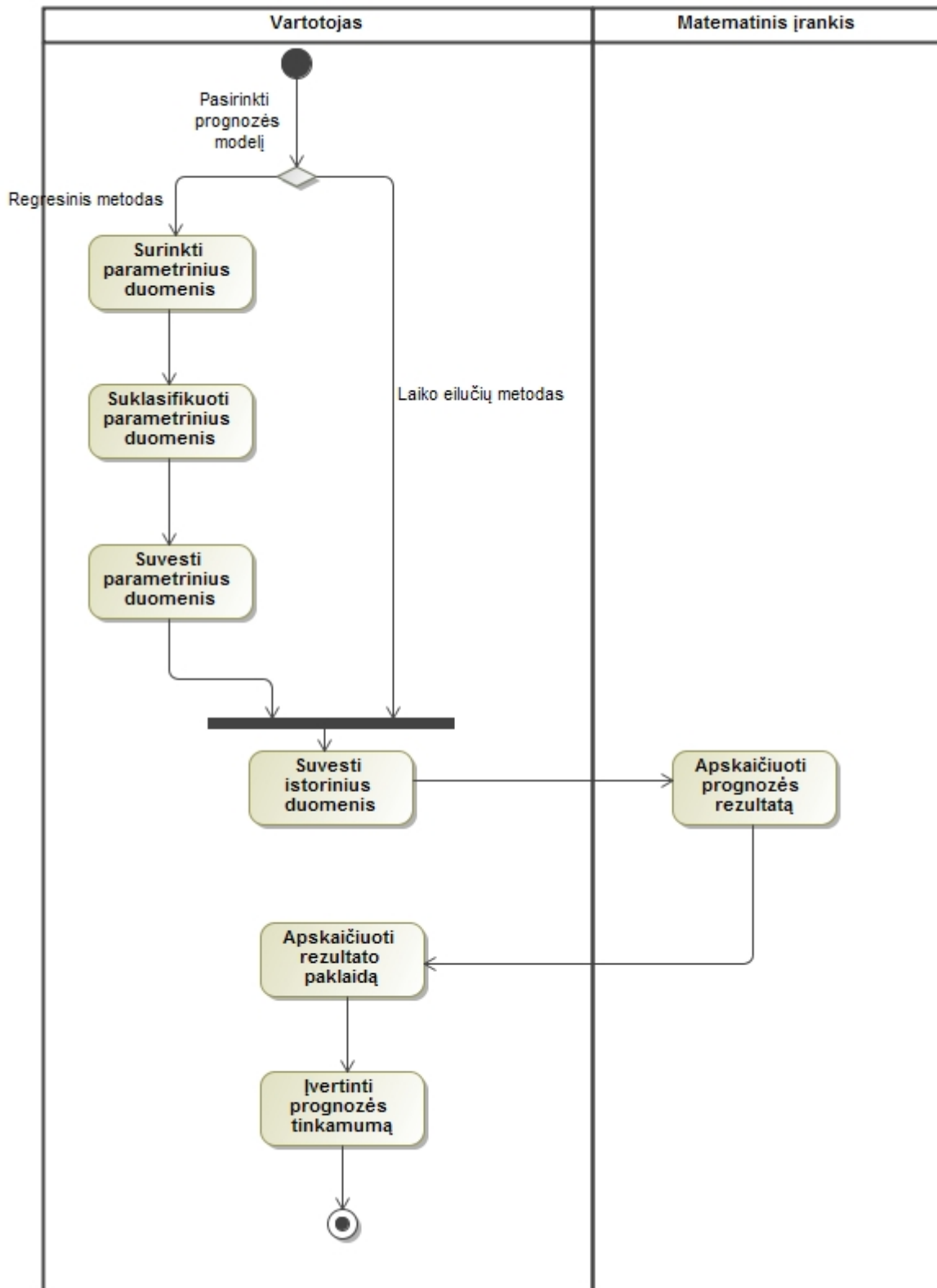
Iliustracijoje Pav. 13 parodyta prognozavimo proceso veiklos diagrama dabartinėje aplinkoje. Minimas matematinis įrankis gali būti tiesiog skaičiuotuvas arba programinis skaičiavimo paketas Microsoft Excel. Nors iš pirmo žvilgsnio procesas gali atrodyti nesudėtingas, tačiau daugumos atskirų žingsnių įgyvendimas gali užtrukti labai ilgai juos atliekant rankiniu būdu. Taip pat labai padidėja atsiradusios klaidos tikimybė, dėl kurios procesą gali tekti kartoti, arba, dar blogesnėje situacijoje, klaida iš viso nepastebima ir toliau organizacijoje priimami svarbūs sprendimai grindžiami netikslinga prognoze.

Pasirinkus regresinį metodą pirmasis žingsnis yra „Surinkti parametrinius duomenis“. Tai vienas iš ilgiausiai užtrunkančių darbų. Paprastai kiekvieną parametą reikia surinkti iš skirtingo šaltinio, taip pat turimus duomenis papildyti reguliariais intervalais. Nėra lengva ir šiuos duomenis saugoti, bėgant laikui jų kiekis sparčiai didėja, tampa sunku juos organizuotai kaupti. Turbūt populiariausias parametrinių duomenų saugojimo būdas yra tekstiniai failai – kableliais atskirtos reikšmės (angl. CSV), apie tai daugiau galima sužinoti šaltinyje [23]. Deja, tai nepatogu dėl kelių priežasčių:

- Kiekvienas parametras saugomas atskirame faile. Esant jų dideliame kiekiui tampa sunku suvaldyti organizuotą saugojimą.
- Atliekant prognozavimo procesą, reikia visų naudojamų parametrų reikšmes perkelti į bendrą skaičiavimo sistemą, nesvarbu ar kompiuterizuotą ar ne.
- Neįmanoma vykdytų užklausų greitai parametrinių duomenų manipuliacijai atlikti.
- Nepatogu, jeigu prognozavimas yra kooperatyvinis ir vyksta dideliu mastu. Didelėje įmonėje prognozavimo ataskaitą gali rengti grupė žmonių, kurios nariai išsibarstę skirtinguose padaliniuose [24]. Tokiu atveju reikalingas centralizuotas parametrinių duomenų valdymas. Taip pat dėl išsiskiriančių nuomonių prognozavimą gali tekti parengti keliomis iteracijomis.

Žingsnis „Suklasifikuoti parametrinius duomenis“ priklauso nuo to kas prognozuojama ir kokie parametrai bus naudojami. Dažniausiai į klasifikavimą įeina: parametrų suskirstymas į grupes pagal numatomą reikšmingumą, visų naudojamų parametrų sulyginimas tuo pačiu laiko periodu. Taip pat, jeigu turimas parametras yra ne skaitinis, reikia jį aprašyti dvejomis būsenomis, apie tai daugiau rašoma poskyryje „Logistinė regresija.“

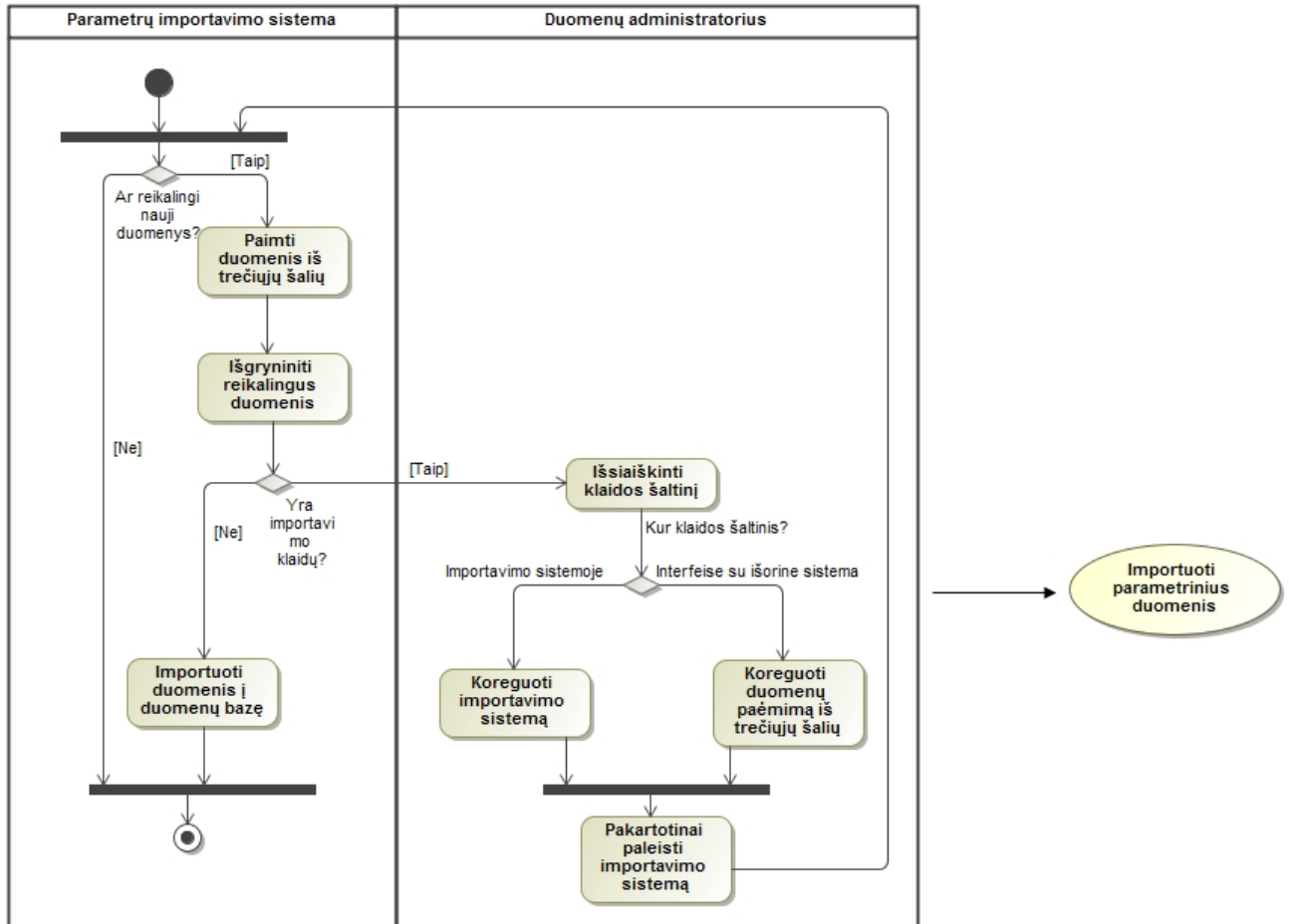
Paskutiniai prognozavimo proceso žingsniai yra prognozės skaičiavimas ir jos įvertinimas. Žingsnių įgyvendinimo galimybės labiausiai priklauso nuo naudojamos prognozavimo aplinkos ir tai atliekančio žmogaus patirties bei žinių.



Pav. 13 Prognozavimo vykdymo veiklos diagrama dabartinėje aplinkoje

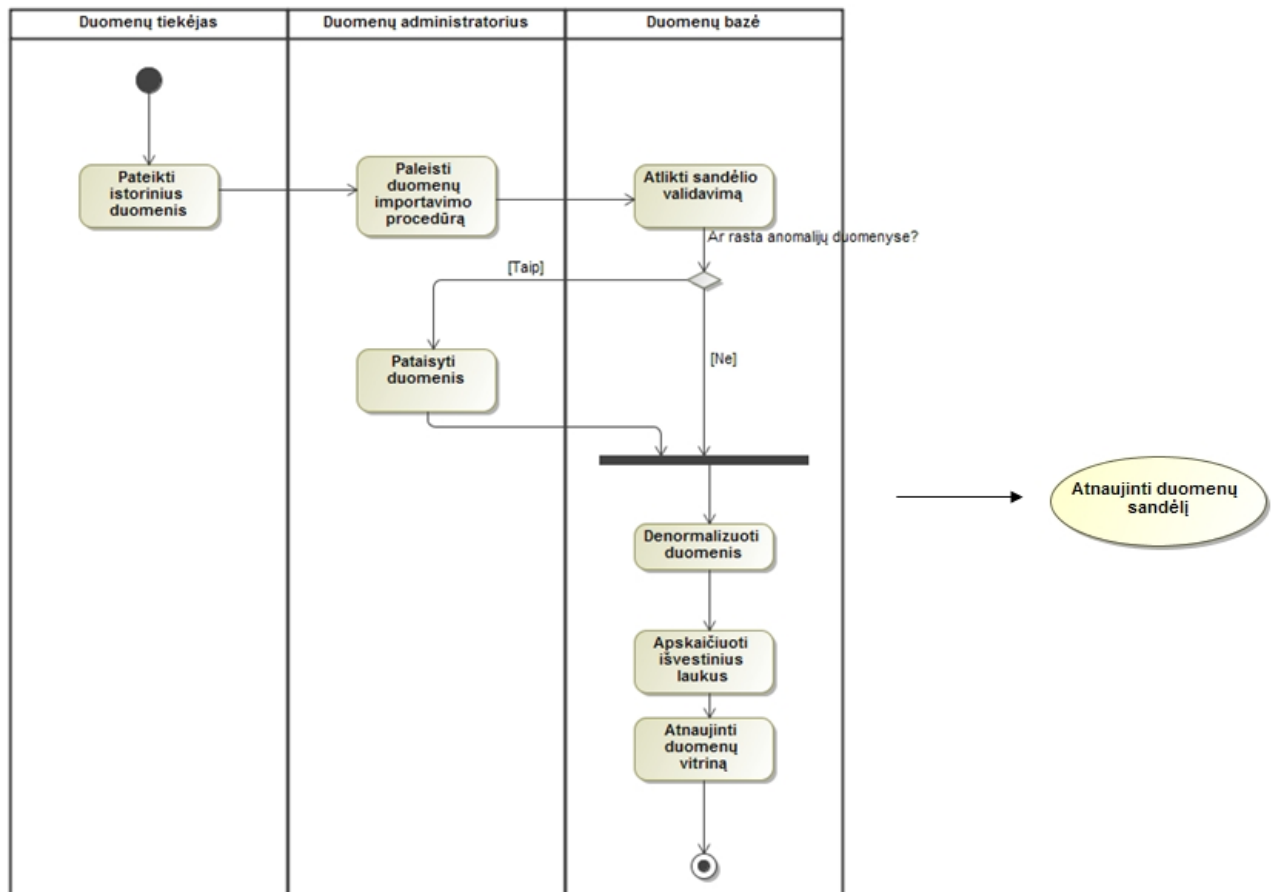
Iliustracijoje Pav. 14 pateiktas sprendimas kaip automatizuoti dalį prognozavimo proceso. Tai būtų atliekama panaudojus automatinio veikimo parametrų importavimo sistemą. Ši sistema

galėtų automatizuoti kelis žingsnius parodytus Pav. 13 diagramoje: „Surinkti parametrinius duomenis“, „Suklasifikuoti parametrinius duomenis“ ir „Suvesti parametrinius duomenis“. Tokios sistemos priežiūra tektų duomenų administratoriui, tačiau jo veiklos įsitraukimo reiktų tik sistemos klaidos atveju. Tikėtina, jog dažnesni klaidos pranešimai galimi tik po sistemos įdiegimo, nes kiekvienas atskiras parametras būtų gaunamas iš atskiro šaltinio – reiktų tvarkyti atskirą bendravimo interfeisą. Tačiau sistemai įsivažiavus ir teisingai sukonfigūravus bendravimą su trečiųjų šalių aplikacijų programavimo sąsaja (angl. API), duomenų administratoriaus darbas turėtų sumažėti iki minimumo.



Pav. 14 Kompiuterizuotos parametru valdymo sistemos veiklos diagrama

Iliustracijoje Pav. 15 pateiktas sprendimas kaip automatizuoti žingsnį „Suvesti istorinius duomenis“ parodytą Pav. 13. Iš pradžių yra vykdomas naujų istorinių duomenų importavimas į sistemą, tai paprastai atliekama kiekvieną mėnesį. Istoriniai duomenys prognozavimui yra reikalingi visada, nepriklausomai nuo pasirinkto modelio, aprašyto srities analizės skyriuje. Sėkmingai suimportavus duomenis į duomenų saugyklą, vykdoma duomenų denormalizacija. Duomenų denormalizavimas pagreitina duomenų išrinkimą, todėl gerai tinka duomenų vitrinoms, kuriose pagrindinis vykdomas procesas yra duomenų skaitymas [25]. Galiausiai yra atnaujinama duomenų vitrina denormalizuotais duomenimis. Duomenų vitriną vėliau galima panaudoti prognozavimo procese, kai yra reikalingi istoriniai duomenys.



Pav. 15 Duomenų importavimo ir vitrinės kūrimo proceso veiklos diagrama

Iliustracijoje Pav. 16 parodyta bendresnė diagrama sujungianti ankstesnius „to-be“ modelius ir atvaizduojanti visą prognozavimo procesą būsimoje sistemoje. Pirmasis sistemos vartotojo esminis žingsnis, lemiantis tolesnį proceso vykdymą, yra „Pasirinkti prognozavimo metodą“. Sistema turi leisti rinktis iš tiesinės regresijos, neuroninių tinklų ir laiko eilučių metodų. Pasirinkus pastarąjį, prognozavimo parametrai nereikalingi ir parametų pasirinkimo žingsnis praleidžiamas. Renkantis skirtingą parametrizuotą metodą (neuroninius tinklus ar tiesinę regresiją) vartotojui matomas procesas beveik nesiskiria, nes sistema taiko juodosios dėžės principą.

Svarbus klausimas lieka kaip teisingiau pasirinkti prognozavimo metodą: parametrizuotą ar ne. Daugiau apie tai rašyta analizės dalyje „Mokymas be priežiūros“ skyriuje.

Sekantis svarbus žingsnis iliustracijoje Pav. 16 yra „Pasirinkti prognozavimo parametrus“ (jeigu ankstesniame etape pasirinktas parametrizuotas metodas). Labai dažna neteisinga nuomonė yra, jog prognozavimo modelis visada gerėja, pridėdant daugiau aiškinamųjų parametų (angl. *explanatory variable*). Renkantis didelį kiekį parametų didėja tikimybė parinkti neesminius parametrus. Jeigu modelyje naudojama per daug nereikšmingų parametų įvyksta per didelis pritaikymas (angl. *overadjustment*) t. y. dalis nereikšmingų kintamųjų būna nustatomi kaip turintys akivaizdų poveikį prognozei, tačiau tai įvyksta tik dėl esančio atsitiktinumo mokymo duomenyse [26]. Šie kintamieji iš tiesų padeda gauti tikslesnę regresijos formulę naudojant mokymo duomenų aibę, tačiau, dėl didesnių atsitiktinių variacijų, formulė gali tapti nebepanaudojama prognozuojant ateities duomenis (už mokymo duomenų aibės ribų). Pagrindinį parametų pasirinkimo tikslą galima apibūdinti taip: [27] rasti formulę su mažiausiu kiekiu parametų, kurios tikslumas per daug nenusileidžia formulės su visais galimais parametrais tikslumui.

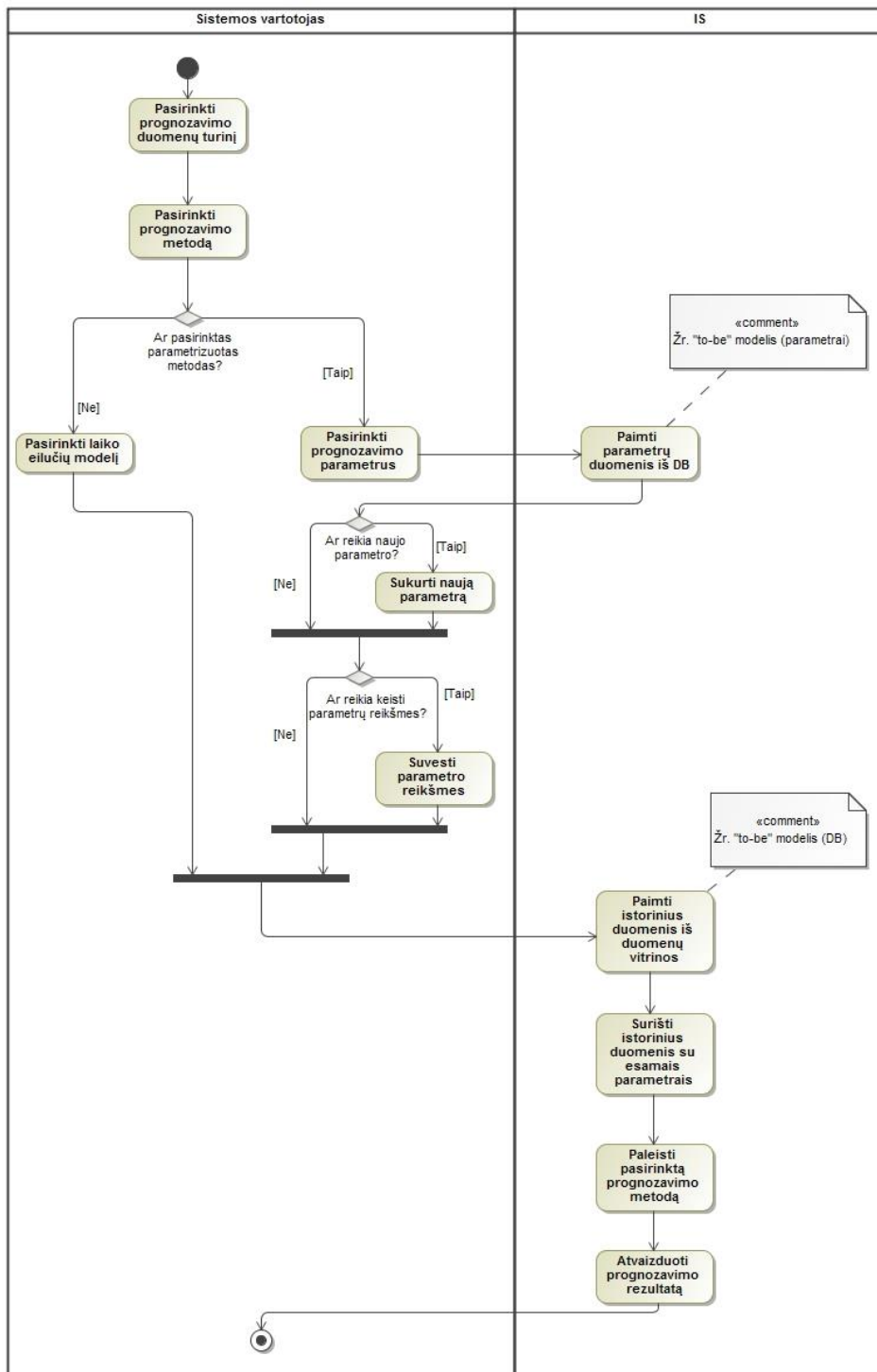
Klausimas išlieka kaip atrinkti šiuos esminius kintamuosius. Galima išskirti tokius būdus [26]:

- Progresinė atranka (angl. *forward selection*). Kiekvienu žingsniu pridedamas naujas parametras, kuris tiksliau apibūdina prognozuojamą narį. Procesas stabdomas, kai nebelieka galimų parametų aiškiai apibūdinančių prognozuojamą narį.
- Atgalinė atranka (angl. *backward selection*). Pradedama turint formulę su visais galimais parametrais. Atrenkamas parametras, kurio išmetimas mažiausiai lemia bendrą tikslumą ir jis pašalinamas iš modelio. Tai tęsiama tol, kol formulėje lieka tik tokie parametrai, kurių išmetimas smarkiai pakenktų modelio tikslumui.
- Žingsninė atranka (angl. *stepwise selection*). Mišrusis metodas. Pradedama nuo tuščios parametų aibės ir jie pažingsniui pridedami, tačiau po kiekvieno pridėjimo tikrinama ar vienas iš ankstesnių parametų netapo neprasmingu. Jeigu toks parametras randamas – jis išmetamas.

