

**VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS**

INFORMATIKOS KATEDRA

Verslo informacijos sistemų studijų programa
Kodas 62103S138

SAULĖ KLIKŪNAITĖ

MAGISTRO BAIGIAMASIS DARBAS

**DUOMENŲ GAVYBOS METODAI RYŠIŲ SU VERSLO KLIENTAIS
SISTEMOSE**

Kaunas 2009

**VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS**

INFORMATIKOS KATEDRA

SAULĖ KLIKŪNAITĖ

MAGISTRO BAIGIAMASIS DARBAS

**DUOMENŲ GAVYBOS METODAI RYŠIŲ SU VERSLO KLIENTAIS
SISTEMOSE**

Leidžiama ginti _____

Magistrantas _____
(parašas)

Darbo vadovas _____
(parašas)

(darbo vadovo mokslo laipsnis, mokslo
pedagoginis vardas, vardas ir pavardė)

Darbo teikimo data _____
Registracijos Nr. _____

Kaunas 2009

TURINYS

| | |
|--|----|
| SANTRUMPŲ SĄRAŠAS | 4 |
| LENTELIŲ SĄRAŠAS | 4 |
| PAVEIKLŲ SĄRAŠAS | 5 |
| SANTRAUKA | 6 |
| ĮVADAS | 7 |
| 1. DUOMENŲ GAVYBOS METODŲ IR JŲ TAIKYMO VERSLO SRITYSE | |
| APŽVALGA..... | 10 |
| 1.1. Gavybos metodų klasifikacija ir apibendrinimas | 11 |
| 1.1.1. Duomenų valdymo problemos | 11 |
| 1.1.2. Duomenų gavybos metodai ir jų klasifikacija..... | 13 |
| 1.2. Pagrindinės duomenų gavybos funkcijos verslo sistemose..... | 16 |
| 1.3. Duomenų analizės specifika ryšių su klientais valdymo sistemose | 19 |
| 1.4. CRM apžvalga ir klasifikacija | 21 |
| 1.5. Duomenų gavybos metodų taikymo CRM vertinimas..... | 24 |
| 2. DUOMENŲ GAVYBOS SPECIFIKA B2B APLINKOJE | 27 |
| 2.1. Pardavimų kanalų optimizavimas..... | 27 |
| 2.2. Klientų aptarnavimo tobulinimas | 28 |
| 2.3. Pirkimo (tiekimo) kanalų tobulinimas..... | 28 |
| 2.4. Vartotojų elgsenos prognozavimas..... | 31 |
| 2.5. Duomenų gavybos metodų taikymo CRM apžvalgos išvados | 32 |
| 3. DUOMENŲ GAVYBOS TAIKYMO CRM B2B METODO SIŪLYMAS..... | 33 |
| 4. DUOMENŲ GAVYBOS METODŲ TAIKYMO B2B APLINKOJE | |
| EKSPERIMENTINIS TYRIMAS | 38 |
| 4.1. Įvesties duomenų aprašymas | 39 |
| 4.2. Įvesties duomenų apdorojimas | 41 |
| 4.3. Pirkimų kreditingumo, nustatomas klientams, tyrimo eiga..... | 43 |
| 4.4. Prekių kiekių pirkimams tyrimo eiga | 49 |
| 4.5. Tvarkaraščių, gabenimo laiko valdymo tyrimo eiga | 65 |
| IŠVADOS..... | 72 |
| PASIŪLYMAI | 74 |
| LITERATŪRA | 75 |
| PRIEDAI | 78 |

SANTRUMPŲ SĄRAŠAS

IM- Informacijos vadyba (angl. Information Management)
IRM - Informacijos išteklių vadyba (angl. Information Resources Management)
ISEC -Informacijos apsauga (angl. Information Security)
ITM - Informacijos technologijų vadyba (angl. Information Technologies Management)
KM - Žinių vadyba (angl. Knowledge Management)
PR - Ryšiai su spauda (angl. Press Relations)
CRM - ryšių su klientais valdymas (angl. Client Relationship Management)
IC - Intelektualus kapitalas (angl. Intelektual Capital)

LENTELIŲ SĄRAŠAS

| | |
|--|----|
| 1 lentelė. Duomenų surinkimo problematikos apibendrinimas..... | 12 |
| 2 lentelė. Duomenų gavybos metodų apibendrinimas | 15 |
| 3 lentelė. Įvesties laukų aprašymas | 41 |
| 4 lentelė. STATISTICA 7 aprašymas | 42 |
| 5 lentelė. Įvesties laukų aprašymas pirkimo kreditavimo tyrimui | 44 |
| 6 lentelė. Įvesties laukai pirkimo kreditavimo tyrimui | 45 |
| 7 lentelė. Sviurių rangavimas | 46 |
| 8 lentelė. Įvesties laukų aprašymas prekių kiekių pirkimams tyrimui | 50 |
| 9 lentelė. Įvesties duomenų fragmentas | 51 |
| 10 lentelė. Metodų tyrimui pasirinkimo apibendrinimas | 52 |
| 11 lentelė. MARSpline parametrai analizei | 55 |
| 12 lentelė. Modelio koeficientai | 56 |
| 13 lentelė. Įtakų pasiskirstymas | 56 |
| 14 lentelė. MARSpline parametrai analizei | 58 |
| 15 lentelė. Įvesties duomenų fragmentas | 61 |
| 16 lentelė. Kiekio tyrimo koreliacinė matrica | 62 |
| 17 lentelė. Kiekio tyrimo statistinė informacija | 89 |
| 18 lentelė. Prognozės tinkamumo analizė statistiniai rezultatai | 64 |
| 19 lentelė. Prognozės tinkamumo analizė | 65 |
| 20 lentelė. Prognozės tinkamumo analizė | 65 |
| 21 lentelė. Įvesties laukų aprašymas gabenimo laiko valdymo tyrimui | 66 |
| 22 lentelė. Bendra statistinė informacija kiekvienai grupei | 69 |

| | |
|--|----|
| 23 lentelė. Literatūros šaltiniai rasti naudojantis paprastomis paieškos sistemomis | 79 |
| 24 lentelė. Informacijos paieška populiariose mokslinės informacijos duomenų bazėse | 81 |
| 25 lentelė. Programinės įrangos su DGM palyginimas | 86 |
| 26 lentelė. Galutinė klasifikacija | 88 |
| 27 lentelė. Analizės rezultatų langas | 88 |
| 28 lentelė. Atstumai tarp sudarytų grupių | 89 |
| 29 lentelė. Grupių parametrai | 89 |