Modelio patikrinimas atliekamas naudojant du būdus: „Forward selection“ ir „Backward selection“. Jeigu abu kartus gaunama tokia pati parametų aibė – modelį galima vadinti tvirtu (angl. *robust*) [26].

Tolesnis žingsnis Pav. 16 diagramoje yra „Paimti parametų duomenis iš DB“. Dėl detalesnės informacijos kaip duomenų bazė užpildoma parametų duomenimis, žiūrėti Pav. 14.

Dar vienas esminis žingsnis yra „Paimti istorinius duomenis iš duomenų vitrinos“. Duomenų vitrinos plėtojimas detalizuotas Pav. 15 diagramoje.



Pav. 16 Prognozavimo proceso veiklos diagrama

3. PARDAVIMŲ PROGNOZAVIMO ĮRANKIO SPRENDIMO ARBA EKSPERIMENTINĖS REALIZACIJOS PROJEKTAS

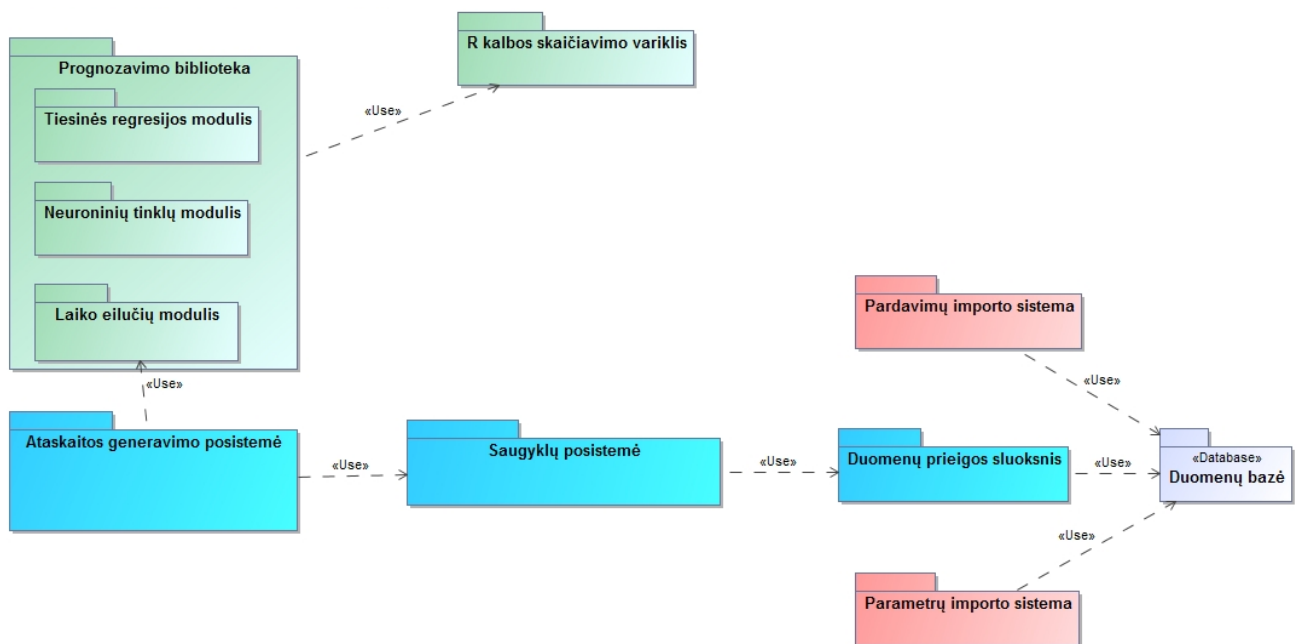
3.1. Sistemos loginė architektūra

Pav. 17 Pateikta sistemos loginės architektūros diagrama. Pagrindinė sistemos dalis yra „Ataskaitos generavimo posistemė“, ji atsakinga už programos logikos vykdymą, darbo organizavimą ir bendravimą su vartotoju per grafinę sąsają. „Saugyklų posistemė“ atsakinga už duomenų valdymą: CRUD funkcijų realizaciją ir vykdymą. „Duomenų prieigos sluoksnis“ yra sisteminis ir palengvina duomenų bazės manipuliaciją, suteikia galimybę lengviau įgyvendinti saugyklų posistemės sluoksnį.

„Pardavimų importo sistema“ yra atskira sistema, nebendruojanti su projektuojama sistema. Jos užduotis yra reguliariais intervalais automatiškai papildyti duomenų saugyklą naujais duomenimis. Savo funkcionalumu labai panaši yra „Parametrų importo sistema“. Vienintelis skirtumas tarp šių sistemų yra jų importuojamų duomenų turinys.

Toliau yra keli elementai atsakingi už prognozavimo įgyvendinimą. „Prognozavimo biblioteka“ yra atsakinga už gaunamų duomenų pertvarką, jų paruošimą prognozavimui ir reikalingų skaičiavimo funkcijų iškvietimą. Bibliotekos sandara priklauso nuo įgyvendinamų prognozavimo metodų, diagramoje šiuo metu jų yra trys ir kiekvienas iš jų turi atskirą modulį. Prognozavimo biblioteka glaudžiai susijusi su R programavimo kalba ir jos skaičiavimo varikliu, bibliotekoje kreipiniai į skaičiavimo variklį turi būti parašyti tam tikra skaičiavimo kalba – šiuo atveju R. Variklis yra atskira nepriklausoma programa, kuri sugeba įvykdyti jai siunčiamas skaičiavimo užklausas, grąžinti rezultatą.

Diagramoje spalvos žymi skirtingus elementų tipus: žalsva spalva pažymėti komponentai atsakingi už prognozavimo vykdymą, mėlsvi elementai atsakingi už ataskaitų generavimą, raudoni elementai yra pagalbinės sistemos importuojančios duomenis, reikalingus užtikrinti kitų sistemų teisingą veikimą. Šiame darbe kuriami komponentai yra „Prognozavimo biblioteka“, „Ataskaitos generavimo posistemė“, „Saugyklų posistemė“, „Duomenų prieigos sluoksnis“, „Parametrų importo sistema“ ir dalis duomenų bazės. „R skaičiavimo variklis“ yra trečiųjų šalių biblioteka atliekanti žemo lygio skaičiavimus.

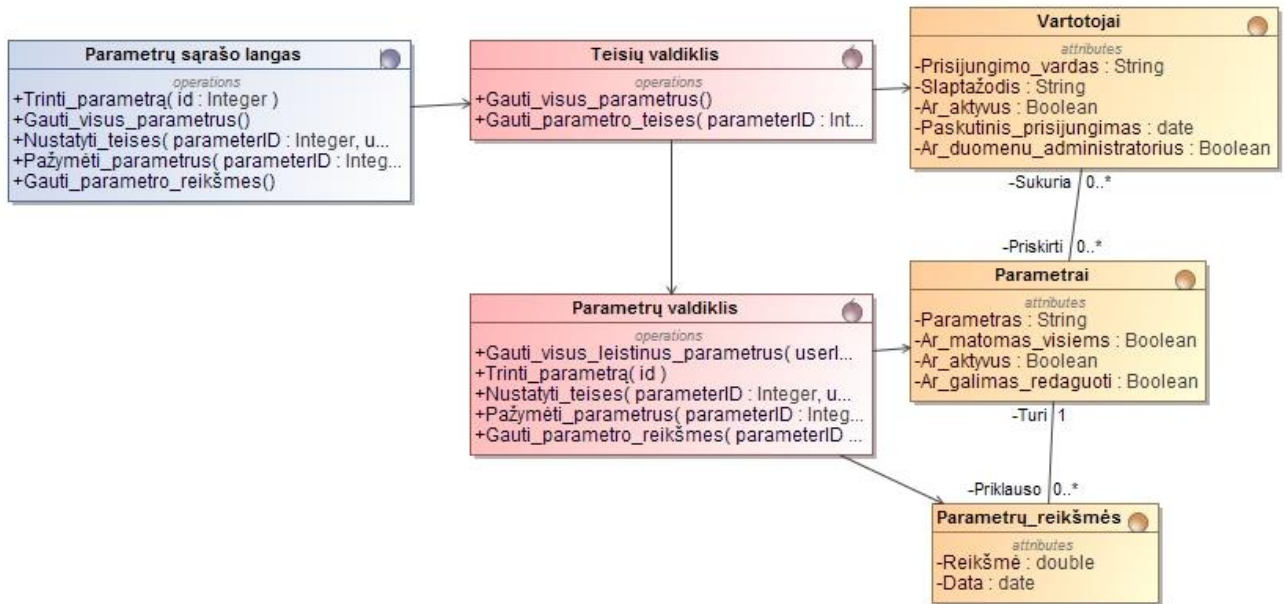


Pav. 17 Sistemos loginės architektūros diagrama

3.2. Reikalavimų analizės modeliai

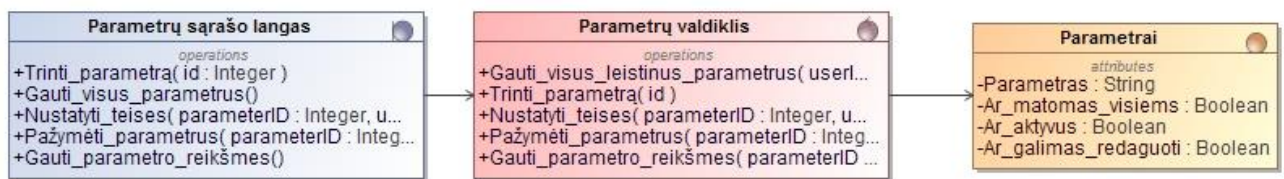
Šiame poskyryje pateiktos robastiškumo diagramos kiekvienam įgyvendinamam panaudojimo atvejui.

Pav. 18 iliustracijoje pateikta panaudojimo atvejo „Ištrinti parametą“ projekto klasių diagrama. Sistemos funkcija vykdoma parametų sąrašo lange, naudojami du valdikliai. Funkcijai įgyvendinti panaudojamos trys duomenų bazės esybės, jos reikalingos sistemos vartotojų ir jų parametų surišimui.



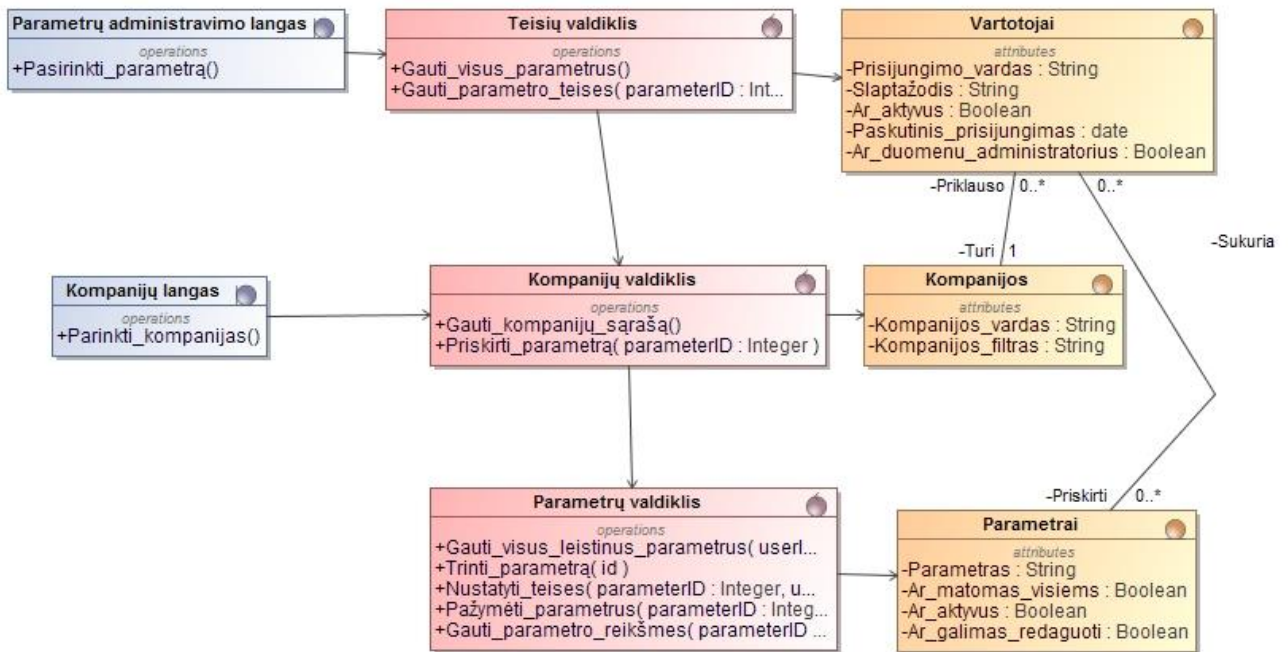
Pav. 18 „Ištrinti parametą“ PA projekto klasių diagrama

Pav. 19 iliustracijoje pateikta panaudojimo atvejo „Nustatyti parametro keitimo teises“ projekto klasių diagrama. Sistemos funkcija vykdoma parametų sąrašo lange, naudojamas vienas valdiklis. Funkcija atvaizduojama į parametų esybę ir yra keičiami parametro atributai, tokie kaip jo matomumas ar visuotinio redagavimo galimybės.



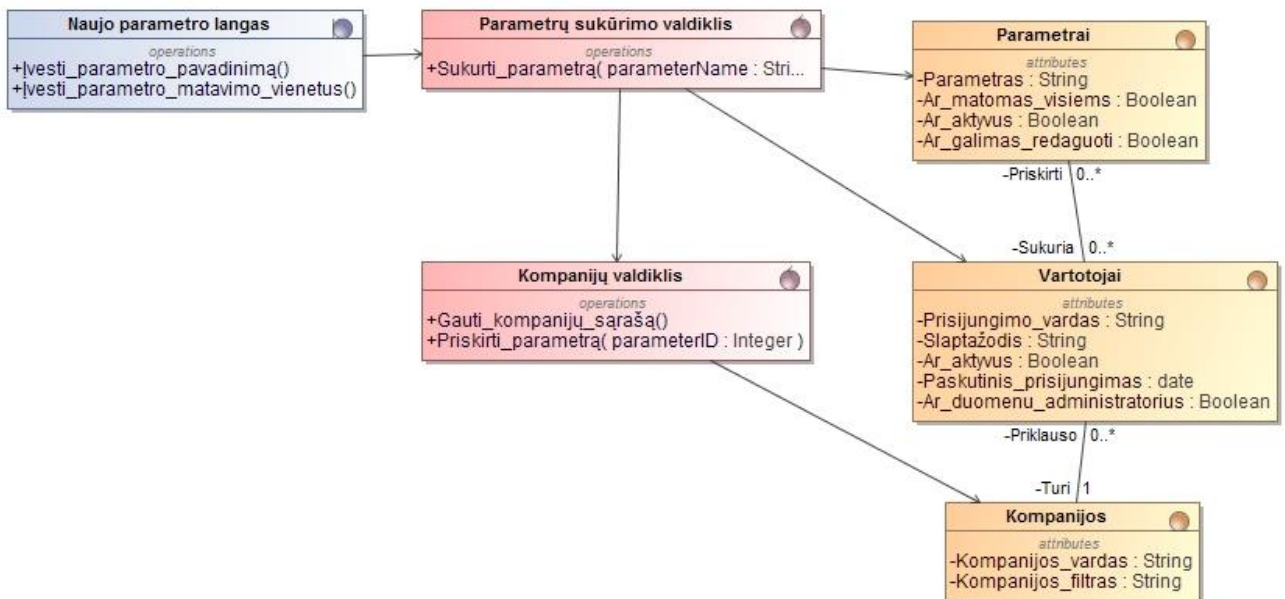
Pav. 19 „Nustatyti parametro keitimo teises“ PA projekto klasių diagrama

Pav. 20 iliustracijoje pateikta panaudojimo atvejo „Priskirti parametą vartotojams“ projekto klasių diagrama. Sistemos funkcija vykdoma per du navigavimo langus: parametų administravimo ir kompanijų. Funkcijos esmė yra sudaryti loginį sąryšį tarp vartotojo, kompanijos ir parametro, dėl to naudojamos šios trys esybės.



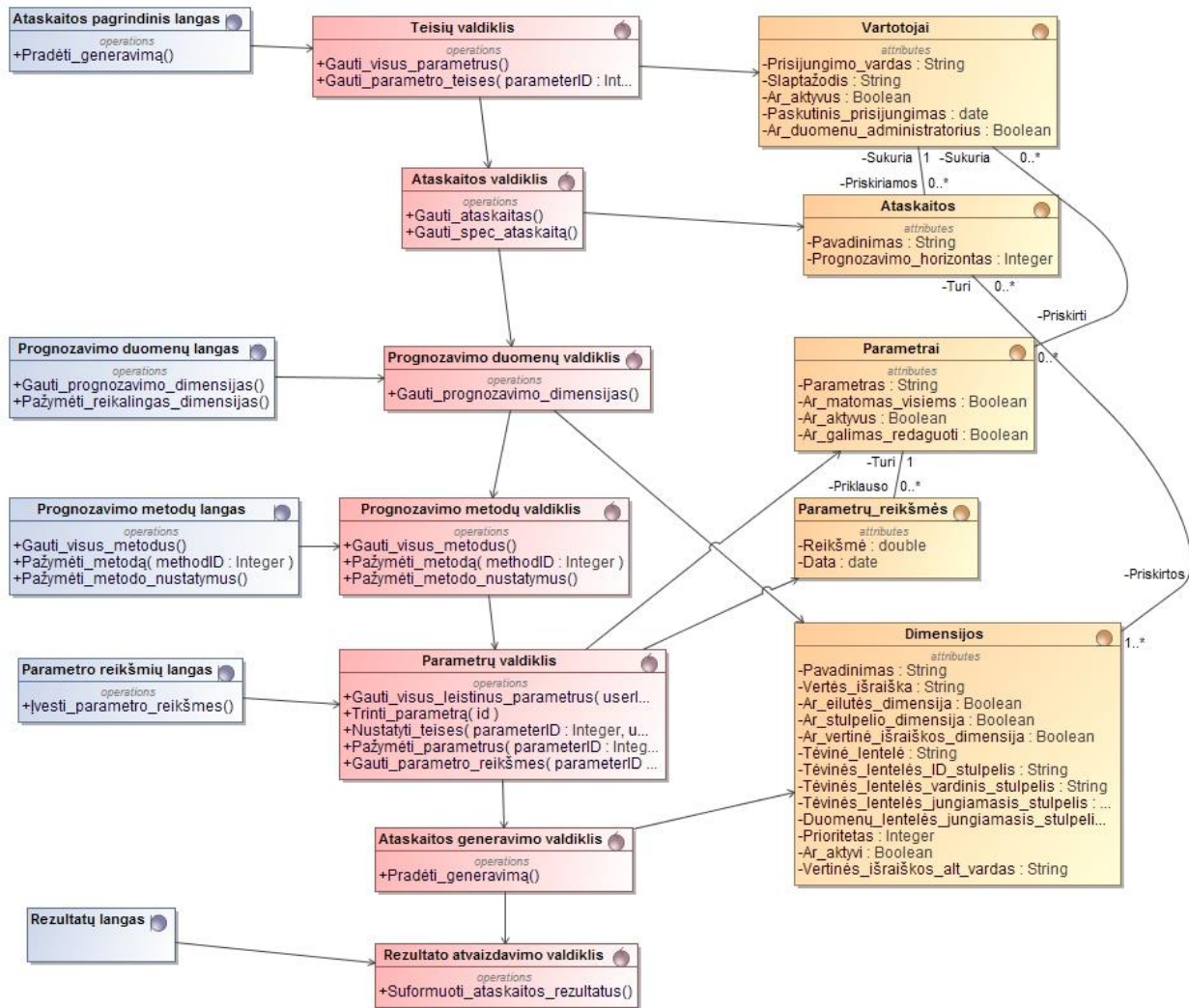
Pav. 20 „Priskirti parametą vartotojams“ PA projekto klasių diagrama

Pav. 21 iliustracijoje pateikta panaudojimo atvejo „Sukurti naują parametą“ projekto klasių diagrama. Sistemos funkcija vykdoma naujo parametro kūrimo lange. Kūrimo metu sudaromas ryšys duomenų bazėje tarp naujojo parametro, jį sukūrusio vartotojo ir kompanijų sąrašo, kurias atstovaujantys vartotojai gali naudoti naują parametą savo ataskaitose.