PAVEIKLŲ SĄRAŠAS

| | |
|--|----|
| 1 pav. Duomenų gavybos metodų klasifikacija | 13 |
| 2 pav. CRM vieta organizacijos intelektualaus kapitalo kūrime | 21 |
| 3 pav. Duomenų gavybos metodo veikimo algoritmas klientų valdymo sistemoje | 25 |
| 4 pav. Duomenų gavybos metodas B2B aplinkoje | 35 |
| 5 pav. Statinis modelio įgyvendinimas | 36 |
| 6 pav. Dinaminis modelio įgyvendinimas | 37 |
| 7 pav. Teorinė tinklo svorių sudarymo iliustracija | 46 |
| 8 pav. Praktinė tinklo svorių sudarymo iliustracija | 47 |
| 9 pav. Duomenų apdorojimo rezultato langas | 47 |
| 10 pav. Kreditavimo kategorijų pasiskirstymo dažnis | 48 |
| 12 pav. Įtaką darančių kintamųjų pasiskirstymas | 57 |
| 13 pav. Kiekio ir prognozuojamo kiekio pasiskirstymas pagal klientų skaičių | 59 |
| 14 pav. Kiekio ir prognozuojamo kiekio pasiskirstymas pagal specifikacijų pilnumą | 60 |
| 15 pav. Vidurkio ir standartinio nuokrypio pasiskirstymas | 63 |
| 16 pav. Vidurkio ir standartinio nuokrypio pasiskirstymas | 64 |
| 17 pav. Duomenų apdorojimo rezultatų langas | 69 |
| 18 pav. Pavėluotų pristatyti dienų pasiskirstymas pagal grupes | 70 |
| 19 pav. Grupių pasiskirstymas pagal maksimalų pristatyti pavėluotų dienų skaičių | 71 |
| 20 pav. CRM procesai ir funkcijos | 84 |
| 21 pav. Programinių paketų rinka pagal valdymo lygmenius bei stipriausios kompanijos gamintojos | 85 |

SANTRAUKA

KLIKŪNAITĖ, Saulė. (2009) *Data mining models in business customers relationship management systems*. MBA Graduation Paper. Kaunas: Vilnius University, Kaunas Faculty of Humanities, Department of Informatics. 90 p.

S U M M A R Y

Latest activity in scientific field has been major in research of algorithms that are able to solve problems related to data analysis using tools of dataset clustering and data mining. Principle of data mining cover analysis of various business environment steps and suggests ways of prognosing, analyzing, classifying data.

Data mining tools are applicable to customer relationship management systems offering key data analysis, increasing customer knowledge.

Scope of thesis is the effective usability of data mining methods for customer relationship management system specifically used by business-to-business companies. Suggested concept leads to new perspective of transactional data analysis through following stages: client credibility valuation, product purchase amount prognosis and product delivery handling. Combined customer orientated data with supply chain inputs gave successful results in real data experiment.

In conclusion, data mining tools improve data analysis in customer relationship management in B2B companies' data analysis routine.

IVADAS

Pastaruosius metus duomenų gavyba tapo populiari mokslinėje tyrimų veikloje. Pirmoji duomenų gavybos metodų tyrimų karta analizavo klasikines duomenų gavybos problemas, aktualias stacionarioms duomenų sistemoms. Augant mokslinių sričių poreikiams, besiplečiant jų riboms bei integruojantis vienas į kitą, pradedami kelti didesni reikalavimai duomenų prognozei, klasifikavimui, faktoringei, regresinei analizei bei kitiems duomenų apdorojimo metodams. Moksliniu požiūriu tyrimai duomenų gavybos srityje galimai veda į naują informacijos sampratą ir duomenų analizę verslo aplinkoje.

Ryšių su klientais valdymo sistemos duoda kompanijoms galimybę koordinuoti informaciją apie klientus, padidinant galimybę planuoti verslo srities veiklą, atveriant perspektyvas tinkliniui verslui.

Verslo procesuose, o ypač tinklinio verslo, B2B aplinkos kompanijoms turi didelę reikšmę operuojamų duomenų analizė. Pjūviai, kuriais galima atlikti rinkos tyrimus ir nustatyti, kiek vartotojų perka tam tikrą prekę, neužtenka vadovautis tik pardavimų duomenimis. Tam, kad ryšių su klientais valdymo sistemos apimtų visapuses žinias apie klientus, reikalinga atlikti finansinių ir vadybinių duomenų analizę, apimančią ne tik marketingo sprendimus ar klientų poreikius, bet ir įtraukčią B2B specifikos - tiekimo grandinės ir elektroninės komercijos, investinius ir išvestinius duomenis.

Susijusios problemos buvo sprendžiamos mokslininko Y.Dezehen (2003). Eksperimentiniai tyrimai parodė, jog vieno tipo duomenys dažnai duoda klaidingus analizės rezultatus, neatlieka funkcijų perspektyvoms numatyti. S.Jaideep savo darbe pagrindė duomenų gavybos metodo taikymą finansinėse kompanijose, išskaidydamas klientų analizę į bendro, segmento ir individo lygius. Duomenų gavybos metodo naudojimo praktikoje klausimus kėlė Ramakrishnan, Agrawal ir Freytag (2005), siekdami pritaikyti juos kasdieniam duomenų apdorojimui. Tačiau moksliniai tyrimai nepalietė tinklinio verslo specifikos bei duomenų, tiesiogiai neteikiamų ryšių su klientais, tačiau galinčių įtakoti verslo sprendimus, pritaikymo.

Specializuotos programinės įrangos kaip Oracle Data Mining (ODM), Tiberius, ClearForest ar SAS, atlieka statistinę duomenų analizę, duomenų gavybos metodais prognozuoja ir klasifikuoja duomenis bei sprendžia atskiras problemas, tačiau neintegruotos į ryšių su klientais valdymo sistemas, nepateikia rezultatų orientuotų į verslo specifiką.

Šiame darbe sprendžiama duomenų analizės, atliktos duomenų gavybos metodais, pritaikymo verslo aplinkoje problematika. Siūlomo modelio rezultatais siekiama įrodyti

informacijos analizės pranašumą, kuri atlikta naudojant duomenis būdingus B2B verslo lygmenyje.

Darbo tikslas yra pasiūlyti duomenų gavybos metodų taikymo modelį ryšių su klientais valdymo sistemose B2B aplinkoje.

Darbo uždaviniai:

1. Ištirti duomenų gavybos metodus, siekiant išsiaiškinti metodų pritaikymą verslo srityje;
2. Atsižvelgiant į ryšių su klientais valdymo specifiką, identifikuoti duomenų analizės bei teikiamų sprendimų problemines sritis;
3. Remiantis B2B verslo specifiką, pasiūlyti modelio konceptą, sprendžiantį klientų klasifikavimo, prekių kiekio analizės ir pristatymo laiko planavimo problemas orientuotas į ryšius su klientais.
4. Atlikti eksperimentinį tyrimą, kurio metu, bus siekiama įvertinti ar siūlomas modelis yra vertas praktinės realizacijos.

Darbas susideda iš keturių dalių. Pirmoje aptariama duomenų gavybos metodų dalykinė sritis, apžvelgiama metodų klasifikacija bei taikymo ypatumai verslo srityse. Šioje darbo dalyje plėtojama ryšių su klientais sistemos tema, pateikiama duomenų gavybos metodų taikymo specifiką. Antra darbo dalis skirta susieti duomenų gavybos metodus taikomus ryšių su klientais sistemose su verslas verslui aplinkos aktualia problematika, iš kurios seka tolimesnis darbo etapas – metodo siūlymas. Metodo koncepcija atsako į mokslinėje literatūroje neanalizuotus klausimus susijusius su B2B verslo klientų valdymo duomenų analize - tai klientų klasifikavimą pagal taikomą kreditų sistemą, pirkimų valdymą ir prognozę bei prekių pristatymo laiko analizę. Trečioje dalyje atliekamas empirinis siūlomo metodo patikrinimas bei pateikiamos išvados.

Darbui atlikti naudoti mokslinės metodologijos metodai:

- dedukcinis metodas – siekiant apžvelgti darbui aktualią teoriją, atliktus mokslinius tyrimus, išskiriant verslo sričiai būdingus aspektus, pereinant prie konkrečios problemos sprendimo bei metodikos siūlymo. Siūloma metodika tikrinama empiriškai, atliekant eksperimentinį tyrimą.
- visuotinio pažinimo metodas naudotas darbo tikslų nustatymui, uždavinių formavimui, informacijos apie duomenų gavybos metodus rinkimui;
- analizės metodai naudoti atskleisti esamą situaciją remiantis pirmine informacija;

- remiantis apibendrinimo metodu buvo grupuojama teorinė medžiaga, dėstoma pagal konkrečią jos reikšmę;
- Abstrakcijos metodu remiantis buvo daromos kiekvienos darbo dalies ir galutinės darbo išvados.

Mokslinis pagrindas, analitinė informacija, eksperimentiniai ir praktiniai duomenys darbui gauti iš internetinių duomenų gavybos metodų ir jų pritaikymo sričių atliktų mokslinių tyrimų. Ieškant informacijos susidurta su konstruktyvios mokslinės informacijos stoka. Dažnai moksliniuose tyrimuose pateiktas teorinis pagrindimas tik pritaikytas tiriamos veiklos dalykinei sričiai.

Naudoti raktiniai žodžiai: duomenų gavyba (*angl. Data mining*), ryšių su klientas valdymas (*angl. Customer relationship management CRM*), duomenų gavybos metodai, ryšių su klientais valdymo sistemos ir metodai.

Darbo apimtis – 73 puslapiai be priedų, su priedais – 90 puslapiai. Darbe yra 29 lentelių, 21 paveikslai bei 5 priedai.

1. DUOMENŲ GAVYBOS METODŲ IR JŲ TAIKYMO VERSLO SRITYSE APŽVALGA

Šiame skyriuje atskleidžiama duomenų gavybos metodų pritaikymo verslo srityse problematika. Darbe pateikiama duomenų gavybos metodų klasifikacija, jų taikymo aspektai bei įtaka verslo sprendimų priėmimui. Apžvelgta mokslinė literatūra bei atlikti eksperimentiniai tyrimai statistinio modeliavimo dalykinėje srityje. Pateikta duomenų gavybos metodų kaip verslo duomenų analizės priemonės taikymo galimybių apžvalga.

Duomenų gavyba atsirado ir vystėsi taikomosios statistikos, vaizdų atpažinimo, dirbtinio intelekto metodų, duomenų bazių teorijos ir t. t. bazėje. Duomenų gavybos metodai taikomi tose srityse, kur duomenų kiekiai ir jų sistemų sudėtingumas išauga iki žmogui neapdoromo lygio ir reikalingas automatinis duomenų valdymas. Duomenų gavybos metodai taip pat naudojami atsiradus duomenų paskirties ar išteklių, duomenų kilmės informacijos poreikiui. Keletą iš tokių sričių galima būtų paminėti: mediciną, ryšių su klientais valdymo sistemas, sprendimų priėmimo ir valdymo sistemas, bankininkystę, mokymo, meteorologines sistemas, kompiuterinių tinklų, apsaugos sistemas. Kita vertus, dauguma mokslinių disciplinų palaipsniui generuoja didelius duomenų kiekius, kuriems apdoroti reikalingi duomenų gavybos metodai (Sunil, Chien-Chih, Puttikan, 2005).