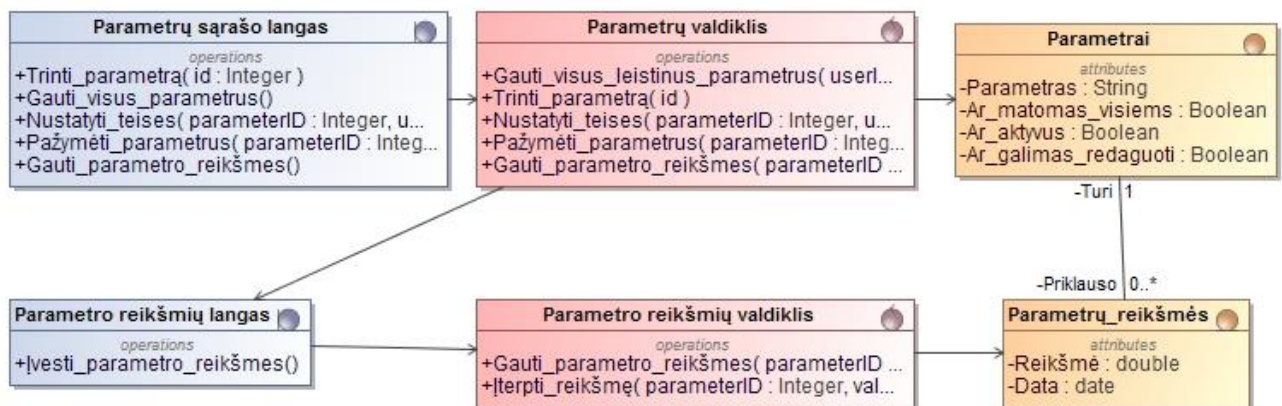
Pav. 21 „Sukurti naują parametą“ PA projekto klasių diagrama

Pav. 22 iliustracijoje pateikta panaudojimo atvejo „Sukurti prognozavimo ataskaitą“ projekto klasių diagrama. Šiame procese dalyvauja didžioji dalis sistemoje esančių valdiklių, langų ir esybių. Tai esminis sistemos funkcionalumas.



Pav. 22 „Sukurti prognozavimo ataskaitą“ PA projekto klasių diagrama

Pav. 23 iliustracijoje pateikta panaudojimo atvejo „Suvesti parametro reikšmes“ projekto klasių diagrama. Funkcija vykdoma parametro reikšmių lange, į kurį patenkama per parametru sąrašo langą ir parametru valdiklį. Suvestos reikšmės priskiriamos parametru reikšmių esybei ir įrašomos į duomenų bazę. Tam, kad būtų žinoma kuriam parametru kurios reikšmės priklauso sukuriama ryšys tarp parametru ir parametru reikšmių esybių.

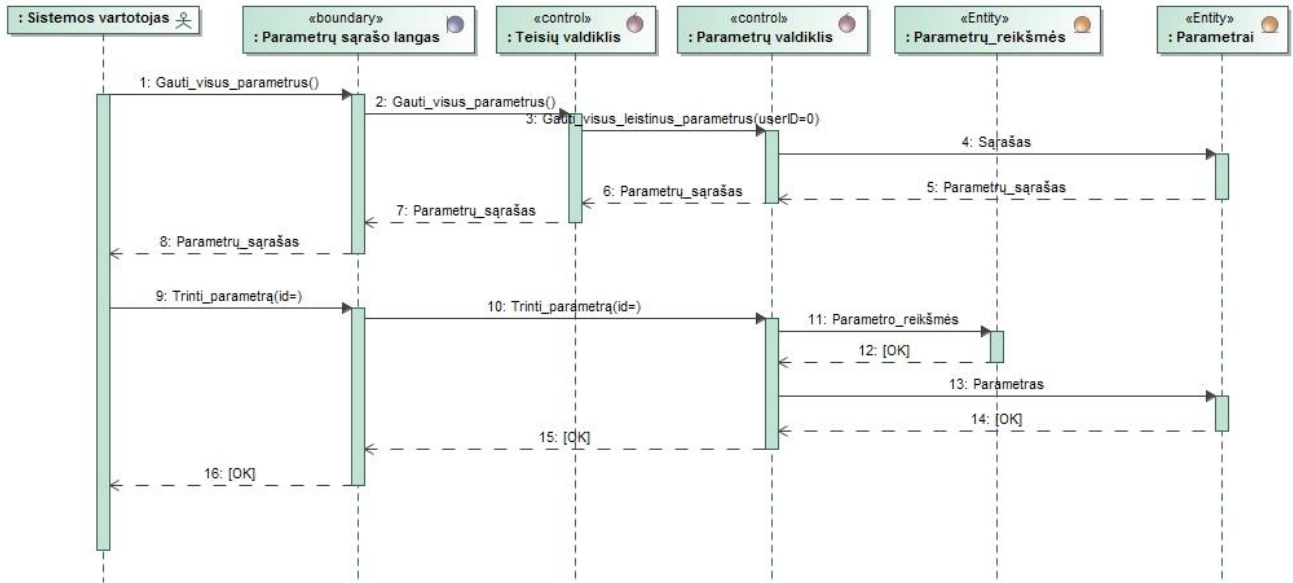


Pav. 23 „Suvesti parametro reikšmes“ PA projekto klasių diagrama

3.3. Sistemos elgsenos modeliai

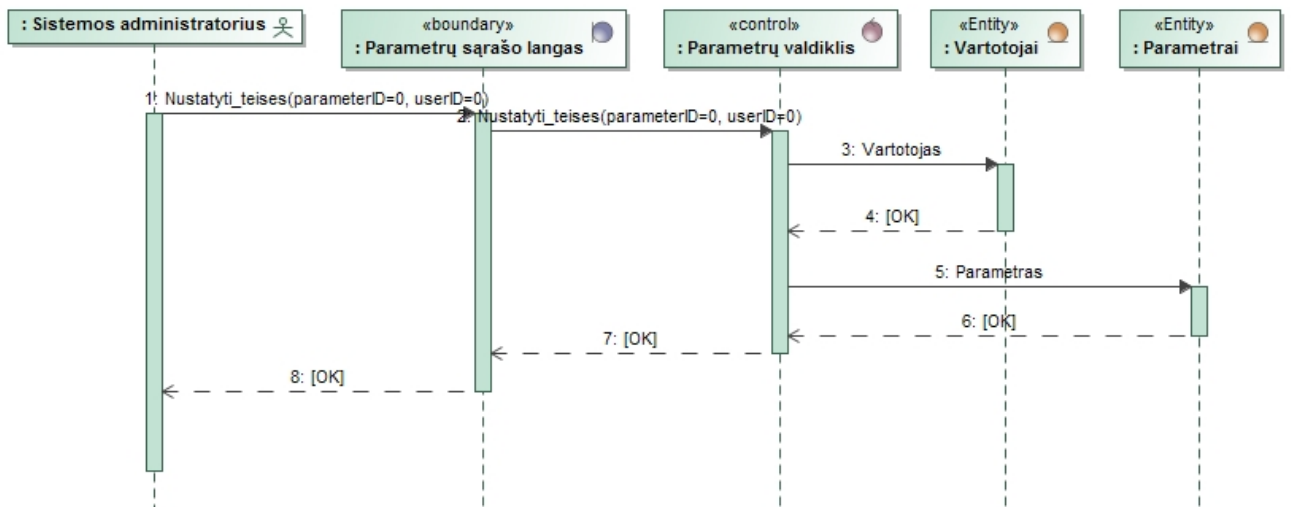
Šiame poskyryje pateikta visų realizuojamų panaudojimo atvejų sekų diagramos.

Pav. 24 iliustracijoje pateikta panaudojimo atvejo „Ištrinti parametą“ sekos diagrama. Pirmoji sekos dalis atsakinga už parametų sąrašo pateikimą vartotojui. Toliau vartotojas renkasi kurį iš ekrane matomų parametų norima ištrinti. Trynimasis vyksta dviem etapais: pirma ištrinamos parametro reikšmės, o toliau ištrinamas ir pats parametras.



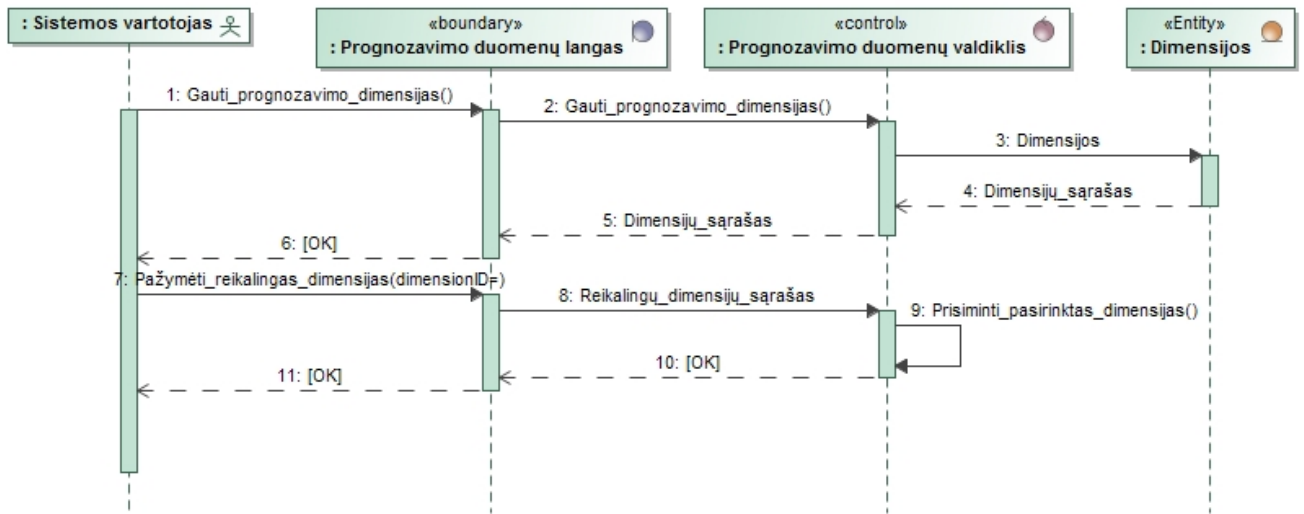
Pav. 24 „Ištrinti parametą“ PA sekų diagrama

Pav. 25 iliustracijoje pateikta panaudojimo atvejo „Nustatyti parametro keitimo teises“ sekos diagrama. Šioje sekoje naudojami du atributai: parametro ID ir vartotojo ID. Jie reikalingi tam, kad būtų žinoma kurio parametro teisės keičiamos ir kurio vartotojo atžvilgiu tai atliekama.



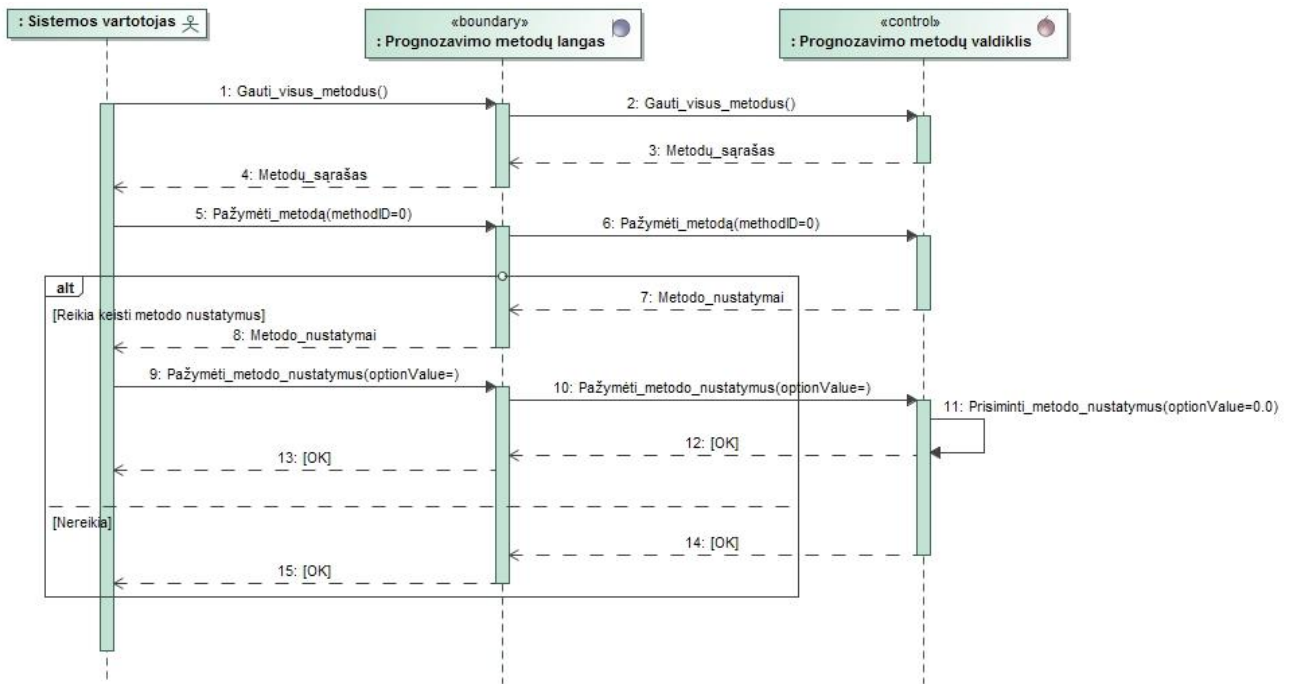
Pav. 25 „Nustatyti parametro keitimo teises“ PA sekų diagrama

Pav. 26 iliustracijoje pateikta panaudojimo atvejo „Nustatyti parametro keitimo teises“ sekos diagrama. Pirmasis funkcijos etapas yra atvaizduoti vartotojui prognozavimo dimensijų sąrašą. Toliau vartotojas matydamas šį sąrašą gali pasirinkti kurios dimensijos jam yra reikalingos. Pažymėtų dimensijų identifikatoriai laikinai užsaugojami prognozavimo duomenų valdiklyje tolesniam naudojimui.



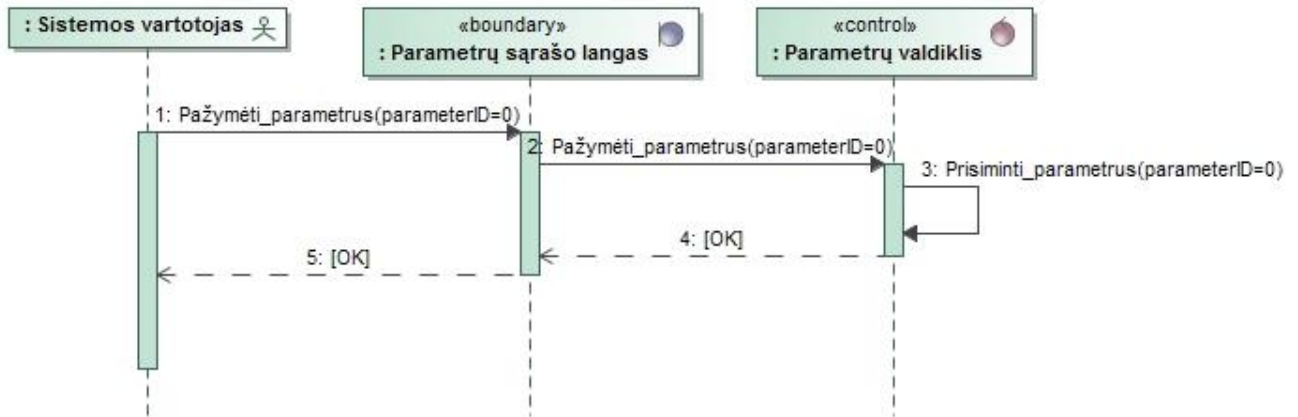
Pav. 26 „Pasirinkti prognozavimo duomenų turinį“ PA sekų diagrama

Pav. 27 iliustracijoje pateikta panaudojimo atvejo „Pasirinkti prognozavimo metodą“ sekos diagrama. Pirmasis žingsnis yra atvaizduoti galimų metodų sąrašą vartotojui. Toliau vartotojas pažymi norimą metodą ir, priklausomai nuo pažymėto metodo, suteikiama, arba ne, galimybė koreguoti metodo nustatymus. Jeigu galimybė suteikiama tai pasirinkti metodo nustatymai laikinai išsaugojami prognozavimo metodų valdiklyje tolesniam darbui.



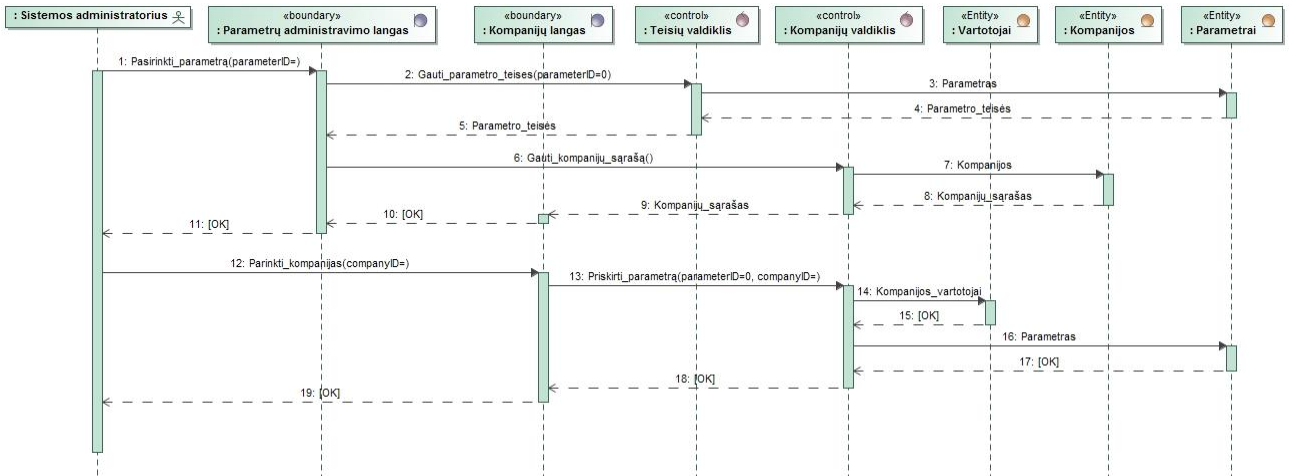
Pav. 27 „Pasirinkti prognozavimo metodą“ PA sekų diagrama

Pav. 28 iliustracijoje pateikta panaudojimo atvejo „Pasirinkti prognozavimo parametrus“ sekos diagrama. Pažymėtų parametrų identifikatoriai išsaugojami parametrų valdiklyje tolesniam naudojimui.



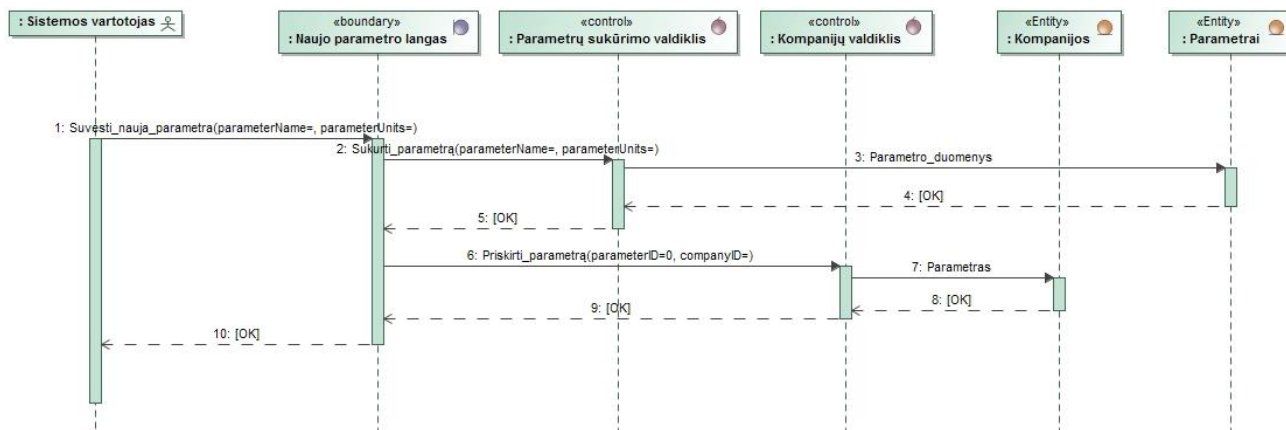
Pav. 28 „Pasirinkti prognozavimo parametrus“ PA sekų diagrama

Pav. 29 iliustracijoje pateikta panaudojimo atvejo „Priskirti parametą vartotojams“ sekos diagrama. Pirmajame etape sistema bando atvaizduoti kompanijų sąrašą sistemos administratoriui. Toliau administratorius iš šio sąrašo pasirenka kompanijas, kurių vartotojai turės galimybę naudoti atitinkamą parametą savo ataskaitose. Administratoriaus koreguota informacija išsaugojama vartotojų ir parametrų esybėse.



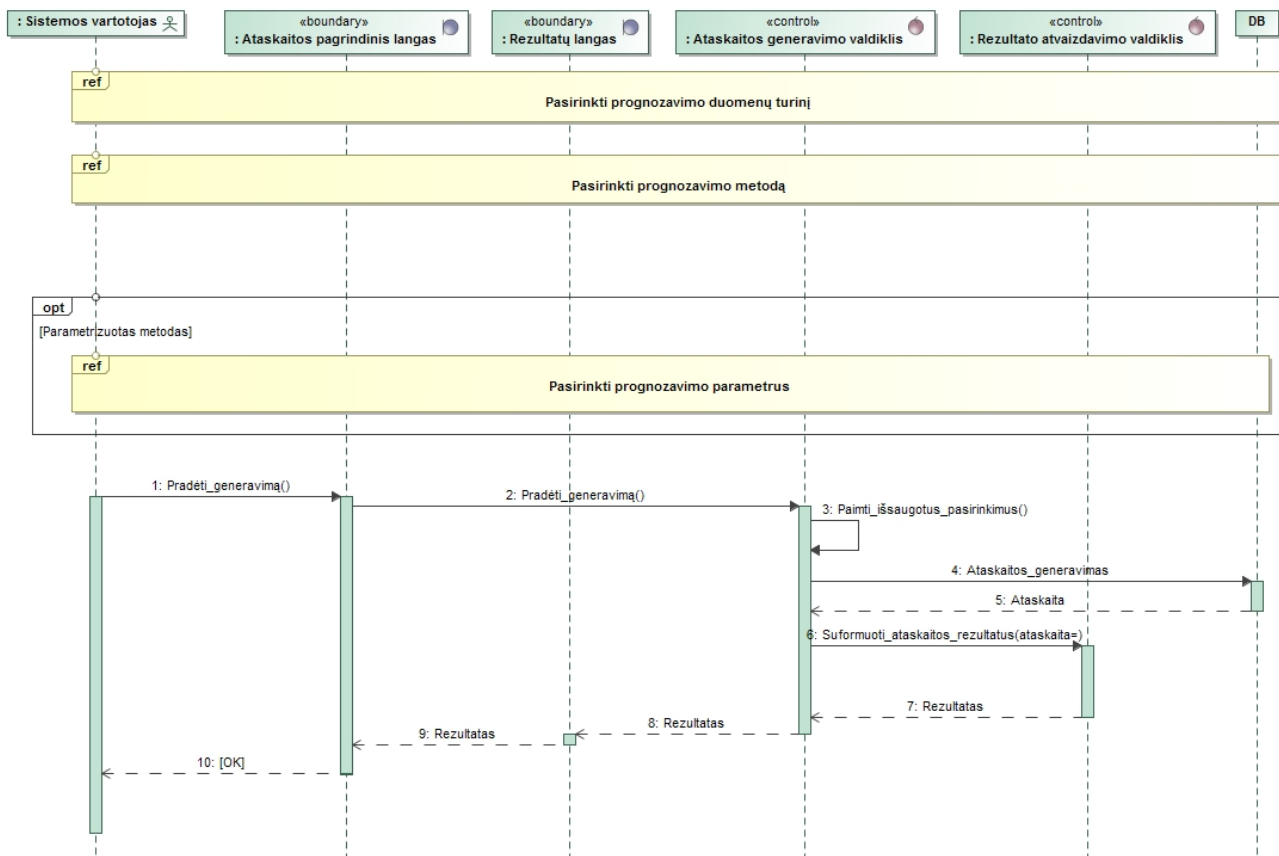
Pav. 29 „Priskirti parametą vartotojams“ PA sekų diagrama

Pav. 30 iliustracijoje pateikta panaudojimo atvejo „Sukurti naują parametą“ sekos diagrama. Parametro kūrimą galima išskaidyti į dvi mažesnes veiklas, tai būtų pačio parametro sukūrimas ir sekantis žingsnis – parametro priskyrimas kompanijų vartotojams. Jeigu parametras nebūtų priskiriamas nei vienam tradiciniam vartotojui, tai jo niekas negalėtų naudoti savo ataskaitose.



Pav. 30 „Sukurti naują parametą“ PA sekų diagrama

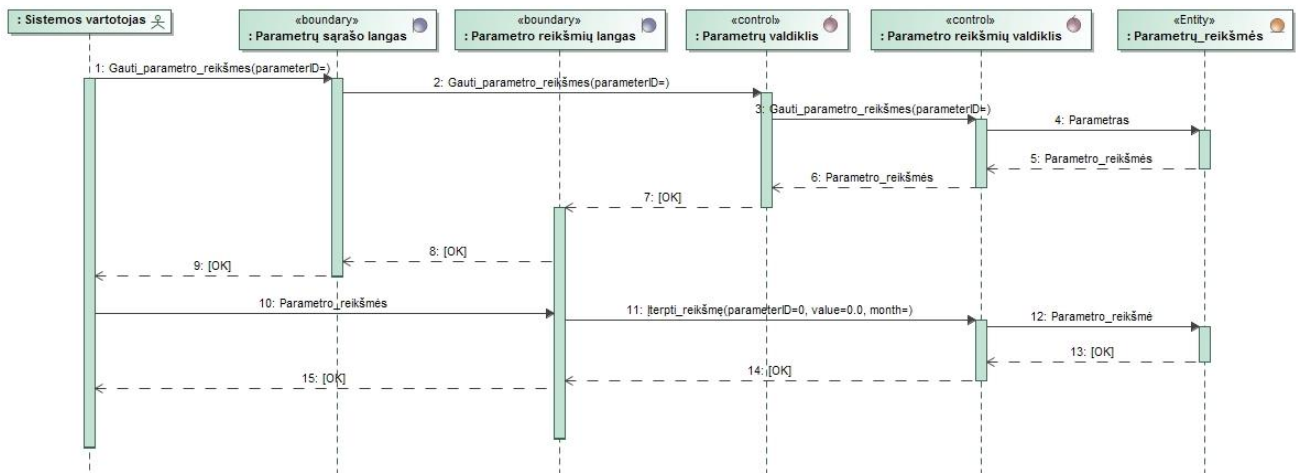
Pav. 31 iliustracijoje pateikta panaudojimo atvejo „Sukurti prognozavimo ataskaitą“ sekos diagrama. Seka apima anksčiau aprašytas ir pavaizduotas mažesnes sekas: prognozavimo duomenų turinio pasirinkimas, prognozavimo metodo pasirinkimas, prognozavimo parametrų pasirinkimas. Atlikus šiuos veiksmus jau turimi visi duomenys reikalingi ataskaitos generavimui. Bėlieka tik šiuos duomenis perduoti į duomenų bazės procedūrą atsakingą už ataskaitos generavimą, sulaukti rezultato ir jį atvaizduoti vartotojui.



Pav. 31 „Sukurti prognozavimo ataskaitą“ PA sekų diagrama

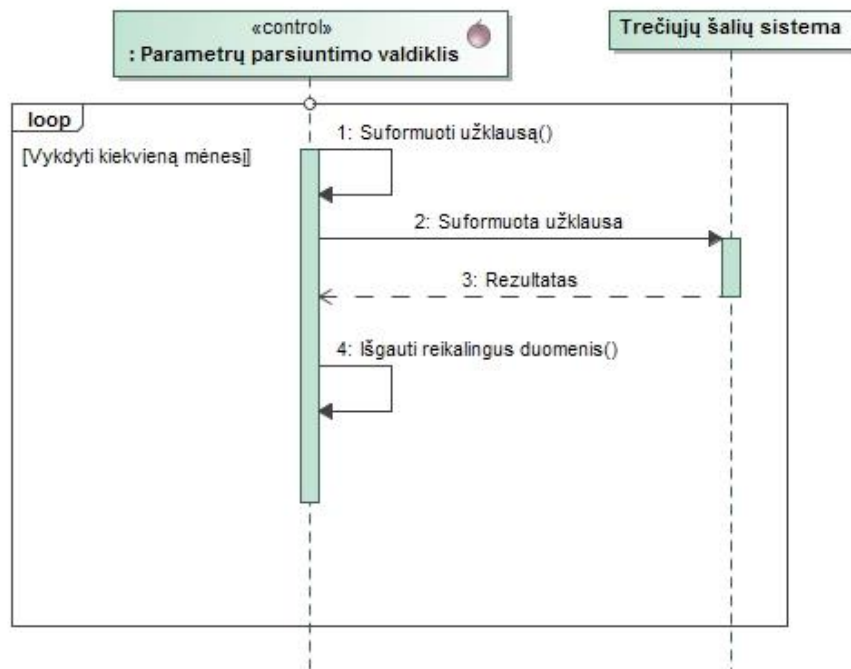
Pav. 32 iliustracijoje pateikta panaudojimo atvejo „Suvesti parametro reikšmes“ sekos diagrama. Pirmasis didesnis žingsnis yra gauti jau suvestas parametro ankstesnes reikšmes ir jas

atvaizduoti sistemos vartotojui. Tai atlikus vartotojas gali rinktis kurias iš jų norima redaguoti, arba, jeigu šios reikšmės dar nebuvo pildytos tai parinkti naujas.



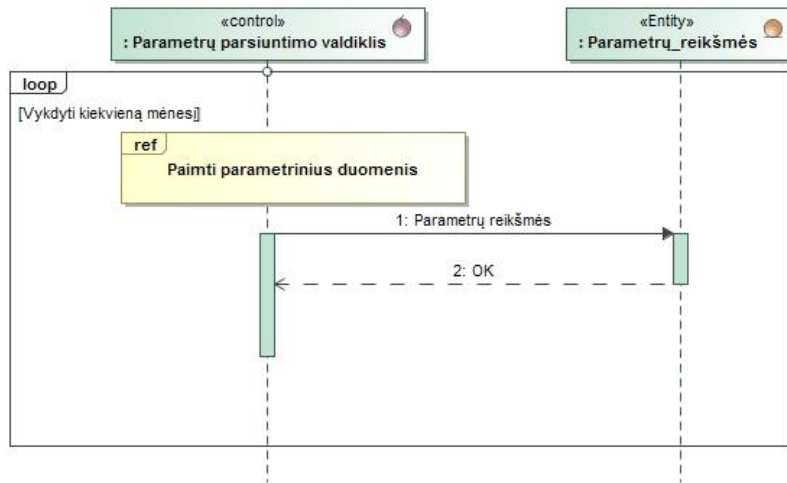
Pav. 32 „Suvesti parametro reikšmes“ PA sekų diagrama

Pav. 33 iliustracijoje pateikta panaudojimo atvejo „Paimti parametrinius duomenis“ sekos diagrama. Ši seka vykdoma periodiškai ir yra automatiškai iškvičiama (numatytu atveju – kiekvieną mėnesį). Iš pradžių suformuojama užklausa, kuri siunčiama į trečiųjų šalių sistemą (pavyzdžiui, meteorologinės stoties sistemą). Gavus užklauskos atsakymą, yra išgaunami tik reikalingi duomenys, jeigu jų gražinama per daug arba netenkina jų formatus.



Pav. 33 „Paimti parametrinius duomenis“ PA sekų diagrama

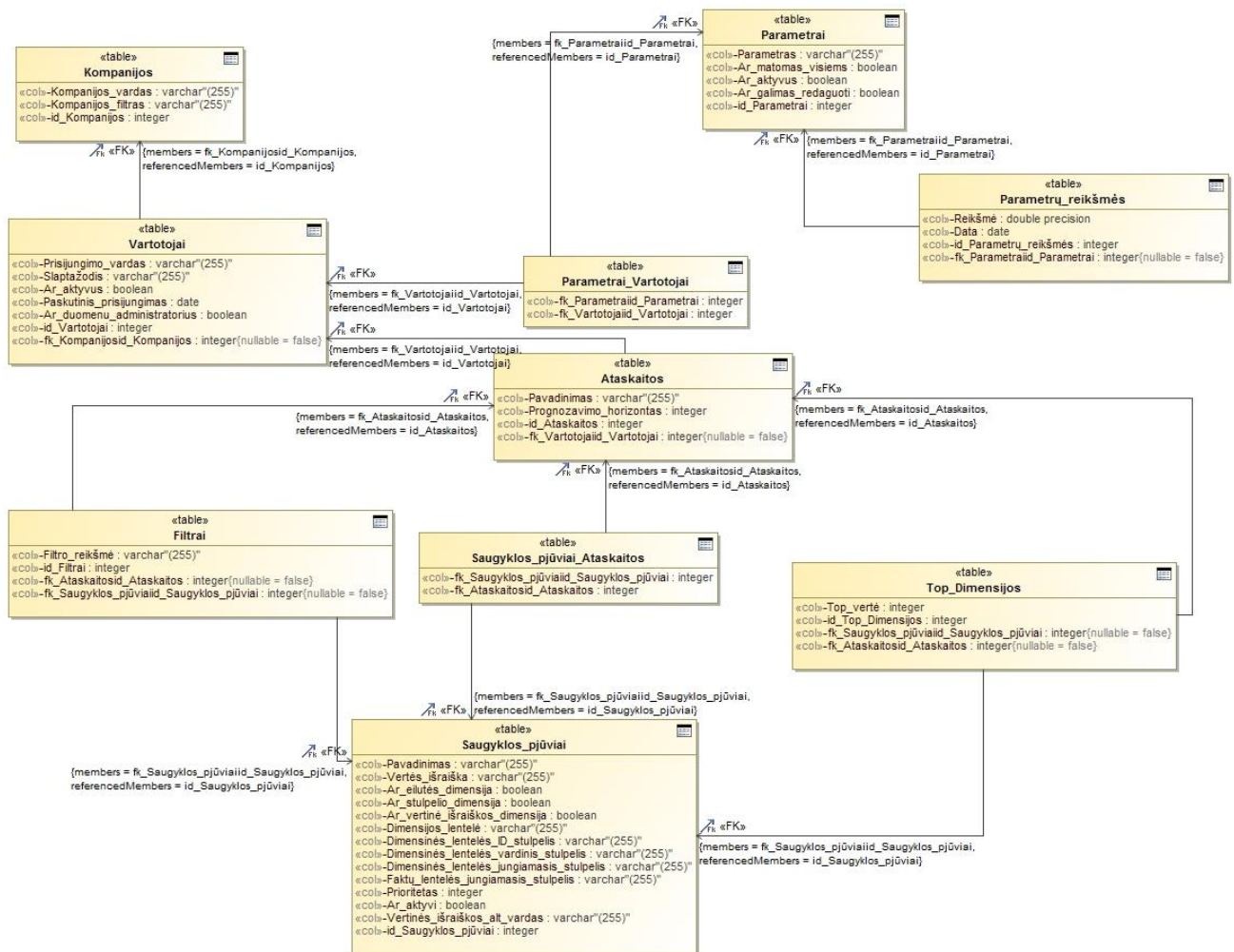
Pav. 35 iliustracijoje pateikta panaudojimo atvejo „Importuoti parametrinius duomenis“ sekos diagrama. Ši seka yra vykdoma kartu su funkcija „Paimti parametrinius duomenis“. Kai duomenys yra išgauti ir pertvarkyti, reikalingas jų importas į turimą duomenų bazę.



Pav. 34 „Importuoti parametrinius duomenis“ PA sekų diagrama

3.4. Duomenų bazės modelis

Šiame poskyryje pateiktas prognozavimo sistemos duomenų bazės modelis gautas iš dalykinės srities esybių klasių modelio transformavimo būdų. Duomenų bazės modelį jau galima naudoti fiziškai kuriant duomenų bazę, automatiškai sugeneruojant SQL užklausą.



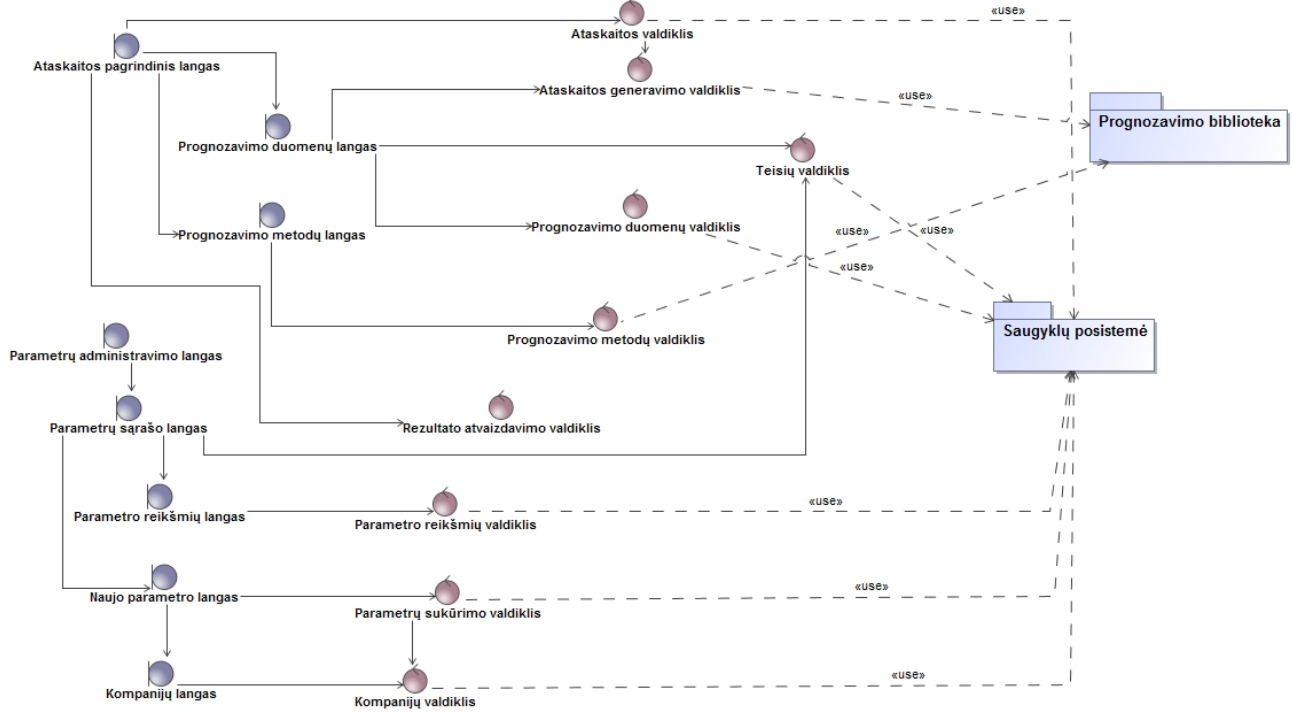
Pav. 35 Prognozavimo sistemos duomenų bazės diagrama

3.5. Sistemos klasių modelis

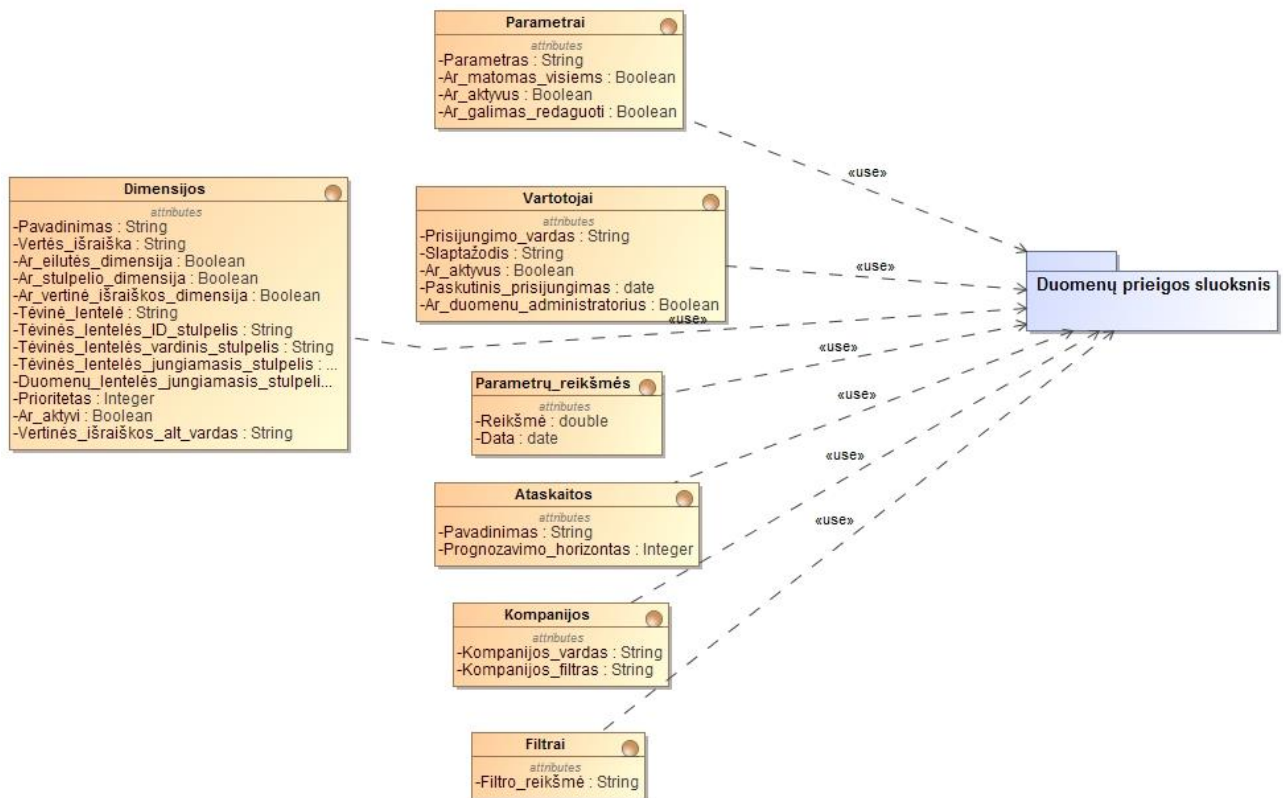
Šiame poskyryje pateiktos sistemos klasių diagramos ir klasių pasiskirstymas posistemėse.

Pav. 36 parodyta ataskaitos generavimo posistemės vidinės klasės ir jų sąveikos tarpusavyje bei su išoriniais komponentais. Kaip ir buvo matoma sistemos loginės architektūros modelyje, klasės bendrauja su dviem išoriniais komponentais: saugyklų posisteme ir prognozavimo biblioteka.

Pav. 37 pavaizduota saugyklų posistemės vidinės klasės ir jų sąveikos su išoriniais komponentais. Vienintelis matomas išorinis komponentas yra duomenų prieigos sluoksnis, per kurį gaunami ar siunčiami duomenys į duomenų bazę.