Apžvelgiant mokslinę medžiagą susijusią su tyrimais ir pasiekimais duomenų gavybos taikymo srityje, reiktų išskirti šias mokslines įžvalgas:

- Duomenų gavybos metodų pritaikymas medicinoje, „bio“ moksle buvo analizuotas mokslininkų Ramakrishnan, Raghu, Agrawal, Rakesh, Freytag, Johann-Christoph ir kt. Darbe „Data mining: The Next Generation (2005)“ analizuojamas duomenų gavybos metodų panaudojimas CRM ir LS (angl. Life science), apibrėžiamos duomenų charakteristikos, aptariami duomenų gavybos rezultatai, analizuojami specifiniai duomenų gavybos metodų klientų valdymo sistemose faktoriai.
- JAIDEEP, Srivastava darbe „Data Mining for Customer Relationship Management (CRM)“ duomenų gavybos metodą naudojamą tam tikros verslo srities ar proceso analizei. Apžvelgiama duomenų gavybos metodų panaudojimo sritis. Pateikiamas požiūris: duomenų gavybos metodų įrankiais atliekama GAP analizė. Išvardinami ir aprašomi duomenų gavybos pagrindiniai ir išvestiniai metodai.
- Autorių Sunil, Arte, Chien-Chih, Lin, Puttikan, Prapai „Mining Internet user survey data Improving Customer Relationship Management (CRM) (2005)“ straipsnyje pateikiamas internetinio duomenų gavybos metodo pritaikymas klientų valdymo sistemoje. Aprašomi

duomenų gavybos metodai bei jų įtaka klientų valdymo sistemų kūrimo procesui, pateikiami eksperimentiniai rezultatai. Plačiau analizuojami duomenų gavybos internetu.

- Mokslininkas Xiaoshan, Du darbe Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship (2006) pateikia fundamentalių duomenų gavybos metodų taikymą. Straipsnyje akcentuojamas duomenų paruošimas duomenų gavybos metodams.

Vertinant mokslinius tyrimus paaiškėjo, jog duomenų gavybos metodų taikymo sritis yra plati ir ne pilnai ištirta orientuojantis į verslo sritis.

Tam, kad būtų galima išvelgti duomenų gavybos metodų pritaikymą verslo specifikos problemoms spręsti, toliau darbe pateikiami ir apibendrinami fundamentalūs bei išvestiniai duomenų gavybos metodai, jų klasifikacija bei taikymo ypatumai.

1.1.GAVYBOS METODŲ KLASIFIKACIJA IR APIBENDRINIMAS

Duomenys naudojami visuose IT sprendimuose, yra valdymo ir prekybos operacijų pagrindas. Valdyti duomenis yra sunku dėl įvairių priežasčių:

- Duomenų kiekis auga eksponentiškai: senus duomenis reikia saugoti, nuolat atsiranda nauji, todėl sunku rasti kiekvienam sprendimui reikalingus duomenis
- Duomenys yra išbarstyti įmonėje, kaupiami skirtingais metodais ir prietaisais. Saugomi keliuose skirtinguose, nutolusiuose serveriuose ar apdorojimo sistemose.
- Naudojami skirtingi formatai ir kalbos (žmonių taip pat);

Didėjant išorinių duomenų naudojimui įmonėje, atsiranda būtinybė užtikrinti duomenų saugumą, kokybę ir vientisumą. Skirtingose valstybėse galioja skirtingi įstatymai reglamentuojantys duomenų saugojimą. Rinkoje egzistuoja daugybė duomenų valdymo sprendimų. Dėl to sunku pasirinkti geriausiai tinkamą orientuotą į tam tikros verslo šakos specifiką.

1.1.1. Duomenų valdymo problemos

Vidiniai duomenų šaltiniai yra dažnai saugojami įmonės duomenų bazėse (kartais skirtinguose skyriuose). Jų turinys – žmonės, produktai, paslaugos ir procesai. Asmeniniai duomenys (Konceptijos, mintys, nuomonės, subjektyvūs pardavimų įvertinimai, nuomonė kokių veiksmų imsis konkurentai) yra įmonės darbuotojų patirties dokumentacija dažnai kaupiama ir prižiūrima pačių darbuotojų. Paties vartotojo sukurtos taisyklės, formulės, modeliai saugomi įmonės žinių bazėje.

Išoriniai duomenų šaltiniai yra skirtingo pobūdžio - komercinės (perkamos) duomenų bazės, davikliai, palydovai, vyriausybės ataskaitos, jų formatai: CD, luistai, žiniatinklio serveriai, vaizdai, diagramos, filmai, garsai. Dauguma duomenų visiškai nesusiję su kiekviena atskirai paimta problema. Todėl būtinas išorinių duomenų monitoringas.

Duomenų surinkimas yra sudėtinga problema (skirtingi šaltiniai, formatai, būdai). Duomenys gali būti fiksuojami darbo aplinkoje (in the field) arba išgaunami iš asmenų. Gali būti naudojami įrankiai ir davikliai arba studijos, apklausos, stebėjimai, ekspertų bendradarbiavimas. Egzistuoja daugybė techninės ir programinės įrangos duomenims kaupti ir perduoti duomenis (OKSL, 2008).

Duomenų surinkimas iš keleto skirtingų išorinių šaltinių yra dar sudėtingesnis uždavinys. Duomenų srautų valdymas (data flow manager – DFM) surenka duomenis iš išorinių šaltinių ir padeda juos ten kur reikia, kada reikia ir naudojimui paruošta forma. Duomenų srautų valdymas susideda iš:

1. Sprendimų paramos sistemos (DSS);
2. Centrinio duomenų poreikių procesoriaus;
3. Duomenų vientisumo tikrinimo komponento.

Sudėtingas duomenų surinkimo procesas sąlygoja duomenų kokybės problemas. Duomenys dažnai būna netikslūs, nepilni, dviprasmiški. Problema ypač aktuali didelėse ir centralizuotose DB. Sprendimai priimti naudojant blogos kokybės duomenis neša nuostolius (OKSL, 2008). Lentelėje pateiktos duomenų valdymo - duomenų surinkimo problemos.

1 lentelė

Duomenų surinkimo problematikos apibendrinimas

| Problema | Priežastis | Galimas sprendimas |
|--|---|---|
| Netikslūs duomenys | Įvedimo klaida | Sukurti sisteminių duomenų patikrinimo būdą. Naudoti automatinį duomenų įvedimą (skeneriai, davikliai) |
| Nesavalaikiai duomenys | Duomenų fiksavimo metodas nėra pakankamai greitas | Keisti duomenų generavimo sistemą. Pereiti prie kliento/serverio technologijos. Automatizuoti. |
| Duomenys fiksuojami ir identifikuojami netinkamai | Duomenys renkami pagal logiką, kuri netinkama analizės tikslams | Sukurti sistemą, kuri apjungtų duomenis reikiamu būdu. Naudoti išmanius paieškos agentus. |
| Reikiamų duomenų nėra | Tokie duomenys tiesiog nefiksuojami | Sukurti IS, kuri kauptų ir saugotų duomenis. Pasinaudoti žiniatinkliu panašių duomenų paieškai. Pasinaudoti ekspertais. |

Šaltinis: sudaryta autoriaus pagal OKSL KTU studijos, (2008). Duomenų valdymas, duomenų sandėliai.