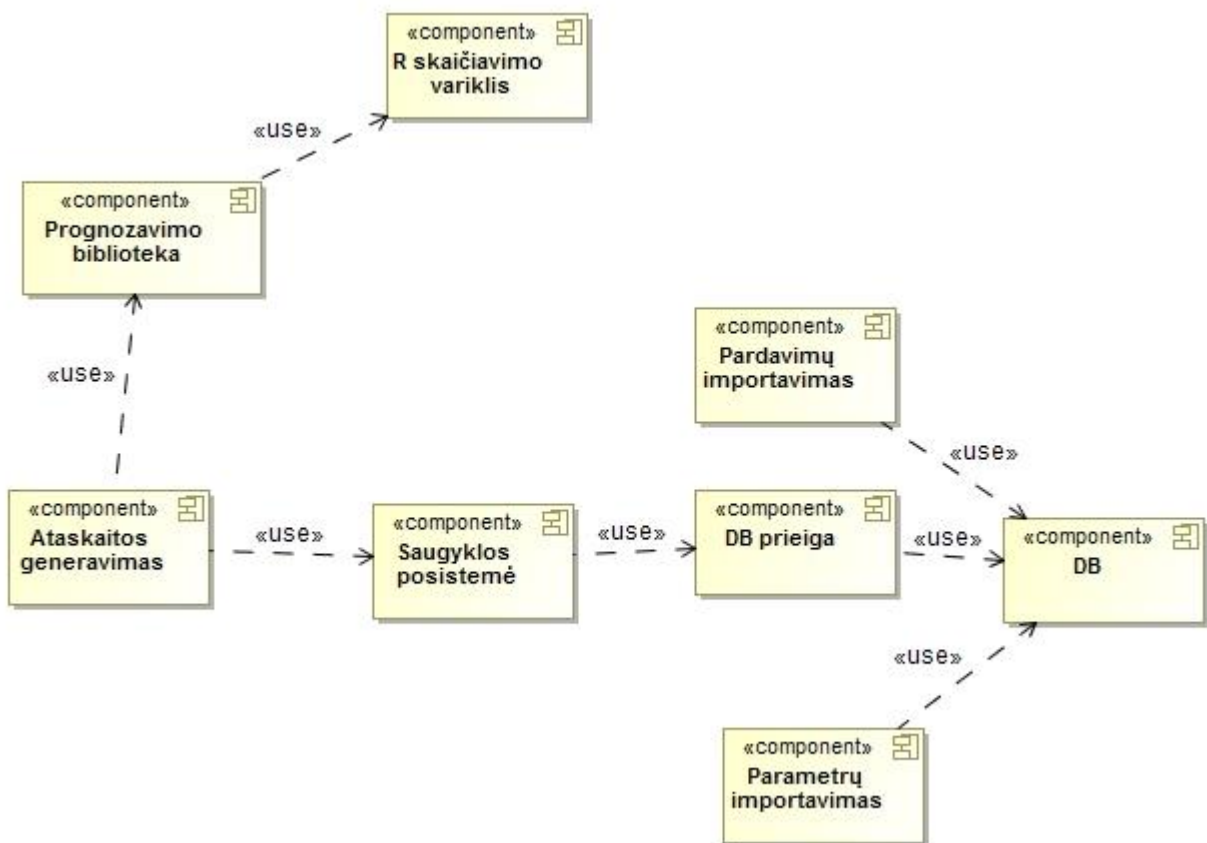
Pav. 36 Ataskaitos generavimo posistemės klasių diagrama



Pav. 37 Saugyklų posistemės klasių diagrama

3.6. Sistemos realizacijos modelis

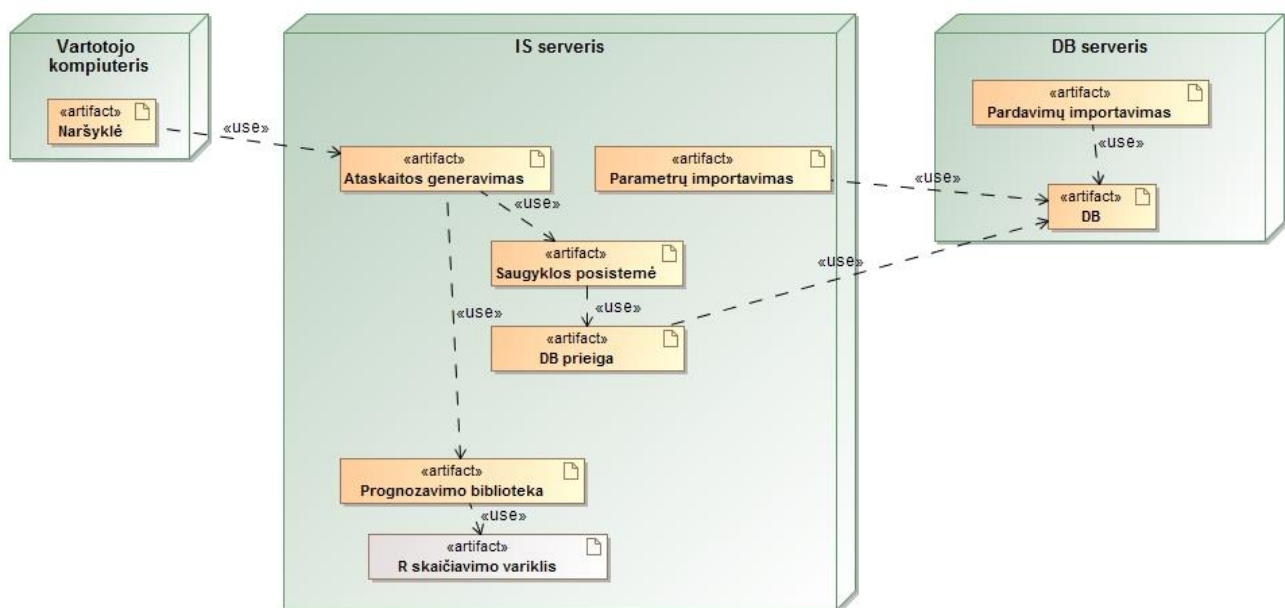
Pav. 38 pateikta pardavimų prognozavimo sistemos realizacijos diagrama. Visos sistemos dalys ir išoriniai komponentai išreikšti komponentais. Patys komponentai ir jų ryšiai atitinka sistemos loginės architektūros modelį.



Pav. 38 Prognozavimo sistemos realizacijos diagrama

3.7. Sistemos diegimo modelis

Pav. 39 iliustracijoje pateikta pardavimų prognozavimo sistemos diegimo diagrama. Pilkai pažymėtas artefaktas „R skaičiavimo variklis“ yra trečiųjų šalių pateikta biblioteka ir darbe nėra kuriamas, tik naudojamas žemo lygio skaičiavimams atlikti.



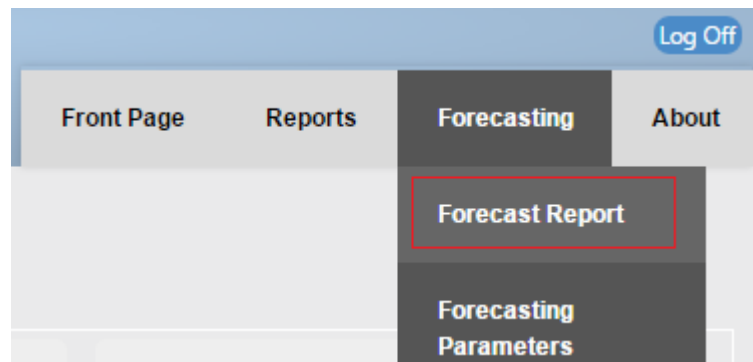
Pav. 39 Prognozavimo sistemos diegimo diagrama

4. SPRENDIMO REALIZACIJA IR SPRENDIMAS

4.1. Sprendimo realizacijos ir veikimo aprašas

Programavimo etape sukurta realizacija – internetinė sistema. Toliau aprašytas jos veikimas, išbandant prognozavimo pavyzdinį skaičiavimą.

Prisijungus prie sistemos, pirmasis uždavinys yra patekti į prognozavimo langą. Tai galima padaryti pagrindiniame meniu spaudžiant ant „Forecasting“ ir, iš atsivėrusio pagalbinio meniu, pasirinkti „Forecast Report“. Iliustracija Pav. 40.



Pav. 40 Patekimas į prognozavimo langą

Patekus į prognozavimo langą matomas vaizdas parodytas paveikslėlyje Pav. 41. Raudonai sužymėti visi pasirinkimai ir žemiau pateikti jų paaiškinimai.

1. Prognozavimo horizontas. T. y. kiek mėnesių į priekį bus atliekama prognozė. Numatytasis pasirinkimas yra trys mėnesiai. Šalia mėnesių pasirinkimo yra vartotoją perspėjantis pranešimas, jog kuo ilgesnis prognozavimo horizontas, tuo didesnė prognozavimo rezultato paklaida.
2. Kiek praities mėnesių duomenų rodyti rezultato lange, šalia prognozavimo rezultato. Istorinių rezultatų pateikimas greta, tikimasi, jog padės aiškiau matyti esančias ir būsimas tendencijas rezultatų lange. Numatytasis pasirinkimas yra 12 mėnesių.
3. Prognozavimo turinio filtrų pasirinkimas. Pasirinkimas neprivalomas, tačiau leidžia susiaurinti ir apibrėžti tik tam tikrų duomenų reikalingų vartotojui prognozę. Nepasirinkus filtrų bus prognozuojama visos rinkos bendra tendencija. Pasirinkus filtrus galima prognozuoti specifinio gamintojo ar produkto pardavimus. Pasirenkamų filtrų kiekis neribojamas.
4. Skaičiuojamos skaitinės išraiškos pasirinkimas. Galimas tik vienas pasirinkimas. Pasirinkimo variantai gali būti įvairių valiutų reikšmės ar ne piniginiai kiekiai.
5. Judėjimo tarp langų mygtukai. Paspaudus perėjimo mygtuką atliekama aktyvaus lango validacija.

1. Content 2. Forecast Type 3. Forecast Configuration

Show Instructions

1. Prediction horizon: 3 ?

2. How many history months do you want to see? 12 ?

Filters ?

- ATC1
- ATC2
- ATC3
- ATC4
- ATC5
- Brand
- Business Unit
- Client Type
- Cluster
- Corporation
- Dispensing Types

Values

- DOT
- Quantity
- Wholesale Price Euros

5. Previous Next

Pav. 41 Pagrindinis prognozavimo langas

Užpildžius pasirinkimus esančius pagrindiniame lange spaudžiamas mygtukas „Next“ ir pereinama į antrą langą, paveikslėlis Pav. 42. Iš pradžių jame yra tik vienas pasirinkimas – varnelė, kuri pažymima jeigu norima prognozuoti su parametrais. Varnelės nepažymėjus, bus naudojamas laiko eilučių prognozavimas. Tokiu atveju yra iš karto aktyvus „Finish“ mygtukas ir, jį paspaudus, pradedamas prognozavimo skaičiavimas. Pažymėjus varnelę, lango struktūra pakeičiama į matomą paveikslėlyje Pav. 43. Raudonai sužymėti pagrindiniai elementai ir žemiau pateikti jų aprašymai.

1. Parametrų pasirinkimo lentelė. Joje vartotojas mato visus jam priskirtus parametrus esančius sistemoje. Iš sąrašo jam reikia išsirinkti parametrus, kurie jo manymu gali lemti galutinį prognozavimo rezultatą ir tokius parametrus pažymėti. Būtina pažymėti bent vieną parametras. Maksimalus parametras kiekis neribojamas.
2. Naujo parametro sukūrimo mygtukas. Sukūrus parametras, jo pasirinkimas ir parametro reikšmės matomos tik jo savininkui. Taip pat, po parametro sukūrimo, jis tampa matomas lentelėje 1.
3. Parametro reikšmių langas. Jame matomi tik tie parametrai, kurie buvo pažymėti lentelėje 1. Reikšmių lange matomos reikšmės užkraunamos iš duomenų bazės, jeigu vartotojui jos atrodo netinkamos, jas galima pasikoreguoti. Reikšmės priskiriamos atitinkamiems mėnesiams, o mėnesių kiekio matymas atitinka pasirinkimą pagrindiniame prognozavimo lange elemente 3 (prognozavimo horizontas).

- Istorinių mėnesių matomumo pasirinkimas. Pažymėjus varnelę, parametro reikšmių langas papildys istoriniais įrašais, kuriuos taip pat, esant reikalui, galima koreguoti.
- Reikšmių užpildymo iš duomenų bazės mygtukas. Jį paspaudus, reikšmių lange iš naujo užkraunamos reikšmės iš duomenų bazės. Šis pasirinkimas reikalingas jeigu vartotojas koregavo parametrų reikšmes, bet galiausiai nusprendė, jog nori naudoti numatytąsias vertes.

1. Content 2. Forecast Type 3. Forecast Configuration

Check if there are any parameter's data which can influence forecast sales.

Previous Finish

Pav. 42 Prognozavimo antro žingsnio langas

1. Content 2. Forecast Type 3. Forecast Configuration

Check if there are any parameter's data which can influence forecast sales.

Show Instructions

Select parameters usage ?

Use in forecasting	Parameter
<input type="checkbox"/>	PSDF for medicines and medical aids
1. <input checked="" type="checkbox"/>	Temperature
<input type="checkbox"/>	GDP Growth
<input type="checkbox"/>	Country Population

2. Create New Parameter ?

4. Edit historical data 5. Load Values

Month	Temperature, °C
2015-10	6
3. 2015-11	2.5
2015-12	-2.5

Previous Next

Pav. 43 Prognozavimo antras žingsnis pasirinkus prognozavimą su parametrais

Kaip ir minėta, antro žingsnio užbaigimas įvykdomas spaudžiant „Next“ arba „Finish“, priklausomai nuo pasirinkimo pavaizduoto iliustracijoje Pav. 42 Prognozavimo antro žingsnio langas“. Jeigu spaudžiama „Finish“ – pradedamas skaičiuoti prognozavimo rezultatas, jeigu spaudžiama „Next“ – patenkama į trečio žingsnio langą (Pav. 44).

Trečiojo žingsnio langas skirtas parametrinio modelio pasirinkimui ir modelio nustatymams sureguliuoti. Žemiau pateikti raudonai sunumeruotų elementų, esančių Pav. 44, paaiškinimai.

1. Modelio tipo pasirinkimas. Galima rinktis tiesinę regresiją arba dirbtinius neuroninius tinklus. Duotame paveikslėlyje Pav. 44 pasirinkta tiesinė regresija, todėl matomi būtent tiesinės regresijos nustatymai.
2. „R Squared“ parametro tenkinanti riba. Jeigu apskaičiuoto modelio „R Squared“ parametras nesiekia šios ribos, modelis bus išmestas kaip nepakankamai tikslus ir vartotojui bus parodomas informuojantis pranešimas. „R Squared“ parametro reikšmės būna tarp 0 ir 1. Nulis sako, jog modelis nepaaiškina jokios paklaidos esančios apie duomenų vidurkį. Vienetas parodo, kad paaiškinama visa paklaida, t. y. prognozavimo kreivė tobulai sutampa su realiais istoriniais duomenimis. „R Squared“ reikšmė nebūtinai turi būti kuo didesnė ir, jeigu jina yra didelė, tai dar nereiškia, jog prognozė bus tikslinga [28]. Numatytoji „R Squared“ riba yra 0.4.
3. „P Value“ reikšmės tenkinanti riba. „R Squared“ yra skaičiuojama visam modeliui, „P Value“ yra skaičiuojama kiekvienam parametru, parinktam Pav. 43. Jeigu reikšmė yra didesnė už nustatytą ribą, tai parametras išmetamas kaip nereikšminis. Jeigu iš pasirinktų parametru visi peržengia šią ribą, tai nereikšminiu laikomas visas modelis ir vartotojui parodomas klaidos pranešimas. „P Value“ reikšmė gali svyruoti nuo 0 iki 1. Skaičius parodo tikimybę, jog parametras yra nereikšminis. Numatyta „P Value“ reikšmė yra 0.05.
4. Rezultatų lange atvaizduojamo grafiko patikimumo rėžių ribos. Patikimumo rėžiai parodo grafiko plotą, kuriame turėtų pakliūti prognozė su tam tikra tikimybe. Ribos gali būti nuo 0 iki 1 (nuo 0% iki 100%). Numatyta reikšmė yra 0.9 (90%) t. y. grafike bus rodomas plotas, kuriame prognozuojama reikšmė patenka 9 kartus iš 10.

1. Content 2. Forecast Type 3. Forecast Configuration

Change forecasting configuration ?

1. Model type: Linear regression

2. Model R Squared threshold: 0.4 ?

3. Parameter P Value threshold: 0.05 ?

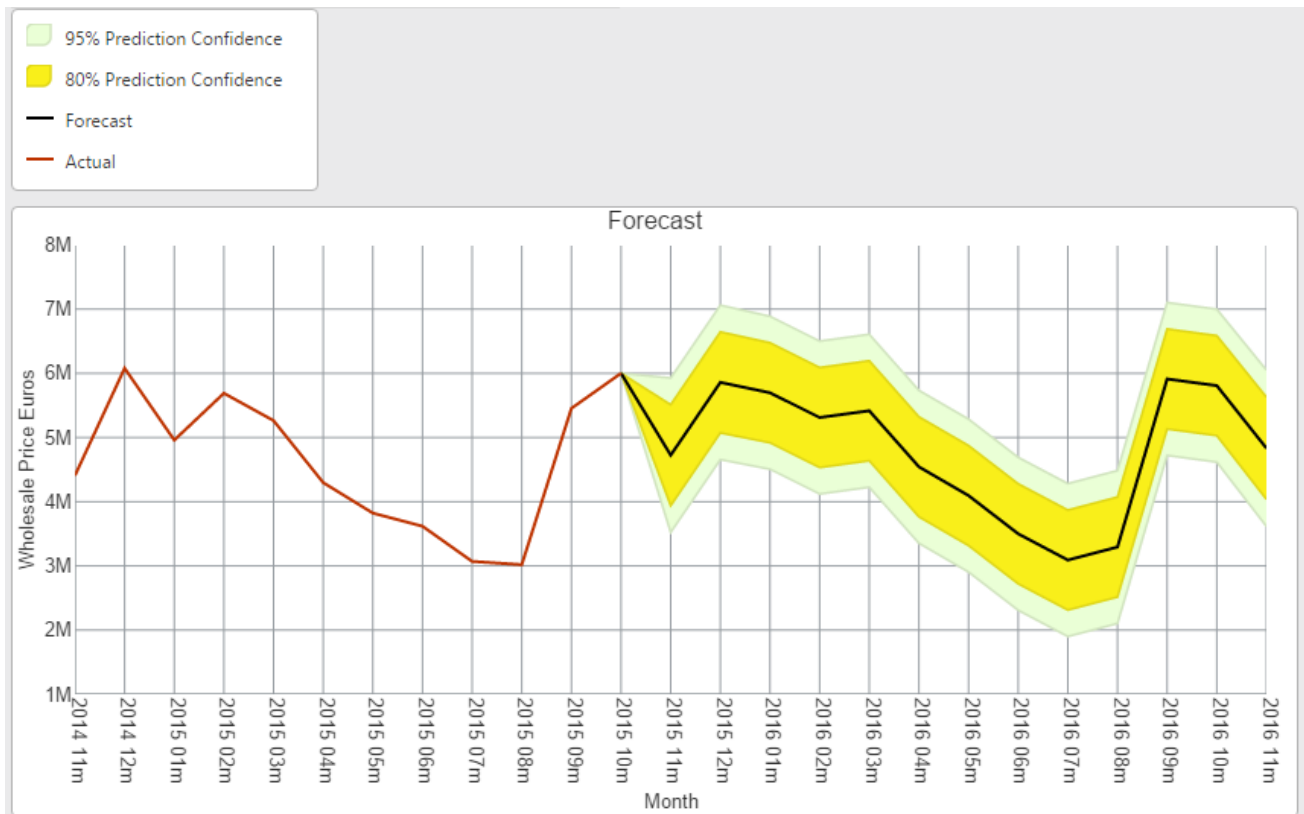
4. Prediction Interval Level: 0.9 ?

Previous Finish

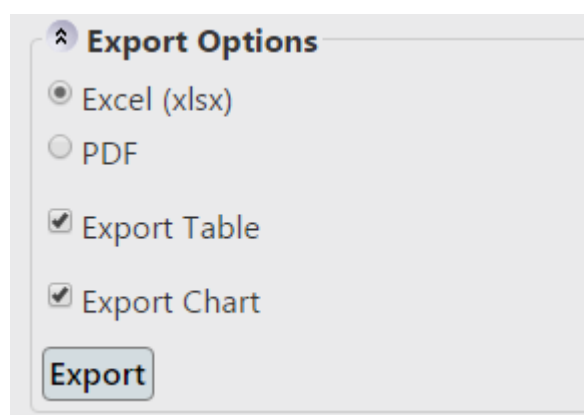
Pav. 44 Prognozavimo trečio žingsnio langas

Atlikus norimus nustatymų pakeitimus, galima spausti „Finish“ mygtuką ir bus pradėtas skaičiavimo procesas. Skaičiavimas gali užtrukti iki kelių minučių, priklausomai nuo istorinių duomenų kiekio ir prognozavimo horizonto, pasirinkto pirmame žingsnyje Pav. 41.

Po pasibaigusio skaičiavimo proceso, vartotojui parodomas rezultatų langas. Prognozė parodoma lentelė arba grafiku Pav. 45. Grafike matoma istoriniai duomenys, pažymėti raudona linija, ir prognozė su patikimumo intervalu. Norint prognozė išsaugoti tolesnėms peržiūroms, galima rezultatą eksportuoti į „Excel“ arba „Pdf“ formatus (Pav. 46).



Pav. 45 Prognozavimo rezultato grafikas



Pav. 46 Prognozavimo rezultato eksportavimo langas

4.2. Testavimo modelis, duomenys, rezultatai

4.1 ir 4.2 lentelėse pateikti reikalavimų testavimo aprašai. 4.1 lentelėje aprašytas nefunkcinių reikalavimų testavimas. Pirmame stulpelyje pateiktas reikalavimo pavadinimas. Daugiau informacijos apie reikalavimą galima rasti skyriuje Nefunkciniai reikalavimai 2.2 Nefunkciniai reikalavimai. Antrame stulpelyje pateikta testavimo eiga, atlikus aprašytus veiksmus turime galėti patvirtinti arba atmesti reikalavimo teisingą įgyvendinimą. Trečiame stulpelyje pateiktas testavimo rezultatas.

4.2 lentelėje aprašytas funkcinių reikalavimų testavimas. Pirmame stulpelyje įvardintas testuojamas elementas. Antrame stulpelyje pateikta testavimo eiga, atlikus aprašytus veiksmus turime galėti patvirtinti arba atmesti reikalavimo teisingą įgyvendinimą. Kiekvienam elementui yra aprašyti keli testavimo atvejai, nes šie elementai paprastai apima kelias smulkesnes sistemos funkcijas. Trečiame stulpelyje pateiktas testavimo rezultatas.

4.1 Nefunkcinių reikalavimų testavimo aprašas

Testuojamas reikalavimas	Testavimo eiga	Testavimo rezultatas
Elementų paaiškinimas	Atliekamas tradicinės prognozavimo ataskaitos kūrimas ir tikrinama ar yra neaiškių elementų, pasirinkimų ir ar prie jų yra pateikta papildoma informacija.	Šalia visų netradicinių elementų yra klaustuko simbolis, ant kurio užvedus parodoma papildoma informacija.
Generavimo laikas	Pasirenkama sudėtinga ataskaita su dideliu kiekiu istorinių duomenų ir atliekamas jos generavimas.	Po maždaug 5 minučių nutraukiamas ataskaitos generavimas ir parodomas pranešimas, jog privalom sumažinti duomenų kiekį.
Informacijos konfidencialumas	Sukuriam naują parametą ir prisijungiam kito vartotojo (iš skirtingos kompanijos) sąskaita.	Naujas parametras nėra matomas pašaliniam vartotojui.
Generavimo lygiagretumas	Su trimis skirtingais vartotojais prisijungiam iš skirtingų naršyklių ir vienu metu spaudžiam ant ataskaitos generavimo.	Ataskaitos generuojamos truputį ilgiau nei paprastai, bet generavimo laikas patenka į numatytas ribas ir visos trys ataskaitos atvaizduojamos rezultatų lange.

4.2 Funkcinių reikalavimų testavimo aprašas

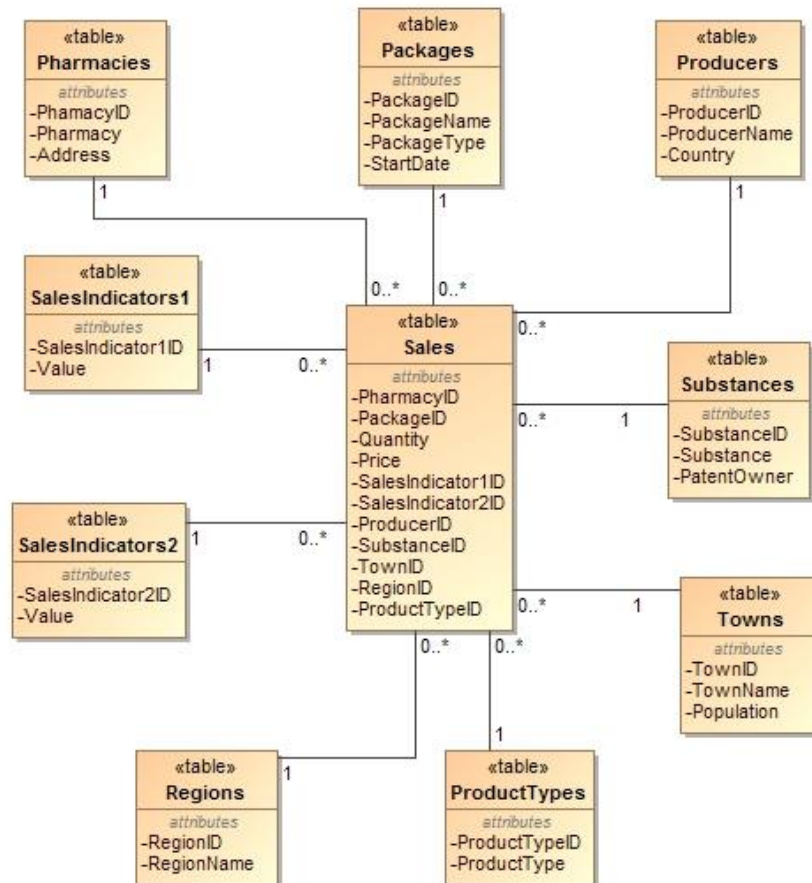
Testuojamas elementas	Testavimo eiga	Testavimo rezultatas
Pagrindinis prognozavimo langas, Pav. 41 Pav. 40	Iš karto spaudžiamas mygtukas „Next“, nepasirinkus jokių papildomų nustatymų.	Ijungiamas sekantis langas. Sistema paima numatytuosius nustatymus ir inicijuoja visos rinkos prognozavimą.
	Spaudžiama ant filtro kategorijos	Atsiranda papildomas langas su filtro kategorijos detalėmis.
	Spaudžiama ant filtro kategorijos detalės.	Esybė atsiranda naujoje pasirinktų filtrų grupėje.
	Spaudžiama ant vieno iš elemento esančio pasirinktų filtrų grupėje.	Filtro elementas išmetamas iš pasirinktų filtrų grupės.
Antras prognozavimo langas, Pav. 42	Keičiama parametų pasirinkimo varnelės būseną.	Priklausomai nuo varnelės būsenos, rodomas ar nerodomas parametų pasirinkimo langas.
	Pažymima parametų varnelė, bet nepasirenkamas nei vienas parametras.	Išmetamas klaidos pranešimas, jog privaloma pasirinkti bent vieną parametą.
	Nepažymima parametų varnelė	„Next“ mygtukas nerodomas, bet rodomas „Finish“ ir tai leidžia iš karto pradėti generavimo procesą.
	Spaudžiama ant „Create New Parameter“	Iššoka naujas langas su naujo parametro duomenų forma.
	Pasirinkto parametro lange spaudžiam ant „Edit historical data“	Išplečiamas reikšmių sąrašas, jį papildant istoriniais įrašais, kuriuos galima redaguoti.
	Pasirinkto parametro lange spaudžiam ant „Load Values“	Atitinkamo parametro reikšmių pakeitimai anuliuojami ir ant viršaus užrašomos numatytosios reikšmės iš duomenų bazės.
	Pasirinktas parametras ir bandoma pereiti į sekantį žingsnį, bet nesuvestos visos parametro reikšmės	Parodomas klaidos pranešimas, kad visos parametro reikšmės turi būti suvestos.
Trečias prognozavimo langas, Pav. 44	Keičiama prognozavimo konfigūracijos varnelės būseną	Priklausomai nuo varnelės būsenos, keičiamas nustatymų bloko matomumas
	Keičiamas prognozavimo modelio tipas.	Priklausomai nuo pasirinkto tipo, parodomi atitinkamo modelio nustatymai.
	Atsitiktinai keičiamos nustatymų reikšmės.	Jeigu reikšmės nepatenka į galimas ribas, parodomas klaidos pranešimas su informacija. Galimos reikšmių ribos parodomos užvedus ant šalia esančių aiškinamųjų simbolių.

5. EKSPERIMENTINIS SPRENDIMO TYRIMAS

5.1. Eksperimento dalykinė sritis

Magistro baigiamojo darbo metu sukurtas sprendimas panaudotas ir eksperimentuojamas įmonėje UAB „SoftDent“. Pagrindiniai įmonės tikslai yra palengvinti farmacijos bendrovių darbą, suteikti galimybę matyti ir analizuoti duomenis įvairiais pjūviais iš skirtingų verslo sričių. Tikslai įgyvendinami atliekant šiuos uždavinius: duomenų kaupimas, jų susistemimas ir importavimas į bendrą duomenų saugyklą, duomenų analizės programinių įrankių kūrimas ir plėtojimas, užsakomųjų rinkos ataskaitų ruošimas.

Iliustracijoje Pav. 47 parodyta „SoftDent“ įmonės pagrindinės duomenų saugyklos struktūros dalis. Struktūra atitinka žvaigždės schemą, kurios platesnis aprašymas buvo pateiktas skyriuje 2.3. Iliustracijoje Pav. 48 parodytas, metaduomenų apie žvaigždės schemą, lentelės realus turinys. Kaip ir buvo rašyta analizės dalyje, šios lentelės duomenys leidžia kurti dinamines SQL užklausas esant ataskaitos pareikalavimui. Metodikos platesnę analizę ir specifikaciją galima rasti skyriuose 2.3 ir 3.4 bei iliustracijose Pav. 8 ir Pav. 9.



Pav. 47 Eksperimentinės duomenų saugyklos dalis

ReportingDimensionID	Pavadinimas	Ar_Vertines_Išraiškos_Dimensija	Dimensijos_Lentelė	Faktu_Lenteles_Jungiamasis_Stulpelis	Dimensines_Lenteles_ID_Stulpelis	Dimensines_Lenteles_Jungiamasis_Stulpelis	Dimensines_Lenteles_Vardinis_Stulpelis	
1	1	Month	0	FullSales	SalesDate_Month	NULL	NULL	SalesDate_Month
2	2	Trimester	0	FullSales	SalesDate_Trimester	NULL	NULL	SalesDate_Trimester
3	3	Quarter	0	FullSales	SalesDate_Quarter	NULL	NULL	SalesDate_Quarter
4	6	Pharmaceutical Form	0	PharmaceuticalForms	PharmaceuticalFormID	PharmaceuticalFormID	PharmaceuticalFormID	PharmaceuticalForm
5	7	Package	0	Packages	PackageID	PackageID	PackageID	Package
6	8	Substance	0	Substances	SubstanceID	SubstanceID	SubstanceID	Substance
7	9	Producer	0	Producers	ProducerID	ProducerID	ProducerID	Producer
8	10	Corporation	0	Corporation	CorporationID	CorporationID	CorporationID	Corporation
9	11	Product	0	Products	ProductID	ProductID	ProductID	Product
10	12	ATC1	0	ATC1	ATC1ID	ATC1ID	ATC1ID	Caption
11	21	Cluster	0	PharmacyIndicators2	PharmacyIndicator2ID	PharmacyIndicator2ID	PharmacyIndicator2ID	PharmacyIndicator2
12	22	Subcluster	0	PharmacyIndicators1	PharmacyIndicator1ID	PharmacyIndicator1ID	PharmacyIndicator1ID	PharmacyIndicator1
13	23	District	0	Districts	DistrictID	DistrictID	DistrictID	District
14	24	Region	0	Regions	RegionID	RegionID	RegionID	Region
15	25	Quantity	1	FullSales	Quantity	NULL	NULL	NULL
16	26	Wholesale Price LTL	1	FullSales	Quantity*Price5 AS wp	NULL	NULL	NULL
17	27	Retail Price	1	FullSales	Quantity*Price2 AS rp	NULL	NULL	NULL
18	28	Wholesale Price Euros	1	FullSales	Quantity*Price6 AS wpe	NULL	NULL	NULL
19	36	Dispensing Types	0	PackageTypes	PackageTypeID	PackageTypeID	PackageTypeID	Package Type
20	37	Week	0	FullSales	SalesDate_Week	NULL	NULL	SalesDate_Week
21	38	Brand	0	Brands	BrandID	BrandID	BrandID	Brand
22	41	DOT	1	FullSales	DOT	NULL	NULL	NULL
23	42	Client Type	0	PharmacyTypes	PharmacyTypeID	PharmacyTypeID	PharmacyTypeID	Pharmacy Type
24	47	Sales Type	0	SalesIndicators1	SalesIndicator1ID	SalesIndicator1ID	SalesIndicator1ID	SalesIndicator1

Pav. 48 Metaduomenų lentelės turinys

5.2. Eksperimento planas

Eksperimentas vykdomas naudojantis farmacijos įmonių vaistų pardavimų duomenimis. Turima duomenų aibė yra nuo 2008 metų pradžios iki 2015 metų pabaigos, tai sudaro apie 300 milijonų faktinių įrašų. Kiekvienas faktinis įrašas turi nemažai informacijos apie save: pirkimo vieta, produkto informacija ir pan. Tai leidžia pardavimus analizuoti ir prognozuoti juos skirstant į grupes, pavyzdžiui, receptinius vaistus, maisto papildus, vaistus nuo širdies ligų ir kitas grupes. Prognozuojant pardavimus, 2015 metų antras pusmetis (kai kuriais atvejais visi 2015 metai) pasirenkamas kaip fiktyvusis prognozuojamas laikotarpis, t. y. prognozuojama atgaline data nuo 2015 metų antro pusmečio pradžios, tai leis palyginti prognozavimo rezultatą su realiais skaičiais ir įvertinti modelius bei paklaidas. Rinkos pardavimai skaičiuojami ir prognozuojami naudojant bendrą piniginę vertę eurais. Toliau esančiuose poskyriuose aprašyti skirtingų modelių eksperimento saviti bruožai.

Eksperimento metu prognozavimo rezultatai gaunami naudojant magistro baigiamojo darbo metu sukurtą sprendimą. Prognozavimo rezultatai tolesnei analizei perkeliama į „RStudio“ programinį paketą, kuriame skaičiuojamos paklaidos ir palyginami rezultatai. Paklaidų apskaičiavimui naudojama R programavimo kalbos funkcija *accuracy()*, kuri gali apskaičiuoti įvairius paklaidų įverčius: vidutinės kvadratinės paklaidos kvadratinė šaknis (RMSE), vidutinė absoliuti procentinė paklaida (MAPE) ir kt.

5.2.1. Tiesinės regresijos ir jos parametru įvertinimas

Norėdami įvertinti tiesinės regresijos modelį pirmiausia turime pasirinkti naudojamus parametrus. Šiame eksperimente naudojami tokie parametrai: vidutinė mėnesio oro temperatūra, metinis BVP augimas procentais, metinė Lietuvos populiacija, metinis PSDF (Privalomojo sveikatos draudimo fondas) biudžetas. Tiesinės regresijos paskaičiavimas atliekamas naudojantis R kalbos *lm()* funkcija. Sudaryto modelio korektiškumas vertinamas ne tik pagal rezultato paklaidą, bet ir pagal modelio ir parametru aktualumo koeficientus „R-Squared“ bei „P-Value“, platesnis jų paaiškinimas pateiktas skyriuje Sprendimo realizacijos ir veikimo aprašas 4.1.

5.2.2. Laiko eilučių metodų paklaidų palyginimas ir įvertinimas

Lyginami skirtingi R programavimo kalboje galimi laiko eilučių metodai ir kurie buvo integruoti į sukurtą įrankį:

- *Auto.arima()* . Geriausio ARIMA modelio automatinis paskaičiavimas, panaudojant AIC, BIC kriterijus.
- *Ets()*. Eksponentinio glodinimo modelis
- *Bats()*. Eksponentinio glodinimo modelis su *Box-Cox* transformacija, ARMA paklaidomis, trendo ir sezoniškumo komponentėmis.
- *HoltWinters()*.

5.2.3. Neuroninių tinklų įvertinimas

Lyginami neuroniniai tinklai su skirtingu kiekiu vidinių neuronų. Vėliau gauti rezultatai lyginami šalia tiesinės regresijos ir laiko eilučių metodų. Neuroninių tinklų paskaičiavimas atliekamas naudojant R programavimo kalbos funkciją *mlp()*.

5.3. Eksperimento rezultatai

Prognozuojamų duomenų aibė pasirenkama visa vaistų rinka, jos neskaidant į terapines grupes ar pardavimo teritorijas. Paveiksle Pav. 49 pateikti realūs pardavimai, kurie buvo 2015 metais pamėnesiui. Toliau esančiuose poskyriuose pateikti šio laikotarpio prognozavimo rezultatai atitinkamais metodais ir paklaidų išskyrimas.

Mėnesis	Rinkos vertė, Eurai
2015 01m	47375233
2015 02m	46508574
2015 03m	52339977
2015 04m	50390334
2015 05m	47843510
2015 06m	51003520
2015 07m	50459260
2015 08m	45645179
2015 09m	51290309
2015 10m	54935865
2015 11m	54477414
2015 12m	57952269

Pav. 49 Tiksliniai vaistų rinkos duomenys 2015 metais

5.3.1. Tiesinės regresijos eksperimento rezultatai

Model R-Squared value - 0.4386 ?

Parameter	P-Value ?
PSDF_for_medicines_and_medical...	0.00055
Temperature	0.00028
GDP_Growth	0.09255
Country_Population	0.00002

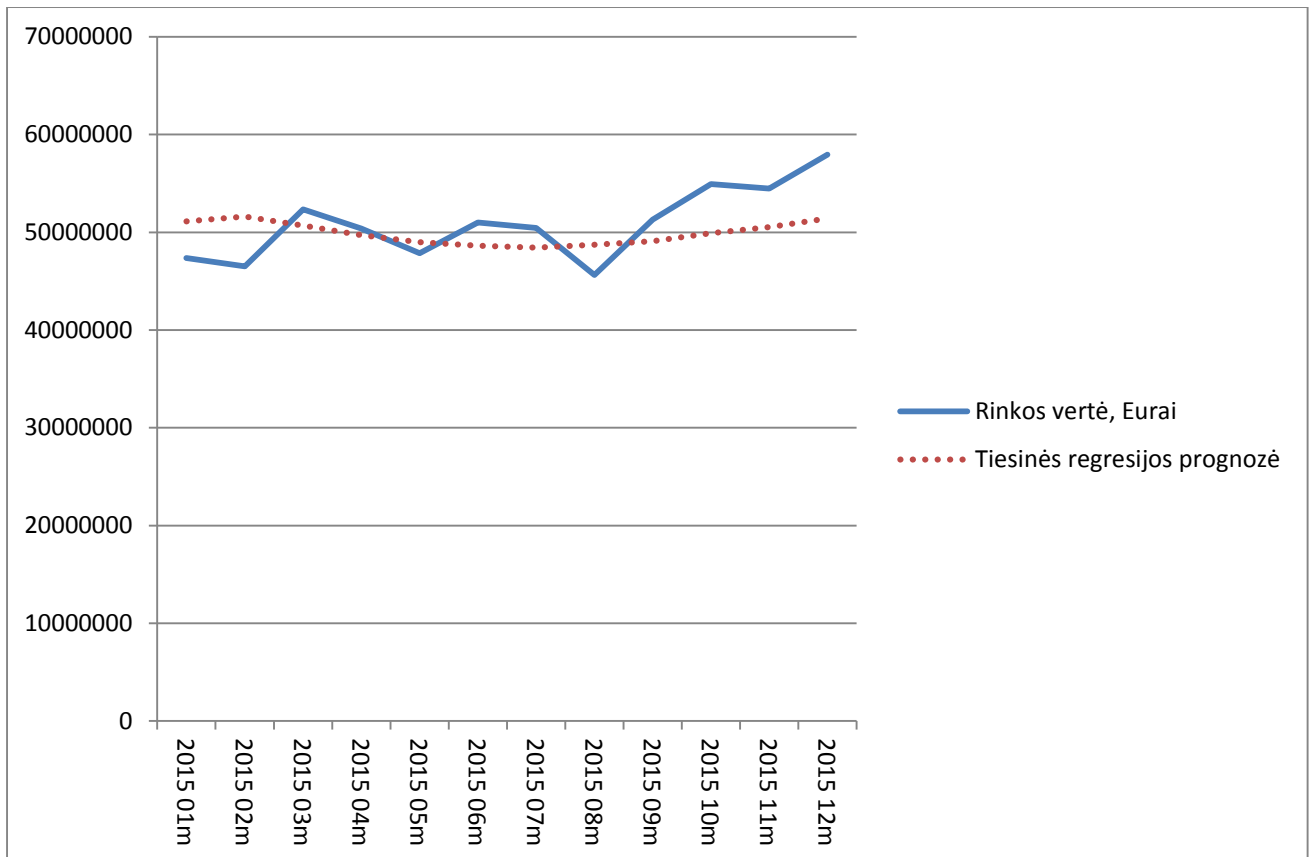
Pav. 50 Tiesinės regresijos rezultatų apžvalga

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
946971	3561501	3125583	1.42593	6.09502	0.33986	0.9534244

Pav. 51 Tiesinės regresijos paklaidų įverčiai

Pav. 50 iliustracijoje pateikta tiesinės regresijos modelio ir parametrų apžvalga. Tiesinės regresijos modelis sukurtas panaudojant šiuos parametrus: vidutinė mėnesio oro temperatūra (*Temperature*), metinis BVP augimas procentais (*GDP_Growth*), metinė Lietuvos populiacija (*Country_Population*), metinis PSDF (Privalomojo sveikatos draudimo fondas) biudžetas (*PSDF_for_medicines_and_medical*). Metinis BVP augimas buvo vienintelis parametras, kuris viršijo rekomenduojamą „P-Value“ reikšmę 0.05 ir dėl šios priežasties parametras iš galutinio modelio išmetamas. Viso modelio „R-Squared“ koeficientas siekė 0.4386, tai yra vos daugiau už apibrėžtą ribą 0.4, tačiau to pakanka, kad galėtume teigti, jog galutinis modelis yra aktualus ir korektiškas.

Pav. 51 lentelėje pateiktos tiesinės regresijos modelio paklaidos. Pav. 52 grafike parodytas prognozės rezultato palyginimas su realiais pardavimais. Matome, jog bendrą metų tendenciją tiesinės regresijos modelis nuprognozuoja, tačiau smulkesnių kasmėnesinių pakitimų neužfiksuoja.