Atliekant duomenų analizę minėtos duomenų rinkimo ir apdorojimo problemos aktualus aspektas atliekant duomenų analizę. Duomenų analizė susideda iš duomenų rinkimo, apdorojimo, duomenų analizės metodo pasirinkimo ir įgyvendinimo, bei rezultatų įvertinimo bei pritaikymo. Tad tai, kokie duomenys bus įvesti sąlygoja visus duomenų analizės etapus.

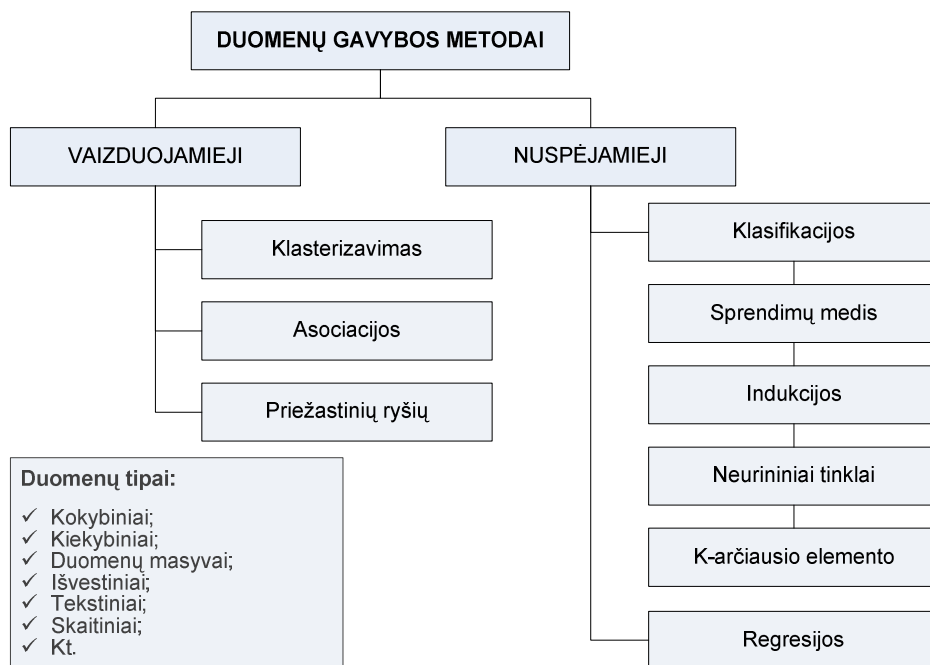
1.1.2. Duomenų gavybos metodai ir jų klasifikacija

Autoriai Sunil, Arte, Chien-Chih, Lin, Puttikan, Prapai, Xiaoshan, pateikia tokią duomenų gavybos metodų klasifikaciją, kurią galima skaidyti į du tipus – vaizduojamieji duomenų gavybos metodai ir nuspėjamieji arba prognozės duomenų gavybos metodai.

Vaizduojamieji duomenų gavybos metodai skirsto duomenis į grupes, atlieka klasifikavimą. Priklausomai nuo to kokie algoritmai yra naudojami metode vaizduojamieji skirstomi į klasterizavimo, asociacijų ir priežastinių ryšių metodus.

Nuspėjamieji duomenų gavybos metodai atlieka duomenų prognozavimą, atlieka kintamųjų priklausomybių tyrimą, vertina kintamųjų faktinės vertes.

Pateiktoje schemeje pavaizduoti fundamentalūs duomenų gavybos metodai. Šių metodų kompleksiškas taikymas veda į naujų jungtinių metodų apibrėžtis.



Šaltinis: sudaryta autoriaus pagal XIAOSHAN, Du.(2006) Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship.

1 pav. Duomenų gavybos metodų klasifikacija

Duomenų gavybos metodų taikymas priklauso nuo taikymo srities specifikos. Todėl norint toliau gilinantis į duomenų gavybos mokslinį pagrindą, pateiksime apibendrintą fundamentalių duomenų gavybos metodų algoritmų apžvalgą.

Asociacijos taisyklės ir klasterizacija. Asociacija naudojama dėsningumams tam tikruose įvykiuose arba procesuose nustatyti, kitaip tariant, ji susieja įvairius vieno įvykio faktus. Turint pardavimų istorinius įvykius, galima suformuluoti pirkimo šablonus

Nepaisant to, kad skaičiavimai yra nesudėtingi, paieškos intervalas gali būti labai didelis, ryšys tarp elementų skaičiaus ir paieškos intervalo didėja eksponentiškai. Skaičiavimui reikia didelio asociacijos taisyklių skaičiaus.

Klasterizaciją galime apibrėžti kaip objektų sujungimą į grupes pagal panašias charakteristikas. Pagrindinis klasterinės analizės tikslas – suskirstyti objektus taip, kad skirtumai klasterių viduje būtų kuo mažesni, o tarp klasterių – kuo didesni. Klasterizacija yra analogiška klasifikacijai, bet skiriasi tuo, kad klasės objektai iš anksto yra nežinomi. Privalumas naudingas tuo atžvilgiu, kad po duomenų klasterizacijos galėsime pritaikyti kitus analizei tinkamus metodus.