Pav. 52 Tiesinės regresijos prognozės ir realių pardavimų palyginimas

5.3.2. Laiko eilučių eksperimento rezultatai

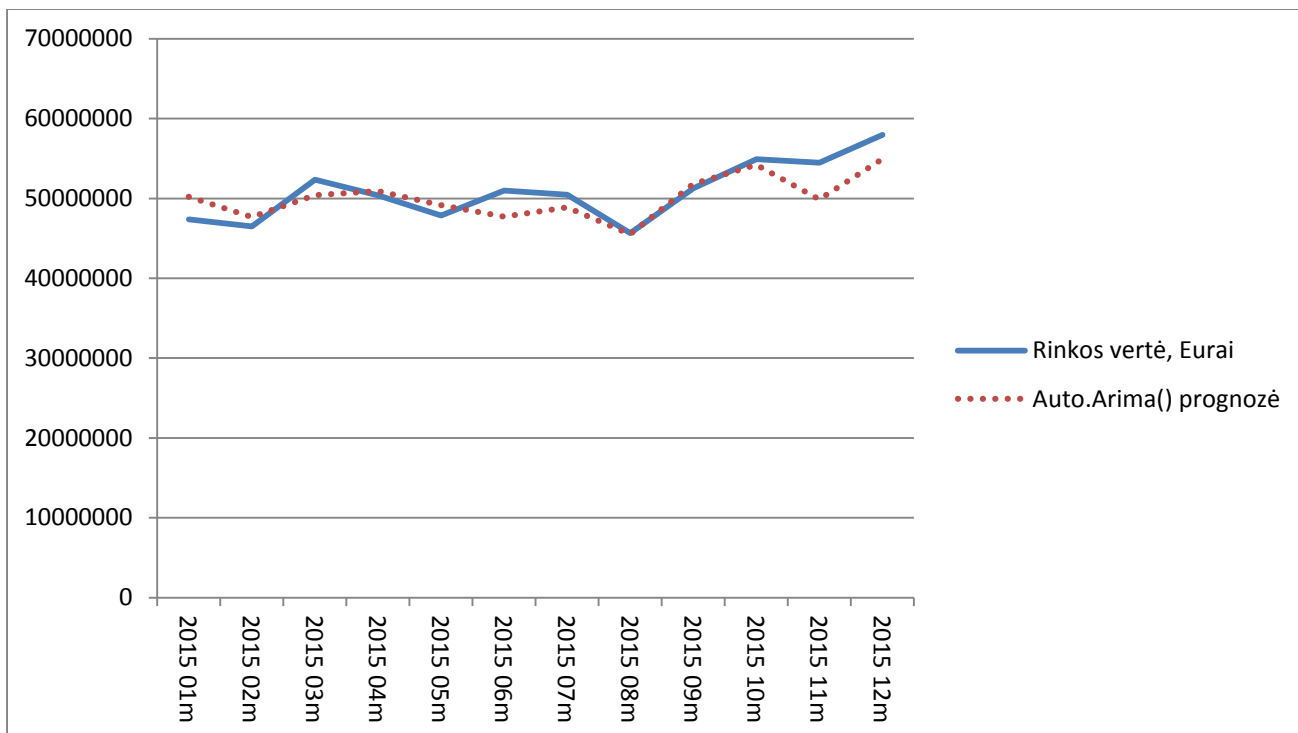
Šiame poskyryje pateikti laiko eilučių skirtingų modelių rezultatų grafiniai atvaizdavimai ir paklaidų įverčiai. Daugiau apie skirtingus modelius galima rasti poskyryje 5.2.2.

Iliustracijose Pav. 53 ir Pav. 54 pateikti *auto.arima()* modelio rezultatai.

Iliustracijose Pav. 55 ir Pav. 56 pateikti *ETS* modelio rezultatai.

Iliustracijose Pav. 57 ir Pav. 58 pateikti *BATS* modelio rezultatai.

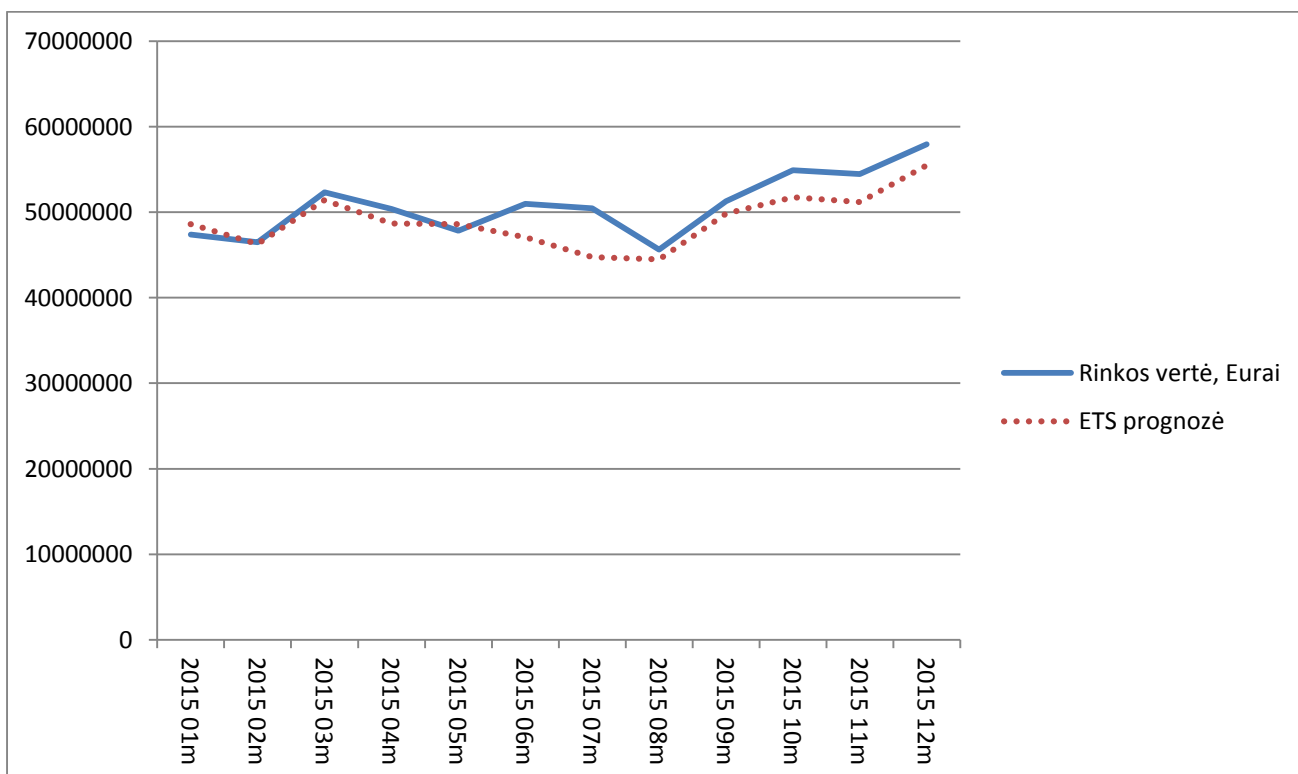
Iliustracijose Pav. 59 ir Pav. 60 pateikti *HoltWinters* modelio rezultatai.



Pav. 53 Laiko eilučių prognozės ir realių pardavimų palyginimas. *Auto.Arima()* modelis

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
742336	2224060	1800381	1.2688	3.48013	0.2172	0.5745647

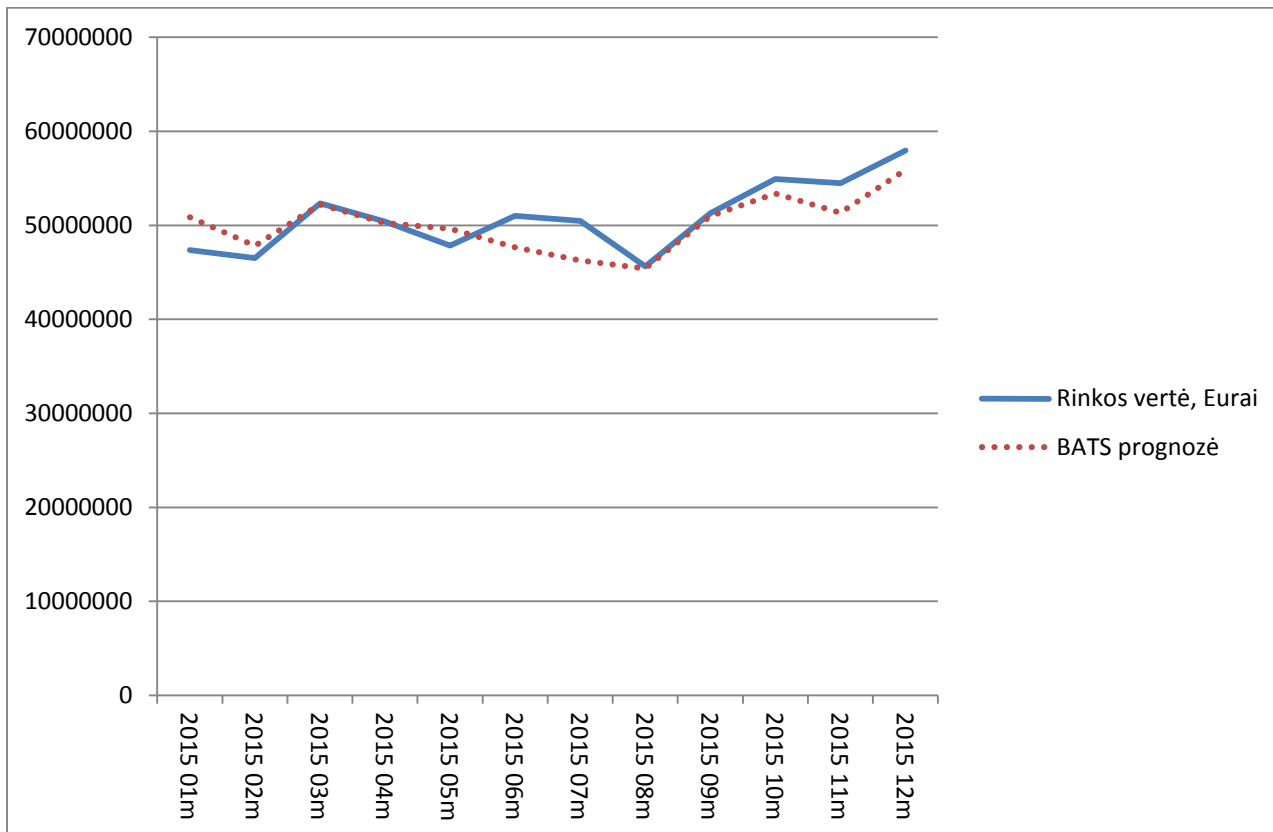
Pav. 54 Laiko eilučių paklaidų įverčiai. *Auto.Arima()* modelis



Pav. 55 Laiko eilučių prognozės ir realių pardavimų palyginimas. *ETS* modelis

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
1837792	2649536	2166764	3.4947	4.1865	0.21391	0.7418951

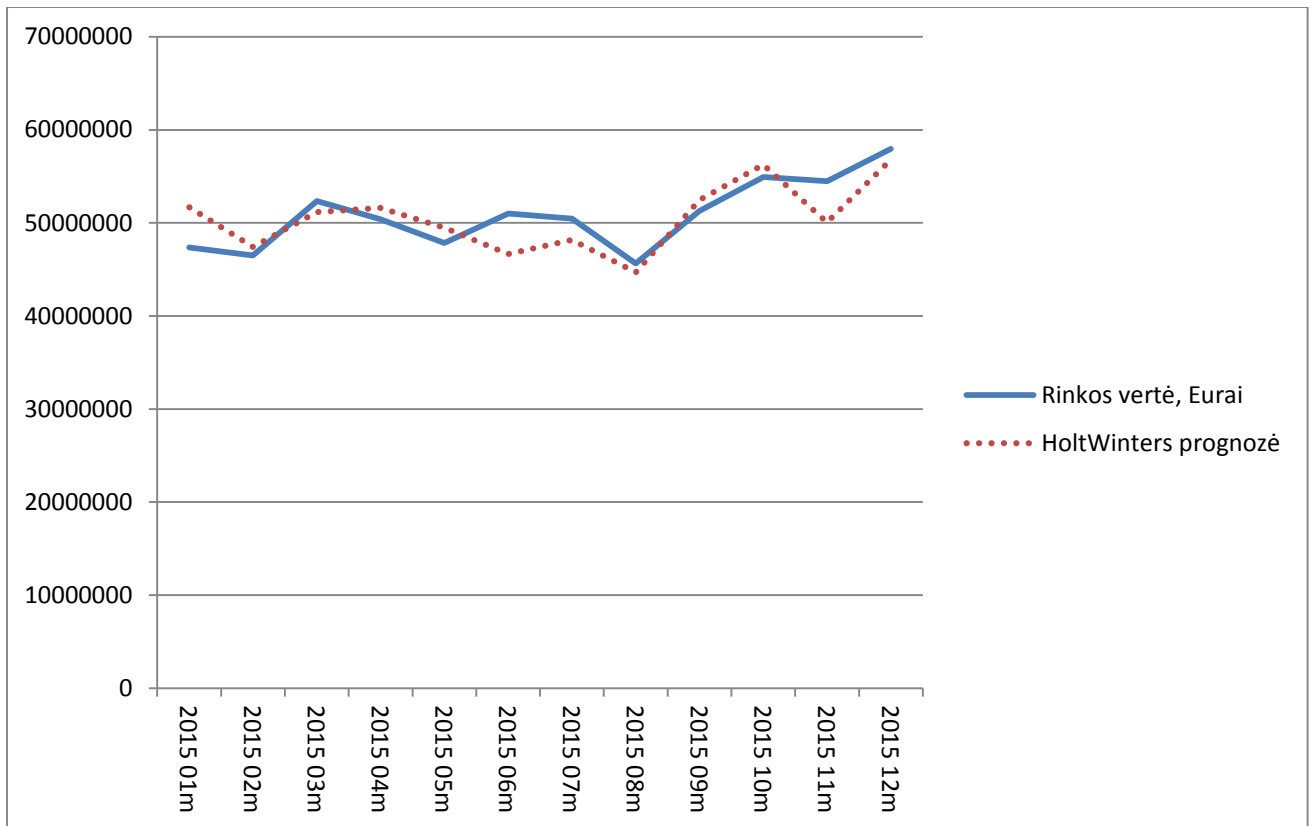
Pav. 56 Laiko eilučių paklaidų įverčiai. ETS modelis



Pav. 57 Laiko eilučių prognozės ir realių pardavimų palyginimas. BATS modelis

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
709944	2282796	1806640	1.22923	3.54653	0.30939	0.5778023

Pav. 58 Laiko eilučių paklaidų įverčiai. BATS modelis



Pav. 59 Laiko eilučių prognozės ir realių pardavimų palyginimas. *HoltWinters* modelis

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
314105	2479111	2070637	0.49308	4.07964	0.07787	0.6077324

Pav. 60 Laiko eilučių paklaidų įverčiai. *HoltWinters* modelis

5.3.3. Neuroninių tinklų eksperimento rezultatai

Eksperimentuojant su dirbtinių neuroninių tinklų modeliu, pirmiausia buvo bandoma gauti prognozę naudojant visus turimus parametrus: šalies populiaciją, BVP metinį augimą, mėnesio vidutinę temperatūrą ir PSDF metinį biudžetą. Gautas rezultatas parodytas Pav. 61, šiuo atveju naudoti trys vidiniai neuronai, tačiau rezultatas išliko nepakitęs bandant visais atvejais nuo vieno iki dešimties neuronų vidiniame sluoksnyje – prognozė visais atvejais išliko nekintanti laike ir buvo smarkiai žemiau realios rinkos vertės.

Tolesniuose žingsniuose bandyta sumažinti dirbtinio neuroninio tinklo kompleksumą. Išmesti parametrai su BVP metiniu augimu ir PSDF metiniu biudžetu, tai daryta atsižvelgiant į poskyryje 5.3.1 gautas parametrų reikšmingumo reikšmes tiesinės regresijos modelyje. Nors buvo gauti ryški koreliacija tarp vaistų pardavimų vertės ir šalies populiacijos, tačiau populiacijos parametras taip pat buvo pašalintas. To priežastis buvo klaidingasis ryšys (angl. *spurious relationship*), kaip yra žinoma, Lietuvos populiacija turi tendenciją mažėti, tuo pačiu metu parduodamų vaistų vertė bėgant metams panašiu procentu didėja. Panaudojus šiuos duomenis matematinės analizės įrankyje, matysime, jog yra aptinkama koreliacija, t. y. kuo mažesnė

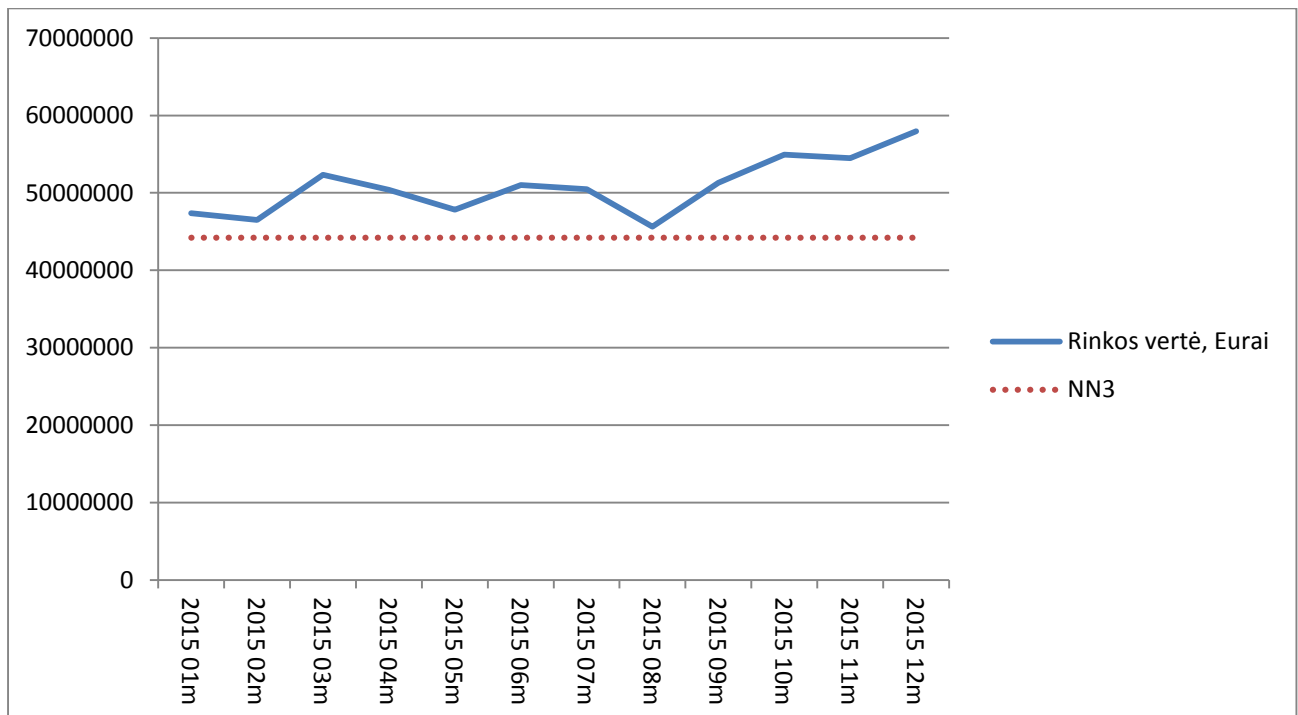
populiacija, tuo daugiau vaistų parduodama. Tačiau, visiems suprantama, jog toks ryšys yra klaidingas ir nelogiškas, šį ryšį lemia kiti nenumatyti vidiniai veiksniai, tokie kaip Lietuvos ir viso pasaulio infliacija ar išaugusios medžiagų kainos, dėl ko didėja pačių vaistų kainos.

Pav. 62 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (2) pateikta prognozė, kai naudojamas tik vidutinės mėnesio temperatūros parametras ir vienas vidinio tinklo neuronas. Šalia iliustracijoje Pav. 63 parodytos gautos paklaidos.

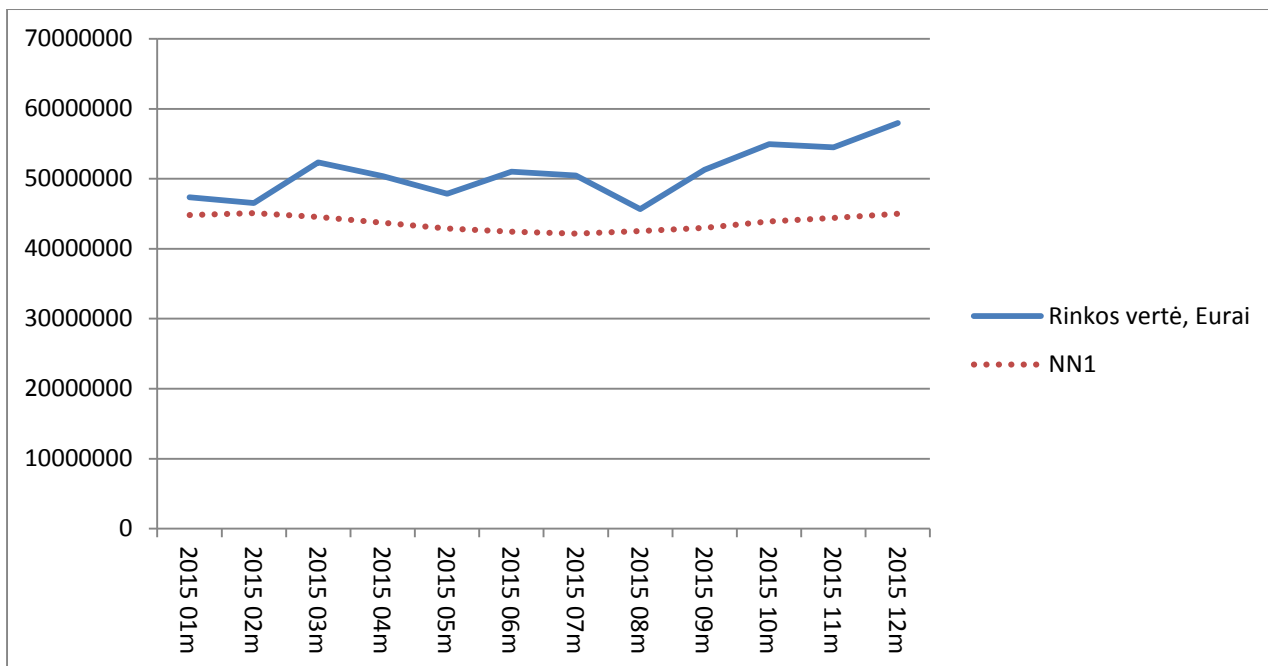
Pav. 64 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (3) pateikta prognozė, kai naudojamas vienas vidutinės mėnesio temperatūros parametras ir du vidinio tinklo neuronai. Šalia iliustracijoje Pav. 65 parodytos gautos paklaidos.

Pav. 66 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (4) pateikta prognozė, kai naudojamas vienas vidutinės mėnesio temperatūros parametras ir trys vidinio tinklo neuronai. Šalia iliustracijoje Pav. 67 parodytos gautos paklaidos.

Iš paklaidų lentelių matosi, jog geriausią rezultatą duoda neuroninis tinklas su dviem vidiniais neutronais, tačiau net ir ši prognozė yra pakankamai netiksli, tai matosi vien tik pažiūrėjus į prognozės grafinį atvaizdavimą. Pati prognozės kryptis labai primena tiesinės regresijos prognozė, kas nėra nieko keisto, nes pagrindinis lemiamas parametras abejais atvejais yra mėnesio temperatūra. Tiesa, pažiūrėjus į specifines paklaidas, matomas ryškus tiesinės regresijos pranašumas prieš dirbtinių neuroninių tinklų modelius.



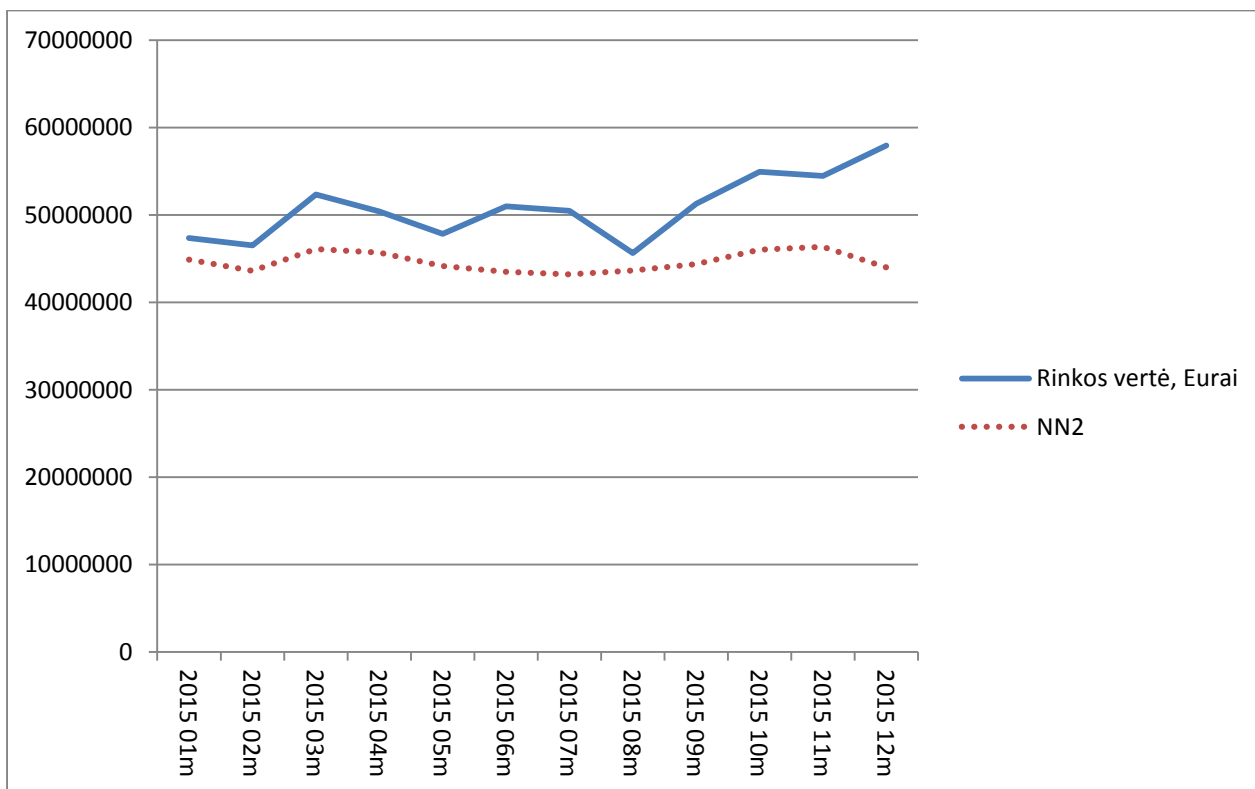
Pav. 61 Neuroninio tinklo prognozės ir realių padavimų palyginimas (1)



Pav. 62 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (2)

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
7142966	7907054	7142966	13.6721	13.6721	0.32953	2.222593

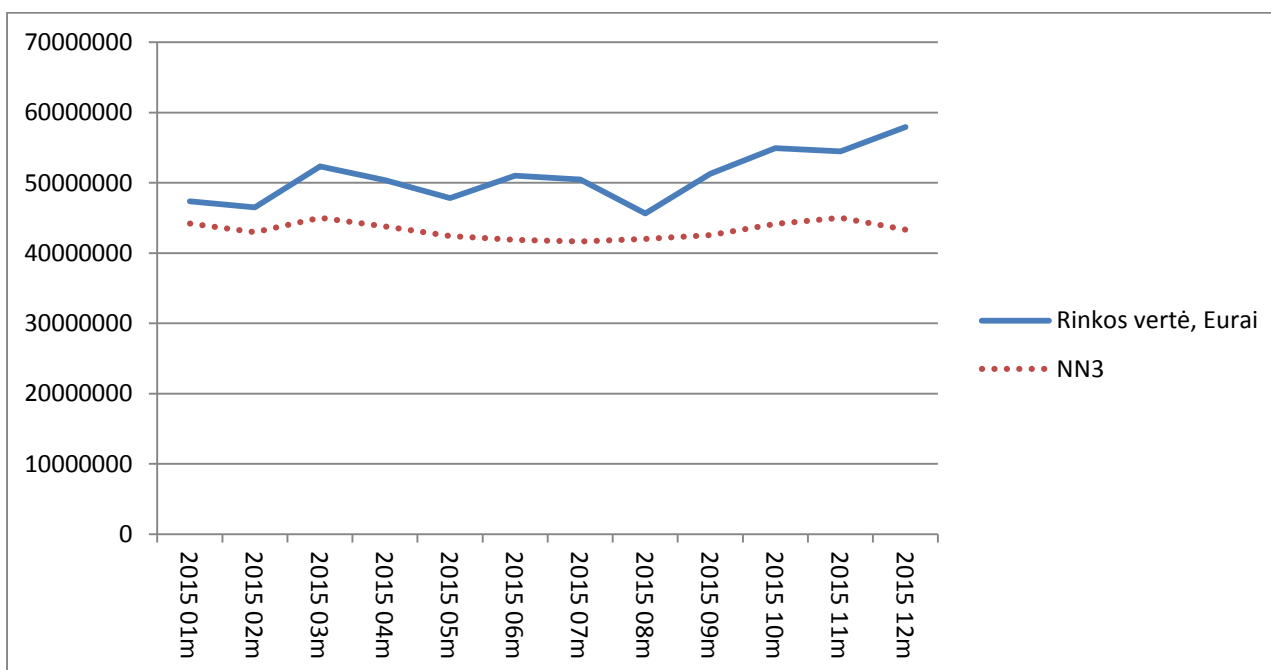
Pav. 63 Neuroninio tinklo paklaidų įverčiai (2)



Pav. 64 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (3)

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
6223364	7014028	6223364	11.884	11.884	0.22689	1.956913

Pav. 65 Neuroninio tinklo paklaidų įverčiai (3)



Pav. 66 Neuroninio tinklo prognozės ir realių padavimų palyginimas. (4)

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
7593729	8252280	7593729	14.5921	14.5921	0.26497	2.314673

Pav. 67 Neuroninio tinklo paklaidų įverčiai (4)

5.3.4. Bendri eksperimento rezultatai

Žemiau pateikiama visų eksperimente naudojamų modelių paklaidų suvestinė Pav. 68 ir pažymėti modeliai su geriausiais įvertinimais atitinkamose kategorijose. Taip pat pateikti paklaidų aprašai ir jų interpretavimo ypatumai.

ME. (angl. *Mean error*). Paklaidų vidurkis. Lengviausiai apskaičiuojamas įvertis, netinkamas lyginti modelius su skirtingu kiekiu paklaidų, taip pat jautrus paklaidų kryptčiai (teigiamai ar neigiamai). Dėl šių priežasčių praktikoje retai naudojamas kaip galutinis modelio įvertinimas.

RMSE. (angl. *root mean squared error*). Vidutinė kvadratinė paklaida. Dažnai naudojamas prognozavimo įvertis, išryškina didesnes paklaidas, privalumą teikia modeliams, kurių paklaidos artimiau pasiskirsčiusios aplink paklaidų vidurkį.

MAE. (angl. *mean absolute error*). Paklaidų absoliutus vidurkis. Skaičiavimo principas labai panašus į anksčiau aprašytą paklaidų vidurkį, išskyrus tai, kad paklaidų absoliutus vidurkis

nejautres paklaidos kryptčiai. Praktikoje naudojamas tik kaip apibendrintas pradinis įvertis, dėl tų pačių trūkumų kaip ir paklaidos vidurkis (angl. *Mean error*).

MPE. (angl. *mean percentage error*). Vidutinė procentinė paklaida. Veikimo principas, privalumai ir trūkumai panašūs į paklaidų vidurkį (ME), tik paklaida išreiškiama procentine dalimi.

MAPE. (angl. *mean absolute percentage error*). Vidutinė absoliuti procentinė paklaida.

ACF1. Pirmos eilės autokoreliacija. Parodo ar skirtinguose laiko momentuose esančios paklaidos neturi dėsningumo tarpusavyje. Jeigu paklaidos dėsningos, tai reiškia, jog nebuvo teisingai įvertinti parametrai arba pats modelis. Jeigu autokoreliacijos nėra, tai rodo, jog paklaidas galima priskirti baltajam triukšmui (angl. *white noise*) ir turimą modelį laikyti korektišku ir pilnu.

Theil's U. Palygina modelio vidutinę kvadratinę paklaidą su ta pačia paklaida naudojant naivų modelį. Theilo įvertį galima išskirti į tris reikšmių grupes:

- Mažiau už 1. Prognozavimo modelis yra reikšmingas ir tikslesnis už atsitiktinį klaidžiojimą.
- Lygu 1. Prognozavimo modelio tikslumas yra toks pats kaip ir atsitiktinio klaidžiojimo.
- Daugiau už 1. Prognozavimo modelis yra bevertis ir mažiau tikslus už atsitiktinį klaidžiojimą.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
auto.arima	742336	2224060	1800381	1.2688	3.48013	0.2172	0.5745647
ETS	1837792	2649536	2166764	3.4947	4.1865	0.21391	0.7418951
BATS	709944	2282796	1806640	1.22923	3.54653	0.30939	0.5778023
HoltWinters	314105	2479111	2070637	0.49308	4.07964	0.07787	0.6077324
Tiesinė regresija	946971	3561501	3125583	1.42593	6.09502	0.33986	0.9534244
Neuroninis tinklas	6223364	7014028	6223364	11.88404	11.88404	0.2268942	1.956913

Pav. 68 Paklaidų įverčių suvestinė

Iš Pav. 68 suvestinės matome, jog tiksliausias prognozes pateikia *auto.arima* ir *HoltWinters* modeliai. *Auto.arima* modelis mažiausias paklaidas turi keturiuose kategorijose, *HoltWinters* – trijose. Ne per daug nuo jų atsilieka ir kitu du laiko eilučių modeliai: *ETS* ir *BATS*. Didžiausias paklaidas pateikia neuroninių tinklų modelis, jo Theilo įvertis siekia net 1.9, kas parodo, jog modelis neturi jokios praktinės naudos. Taip pat prastokus rezultatus parodė tiesinės regresijos modelis, jo Theilo įvertis nors ir buvo mažesnis už vienetą, tačiau prognozė buvo labai menkai tikslesnė už atsitiktinį klaidžiojimą, kiti paklaidų įverčiai gan ryškiai atsiliko nuo laiko eilučių modelių.

5.4. Sprendimo taikymo rekomendacijos

Eksperimento metu buvo nustatyta, jog duomenų saugyklos projektavimas paprastą vieno lygio žvaigždės schema pilnai atitinka keliamus reikalavimus. Duomenų saugykla, kurta pagal šią schemą, ir kartu ją aprašanti metaduomenų lentelė, leidžia programiškai kurti dinamines SQL užklausas ir parodyti vartotojui reikiamą ataskaitą jos pareikalavimo metu. Reikia paminėti, jog tokio ataskaitų generavimo būdo trūkumas yra duomenų ištraukimo iš saugyklos greitis. SQL užklausos kiekvieną kartą paleidžiamos su pilnais ir neagreguotais duomenimis, dėl ko greitis tampa neoptimaliu ir labai priklausomu nuo ataskaitos sudėtingumo ir apsibrėžtų laiko periodų. Galima

daryti išvadą, kad darbo metu aprašytas saugyklos ir metaduomenų metodas geriausiai tinka dirbant su nepastoviais, dažnai atnaujinamais ir dažnai besikeičiančiais duomenimis. Dirbant tokioje aplinkoje užklausų generavimas su pilnais duomenimis yra neišvengiamas ir veikimo greitis bet koku atveju nebus optimalus.

Sekantis žingsnis yra aptarti prognozavimo rekomendacijas. Eksperimentas parodė, jog pardavimams prognozuoti geriausiai tinka laiko eilučių modeliai. Tiesinės regresijos ir neuroninių tinklų modeliai pateikė dideles paklaidas ir jų naudojimas yra nerekomenduojamas. Kaip teigiama [29], neuroninių tinklų panaudojimas pardavimams ar kitiems, smulkiai laike išsidėsčiusiems, duomenims prognozuoti yra abejotinas. Šiuo klausimu yra vykdyta daug tyrimų ir jų pateikiamos išvados labai įvairios, galima daryti išvadą, jog neuroninių tinklų panaudojimo nauda labai priklauso nuo specifinės tyrimo srities. Tiesa, verta paminėti, jog šiame darbe pastariesiems modeliams buvo pateikiamas minimalus parametrų kiekis, tikėtina, kad, prognozuojant specifinės srities pardavimus su didesniu kiekiu tą sritį aprašančių parametrų, rezultatai būtų geresni. Deja, nėra labai tikėtina, jog paprastas sistemos vartotojas turės šiuos siauros srities parametrus ir juos panaudos, tai būtų aktualiau prognozavimo specialistui. Atsižvelgiant į tai, kad sistemos vartotojas greičiausiai neturės gilių prognozavimo žinių ir prognozavimo procesas turėtų būti kuo paprastesnis, galima daryti išvadą, jog tokioje aplinkoje rekomenduotina įdiegti laiko eilučių prognozavimo metodus. Kaip parodė paklaidų palyginimas, šį procesą taip pat galima lengvai automatizuoti panaudojant *auto.arima* laiko eilučių modelį, kuris pateikė mažiausiais paklaidas tarp visų tiriamų modelių. Tiesa, kaip teigiama [28], automatizuotą prognozavimo procesą reikia naudoti atsargiai, toks procesas negali matyti ir atsižvelgti į unikalios veiklos srities duomenų savybes, kurias žino srities specialistas.

6. REZULTATŲ APIBENDRINIMAS IR IŠVADOS

1. Darbe buvo išanalizuoti ir palyginti prognozavimo modeliai, mokomi su priežiūra ir be priežiūros, bei nuspręsta didesnę dėmesį skirti pirmajam tipui, dėl jo geresnio tinkamumo probleminei sričiai (galimybės gauti klaidos ar teisingo veikimo signalą, kuris įvertintų potencialų sprendimą).
2. Atlikus prognozavimo atskirų modelių analizę, nuspręsta įgyvendinti laiko eilučių, tiesinės regresijos ir neuroninių tinklų prognozavimo modelius, dėl jų automatizavimo galimybių ir tinkamumo taikymui pardavimų srityje. Tuo tarpu logistinės regresijos modelis atmestas dėl to, kad spėjamas kintamasis gali įgyti tik dvi reikšmes, kas nėra tinkama prognozuojant pardavimus.
3. Darbe išanalizuotos dabartinių prognozavimo įrankių kategorijos, nustatyti jų pagrindiniai trūkumai – veikimo sudėtingumas, specifinių žinių reikalingumas. Darbe sukurtas sprendimas, padengiantis aprašytus konkurentų trūkumus.
4. Siekiant automatizuoti SQL užklausų formavimą ir sukurti prognozavimo modulį, lengvai pritaikomą konkrečiai dalykinei sričiai ar duomenų saugyklai, sudarytas žvaigždės schemas struktūra pagrįstas saugyklos metamodelis.
5. Eksperimento metu, prognozuojant realius vaistų pardavimus ir įvertinus naudotų modelių paklaidas, buvo nustatyta, jog tiksliausi prognozavimo modeliai vaistų pardavimų srityje – laiko eilučių modeliai.
6. Analizuojant eksperimento rezultatus pastebėta, kad tarp laiko eilučių modelių, su mažiausiomis paklaidomis prognozavo *auto.arima* modelis – tai leidžia sistemos vartotojams pasiūlyti automatizuotą, specifinei duomenų aibei skirtą, prognozavimo modelio parinkimą.
7. Eksperimento rezultatai rodo, kad dėl labai plačios prognozuojamos srities ir mažo kiekio aiškinamųjų parametrų, tiesinės regresijos ir dirbtinių neuroninių tinklų modelių paklaidos buvo gerokai didesnės už laiko eilučių modelių paklaidas.
8. Rekomenduojamas kitas žingsnis ateities tyrimams – pakartoti eksperimentą, parenkant siauresnę sritį su didesniu kiekiu parametrų, ir rezultatų tikslumą vėl palyginti su laiko eilučių modeliais.

LITERATŪRA

- [1] C. B. J.T Mentzer, *Sales Forecasting Management*, CA: Thousand Oaks, 1998.
- [2] C.-W. Chu ir G. P. Zhang, A comparative study of linear and nonlinear models for aggregate retail sales forecasting, *Science Direct*, 2003.
- [3] M. Kantardzic, *Data Mining– Concepts, Models, Methods, and Algorithms*, New Jersey: John Wiley & Sons, 2011.
- [4] Z. R. Yang, *Machine Learning Approaches to Bioinformatics*, World Scientific Publishing Co, 2010.
- [5] G. Shmueli, N. R. Patel ir P. C. Bruce, „Supervised and Unsupervised learning,“ įtraukta *Data Mining for Business Intelligence*, Hoboken, John Wiley and Sons, 2011, pp. 15-16.
- [6] B. Clarke, E. Fokoue ir H. H. Zhang, „Unsupervised learning: Clustering,“ įtraukta *Principles and Theory for Data Mining and Machine Learning*, New York, Springer Science & Business Media, 2009, pp. 405-408.
- [7] J. Armstrong, „Selecting Forecasting Methods,“ įtraukta *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer, 2001.
- [8] C. PRABHU, „Characteristics Of A Data Warehouse,“ įtraukta *DATA WAREHOUSING– Concepts, Techniques, Products and Applications*, New Delhi, PHI Learning Private Limited, 2008, p. 9.
- [9] L. Silverston, „Adding Derived Data,“ įtraukta *The Data Model Resource Book: A Library of Universal Data Models for All Enterprises*, Indianapolis, John Wiley & Sons, 2011.
- [10] N. Bhansali, „Metadata,“ įtraukta *Strategic Data Warehousing– Achieving Alignment with Business*, Boca Raton, CRC Press, 2009, pp. 50-52.
- [11] G. G. K., „Data Warehouse Metadata,“ įtraukta *Introduction to Data Mining with Case Studies*, New Delhi, PHI Learning Pvt, 2012, pp. 375-376.
- [12] F. Silvers, „Data Mart,“ įtraukta *Building and Maintaining a Data Warehouse*, Boca Raton, CRC Press, 2008, pp. 128-130.
- [13] L. J. ir C. S., „Best practices in data warehousing to support business initiatives and needs,“ įtraukta *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, Waikoloa, 2004.
- [14] D. W. Hosmer, J. S. Lemeshow ir R. X. Sturdivant, *Applied Logistic Regression*, Hoboken: John Wiley & Sons, 2013.
- [15] P. Braspenning, F. Thuijsman ir A. Weijters, „Criteria in Choosing for a Neural-Network Solution,“ įtraukta *Artificial Neural Networks: An Introduction to ANN Theory and Practice*, Berlin, Springer Science & Business Media, 1995, pp. 235-236.
- [16] S. N. Sivanandam ir S. N. Deepa, „Advantages and Disadvantages of Neural Network,“ įtraukta *Introduction to Neural Networks Using Matlab 6.0*, New Delhi, Tata McGraw-Hill Education, 2006, pp. 368-369.
- [17] D. C. Montgomery, E. A. Peck ir G. G. Vining, *Introduction to Linear Regression Analysis*, Hoboken: John Wiley & Sons, 2012.
- [18] J. I. Gallin ir F. P. Ognibene, „Simple and multiple linear regression,“ įtraukta *Principles and Practice of Clinical Research*, Burlington, Academic Press, 2011, p. 185.
- [19] R. J. Rossi, „The simple linear regression model,“ įtraukta *Applied Biostatistics for the Health Sciences*, Hoboken, John Wiley & Sons, 2009, p. 342.
- [20] I. Balabonienė, R. Bliekienė ir A. Stundžienė, „Laiko eilutės ekonometrinė analizė,“ įtraukta *Ekonometrija. Praktinis regresijos ir laiko eilučių modelių taikymas*, Kaunas, leidykla

- "Technologija", 2013, pp. 177-217.
- [21] C. Chatfield, „Types of Variation,“ ĩtraukta *The Analysis of Time Series*, Boca Raton, Florida, Taylor & Francis e-Library, 2013, p. 11.
- [22] B. J. Goldenberg, „A CRM Primer,“ ĩtraukta *CRM in Real Time: Empowering Customer Relationships*, New Jersey, Information Today, 2008, p. 3.
- [23] R. Nau, „How to move data around,“ [Tinkle]. Available: <http://people.duke.edu/~rnau/411move.htm>. [Kreiptasi 16 4 2015].
- [24] P. S. Foote ir M. Krishnamurthi, „Forecasting Using Data Warehousing Model: Wal-Mart's Experience,“ *The Journal of Business Forecasting Methods & Systems*, 2001.
- [25] S. M. Carlos Coronel, „Denormalization,“ ĩtraukta *Database Systems: Design, Implementation, & Management*, Stamford, Cengage Learning, 2014, pp. 217-220.
- [26] A. Schneider, G. Hommel ir M. Blettner, „Linear Regression Analysis: Part 14 of a Series on Evaluation of Scientific Publications,“ *Dtsch Arztebl Int*, 2010.
- [27] M. J. Miket, „Multiple Regression - Selecting the Best Equation,“ [Tinkle]. Available: <http://math.usask.ca/~miket/S344D..pdf>. [Kreiptasi 16 4 2015].
- [28] R. Nau. [Tinkle]. Available: <http://people.duke.edu/~rnau/rsquared.htm>. [Kreiptasi 5 Lapkritis 2015].
- [29] J. Faraway ir C. Chatfield, „Time series forecasting with neural networks: a comparative study using the airline data,“ *Appl. Statist.*, 1998.