Klasifikacijos medžiai. Dažniausiai naudojamas intelektualinės analizės metodas. Jis sutelkia dėmesį į egzistuojančius atributus ir jų elgesį. Klasifikacijos algoritmų pagalba galima klasifikuoti objektus pagal žinomas charakteristikas. Šių algoritmų rezultatą galima išivaizduoti kaip panašią į medį struktūrą. Kiekvienas išsišakojimas reiškia vieną ar kitą atsakymą į klausimą. Tokiu būdu sudaromos taisyklės, kurios nagrinėjamą duomenų aibę klasifikuoja atsižvelgdamos į elementų savybes. Jie sukuria hierarchinę klasifikavimo taisyklių struktūrą. Tam, kad būtų galima nuspręsti, kokiai klasei priskirti tam tikrą objektą ar situaciją, reikia atsakyti į klausimus medžio mazguose, pradedant nuo jo šaknies. Pagal sąlygą, pasirenkama į kurią pusę išeis atsakymas, prie kito mazgo, ir taip kol randamas sprendimas. Duomenų aprašymas – sprendimų medžiai leidžia saugoti informaciją apie duomenis kompaktiškoje formoje, vietoj duomenų galime saugoti patį sprendimo medį. Klasifikacija – sprendimų medžiai tinka klasifikacijos uždaviniams spręsti.

Regresija. Jei kintamasis turi nepertraukiamas reikšmes, sprendimo medžiai leidžia nustatyti kintamojo priklausomybę nuo nepriklausomųjų kintamųjų. Regresija leidžia tyrinėti ryšio formas. Dažniausiai naudojama prognozės modeliams kurti.

Neuroniniai tinklai. Tai netiesiniai modeliai, sėkmingai naudojami klasifikavimo ir prognozavimo uždaviniams spręsti. Jų struktūra primena biologinius neuroninius tinklus. Tinklas sudarytas iš kelių sluoksnių: įvesties, išvesties ir vieno ar daugiau paslėptų. Įvesties sluoksnyje pateikiama elementą aprašanti informacija, o išvesties sluoksnyje gaunamas

rezultatas – priklausymas klasėms (klasifikavimo uždaviniuose) ar numanomas rezultatas (prognozavimo uždaviniuose). Paslėpti sluoksniai realizuoja uždavinio logiką. Kiekvienas tinklo sluoksnio mazgas yra susijęs su kiekvienu šalia esančio sluoksnio mazgu. Neuroniniai tinklai turi galimybę mokytis.

Neuroninių tinklų metodas yra pakankamai galingas ir produktyvus, kai reikia pasiekti aukšto analizės tikslumo. Tačiau aukšto tikslumo siekimas reikalauja didelių išteklių, kurie auga su priklausomybe panašia į geometrinę progresiją. Todėl šis metodas rekomenduotinas, kai reikalinga ypač tiksli analizė, o žadamas pelnas iš rezultatų gavimo leidžia investuoti į brangius skaičiavimų išteklius.

Duomenų gavybos technologijos taikomos klientų valdymo sistemose aprašytos 1 lentelėje.

2 lentelė

Duomenų gavybos metodų apibendrinimas

| | Duomenų tipas | Sprendimas | Privalumai | Trūkumai |
|---|---------------|--|---|---|
| K-arčiausias elementas. <i>Modelio tipas:</i> Nuspėjamas | Numeruotas | Metodo tipas, kurio pagrindas - klasterizavimas. Naudojamas klientų segmentavime | Lengvai pritaikomas | 1. Negali būti naudojamas kelių dimensijų duomenų struktūrose; 2. Galimos klaidos; 3. Duomenų rūšiavimas trunka sąlyginę laiko dalį. |
| Savaiminis žemėlapių sudarymas. <i>Modelio tipas:</i> Vaizduojamas | Numeruotas | Klientų segmentavimui | Gali būti taikoma kelių dimensijų duomenų struktūrose | 1. Susideda iš sudėtingų technikų; 2. Gaunami sunkiai paaiškinami rezultatai. |
| Priežastiniai ryšiai. <i>Modelio tipas:</i> Vaizduojamas | Kategorijos | Pirkinių krepšelių analizei. Geriausiai parduodamų produktų tyrimui. | Lengvai suprantamas | 1. Modeliavimo rezultato duomenys užima daug vietos. Susideda iš sudėtingų technikų; |
| Nuoseklumo įgyvendinimas. <i>Modelio tipas:</i> Nuspėjamas | Kategorijos | Pardavimų numatymui, atitinkamos reklamos, nuolaidų formuluotei. | Lengvai suprantamas | 2. Kuo daugiau transakcijų, tuo daugiau kombinacijų; 3. Modeliavimo rezultato duomenys užima daug vietos. 4. Rezultatai apdorojami rankiniu būdu. |

3 lentelės tęsinys

| | Duomenų tipas | Sprendimas | Privalumai | Trūkumai |
|--|---------------|---------------------------------------|---|---|
| Neuroniniai tinklai. <i>Modelio tipas:</i> Nuspėjamas | Numeruotas | Naudojamas klientų veiksmams numatyti | 1. Tinkamas dideliems duomenų kiekiams apdoroti; 2. Apdoroja duomenis su trūkstamomis reikšmėmis | 1. Sunkiai paaiškinami rezultatai; 2. Skaičių aibė generuojama į aibę nuo 0 iki 1. |
| Sprendimų medis. <i>Modelio tipas:</i> Nuspėjamas | Kategorijos | Naudojamas klientų veiksmams numatyti | 1. Lengvai suprantamos technikos; 2. Apdoroja duomenis su trūkstamomis reikšmėmis | 1. Sudėtingai įgyvendinamas laiko atžvilgiu; 2. Reikalauja skirtingų modelių pritaikymo atskiroms duomenų kategorijoms |

Šaltinis: sudaryta autoriaus pagal Sunil, Arte, Chien-Chih, Lin, Puttikan, Prapai.(2005) Mining Internet user survey data Improving Customer Relationship Management (CRM).

Lentelėje pateikti duomenų gavybos metodų aprašymai tolimesnėje analizėje leis lengviau suvokti duomenų gavybos veikimo principus. Renkantis, kokie metodai tinkamiausi duomenų analizei ryšių su klientais valdymo sistemos aktualijoms spręsti bus remiamasi apibendrintomis duomenų gavybos metodų savybėmis.

1.2.PAGRINDINĖS DUOMENŲ GAVYBOS FUNKCIJOS VERSLO SISTEMOSE

Vystant duomenų gavybą verslo procesuose, analizuojant duomenis ir kuriant metodus svarbu įvertinti duomenų gavybos metodų sudarymo principus.

Netrivialūs rezultatai duomenų gavybos metodų kombinacijose. Klientų valdymo sistemose taikomi kombinuoti, susidedantys ir perdengiantys keletą duomenų gavybos metodų. Kadangi klientų valdymo sistemos yra sudėtingos ir reikalauja įvairių duomenų pjūvių analizės, vieno duomenų gavybos metodo nepakanka. Skirtingais aspektais analizuojami duomenys gali būti naudojami sprendimų priėmimui, pardavimų, klientų elgesio nuspėjimui ir kitai verslo požiūriu aktualiai analizei. Tokiu būdu duomenų pjūviai ir naudojami duomenų gavybos metodai sudaro vieną analizės planą. Kompozicionavimas tai metodas, kuris techniškai realizuoja kompleksines analizes, naudojant standartines operacijas bei informacijos šaltinius. Pavyzdžiui ryšių algebra susideda iš operacijų, kurios *atrenka* eilutes, *suprojektuoja* lentelės stulpelius, *apjungia* kelias lenteles, *sutraukia* eilutes iš panašios

lentelės. SQL, standartinė užklausių kalba, padidina šių operacijų rinkinį, išsaugodama raktines reikšmes, tokiu būdu išsaugoja lenteles (Ramakrishnan, Agrawal, Freytag, 2005).

Duomenų integracija duomenų gavyboje. Duomenys, reikalingi duomenų gavybos procesui, yra skirtingų šaltinių. Daugumoje atvejų sistemose, talpinančiose daug duomenų, informacija yra renkama struktūrizuotose lentelėse, netgi skirtingose duomenų bazėse, o taip pat skirtingose kompanijose. Duomenų gavybos užduotimi tampa šių duomenų rinkimas, sąsajų tarp skirtingų modulių sudarymas. Ypač tai aktualu B2B aplinkoje, tai serveriai gali būti nutolę (Jaideep, 2005).

Nesuderinamų duomenų suderinamumas. Klientų valdymo sistemose dažnai duomenys gali neturėti bendrų atributų ar tarpusavio ryšio, tačiau gavybos metodų rezultatai koreliuoja skirtingus atsitiktinius duomenys, tam, kad gauti reikšmingą ir efektyvų verslui rezultatą. Taip pat duomenų gavybos metoduose turi būti numatomas principas interpretuojant trūkstamus duomenis (Jaideep, 2005).

Privatumo ir konfidencialumo aspektai duomenų gavybos metoduose. Verslo struktūros bei valstybinės institucijos yra atsakingos už duomenų panaudojimą, ypač tai liečia demografinę informaciją, tokia kaip elektroniniai laiškai, skambučių išklotinės ir panašiai. Susirūpinimas duomenų konfidencialumu kelia vartotojo sąsajos klientų valdymo sistemose ribojimą. Šis aspektas yra svarbus būtent B2B aplinkose, kai duomenų bazių vartotojų yra daug (Jaideep, 2005).

Duomenų panaudojimo aspektai. Kai kuriose šalyse yra draudžiama naudoti informaciją tam, kam jos rinkimas nebuvo skirtas. Jei yra renkami apie kiekvieno kliento banke kreditingumą, tai tokia informacija negali būti panaudota jokiems kitiems tikslams. Tad duomenų gavybos metoduose svarbu atsižvelgti į informacijos paskirtį ir galimą naudojimą, bei numatyti pakeitimo galimybes bei duomenų naudojimo sąlygos pasikeistų (Ying, Dezheng, 2003).

Duomenų gavybos metodų atitikimo realiam pasauliui aspektai. Klientų valdymo sistemose informacija gali būti generuojama sprendimų valdymui. Priklausomai nuo organizacijos veiklos srities informacija gali būti svarbi kompanijos egzistavimui. Taikant duomenų gavybos metodu, svarbu įvertinti duomenų, o taip pat ir gaunamų rezultatų rizikingumą. Šis aspektas aktualus finansinėse kompanijose ir medicinos institucijose. Pavyzdžiui: Jungtinėse Amerikos Valstijose yra nelegalu priimti investicinius sprendimus remiantis tik automatizuotais procesais – programomis (Jaideep, 2005).

Klientų elgesio gilios analizės aspektai. Vienas iš pagrindinių klientų valdymo sistemų principų yra – kaip suprasti klientus. Dabartiniai metodai pagrįsti tik pirkinų rezultatais. Šie metodai yra paviršutiniai ir neatspindi gilios klientų analizės. Apibendrinta analizė dažnai duoda klaidingą suvokimą apie rinką bei atvaizduoja klaidingus rezultatus. Duomenų gavybos modeliai numato šios problemos sprendimą klientų modelių kūrimu. Segmentuojami klientai atpažįstami pagal tam tikras bendras savybes. Savybių rinkiniai analizuojami taikant duomenų gavybos algoritmus. Surinktos klientų grupės sudaro klientų modelius. Vieni iš tokių modelių sudarymo kriterijų yra marketingo ir psichologijos mokslo rekomenduojami kriterijai. Pirmiausia analitikas turi pasirinkti kokio gilumo klientų analizė bus atliekama: bendro lygio, segmento lygio, individo lygio. Toliau nustatoma kokiomis dimensijomis bus skaidomi duomenys: demografinėmis, psichografinėmis, makro – mikro elgesio (Jaideep, 2005).

Duomenų gavybos rezultatų įvertinimas. Duomenų gavybos rezultatų sudėtingumas priklauso nuo to, kokie duomenų gavybos metodai, funkcijos, algoritmai buvo panaudoti. Tačiau šiuo metu nėra vienos matavimo priemonės, tam, kad nustatyti, kuris metodas duoda geresnius rezultatus, yra efektyvesnis ir realesnis, atspindintis tikrą klientų elgesį. Taikant kompleksinius metodus, tampa sudėtingiau suprasti kuris metodas turėjo didesnę įtaką duomenų teisingumui ar klaidingumui, todėl labai svarbu griežtai specifiuoti, kokie duomenys bus naudojami, kokio tikimasi rezultato (Jaideep, 2005).

Pastovaus duomenų gavybos naudojimo aspektai. Kai duomenų gavybos metodai pradėjo duoti teisingus rezultatus, kai jie buvo pradėti naudoti praktikoje, iškilo klausimas: kaip vystyti ir pritaikyti metodus kasdieniniam duomenų apdorojimui. Atsižvelgiant į skirtingus dienos duomenų analizės poreikius duomenų gavybos metodų pritaikyme reikalinga interaktyvi vartotojo sąsaja. Taip pat aktualu ir tai, kad duomenų gavybos metodai nebūtų užimantis daug vietos, ar reikalaujantys didelių kompiuterinių resursų. Kasdienis duomenų gavybos metodų naudojimas skatina metodus plėtoti taip, kad jie būtų kompaktiški bei jų apdorojimo laikas trumpas (Ramakrishnan, Agrawal, Freytag, 2005).

Naudojant duomenų gavybos metodus, kompanijos gali modeliuoti prarastus klientų kiekius, analizuoti sprendimus, kurie padėtų klientus išlaikyti. Sprendimo medžio analizės metodai gali padėti išspręsti šia problemą.

Nauda, gaunama iš klientų gali būti skirtinga ir sunkiai prognozuojama. Duomenų gavybos metodai gali numatyti naudos gaunamos iš klientų pokyčius skirtingomis rinkos

sąlygomis. Minėti metodai analizuoja kompanijos istorinius duomenis, rinkos situaciją, randa sprendimus naujiems klientams įgyti (Ying, Dezheng, 2003).

Klientų valdymo sistemos atlieka svarbų vaidmenį kompanijų veikloje. Augant informacijos kiekiams ir jos panaudojimo sritims, didėja statistinio modeliavimo poreikiai. Sprendžiant duomenų analizės problemas, kurios aktualios verslo procesuose, naudojami duomenų gavybos metodai. Todėl toliau darbe analizuojama duomenų analizė ryšių su klientais sistemose

1.3. DUOMENŲ ANALIZĖS SPECIFIKA RYŠIŲ SU KLIENTAIS VALDYMO SISTEMOSE

Šiame skyriuje apžvelgiama ryšių su klientais valdymo sistemų dalykinė sritis. Darbe išskiriama klientų valdymo specifika. Sprendžiamas duomenų gavybos integravimo į klientų valdymo sistemas klausimas. Skyriuje pateikiamas esamų duomenų gavybos metodų ryšių su klientais sistemose vertinimas.

Žinios, kaip intelektualus kapitalas yra svarbiausias įmonių (organizacijų), gaminančių prekes ar teikiančių įvairias paslaugas, turtas. Šio turto strateginis valdymas turi būti nukreiptas jo vertės ir naudingumo didinimo linkme. Fizinio ir finansinio turto valdymas organizacijose yra įprasta kasdienė veikla, tačiau pastaruoju metu vis didesnė svarba teikiama žinioms, kurios yra būtinos įmonių konkurencingumui didinti, inovacinei veiklai plėtoti bei prisitaikyti prie išorinės aplinkos pokyčių. Sparčiai kintančių technologijų eroje didėja įmonių gebėjimai priimti, kaupti, analizuoti ir perduoti informaciją, tačiau vis dar problemiškas lieka šios informacijos pavertimas žiniomis, informacijos gavyba, jos struktūrizavimas, pritaikymas ir išsaugojimas (Klein, 1998).

Intelektualaus kapitalo kūrimas gali būti dvejopo pritaikymo:

- Vidinė informacija organizacijos valdymui;
- Išorinė informacija suinteresuotoms šalims – investuotojams, partneriams, klientams.

Pirmuoju atveju kaupiama intelektualaus kapitalo informacija toliau naudojama valdymo sprendimams priimti. Intelektualaus kapitalo ataskaitose matoma, kokiais žmogiškaisiais ištekliais disponuoja organizacija, kaip yra įgyvendinama žmoniškųjų išteklių plėtra, kokia yra klientų dinamika, darbo organizavimo, išteklių panaudojimas ir kita. Lyginant skirtingų laikotarpių intelektualaus kapitalo ataskaitas organizacijų vadovai gali matyti ar sėkmingai

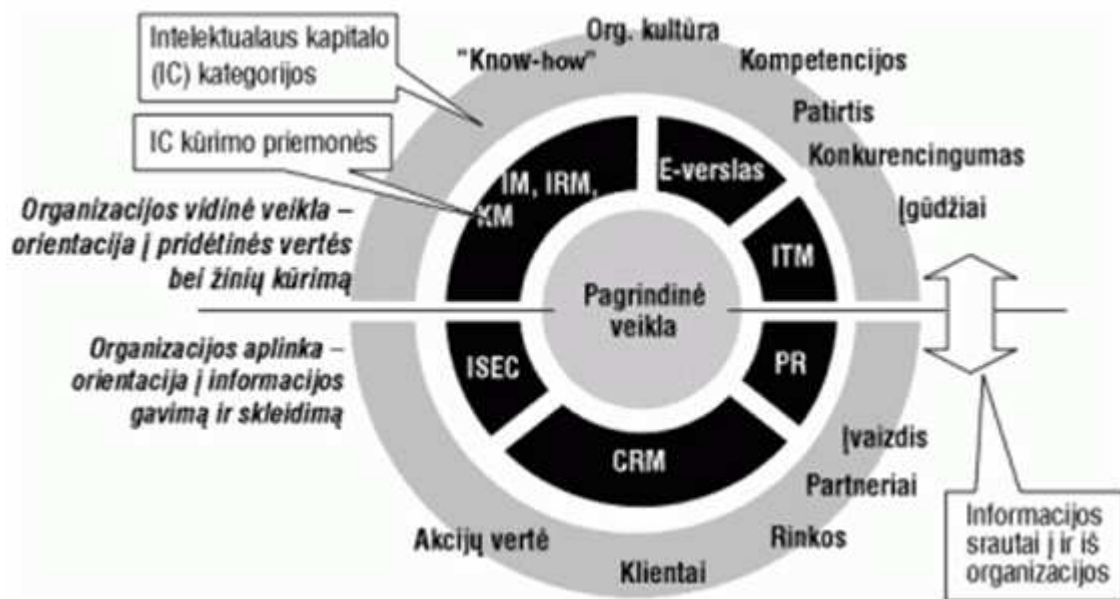
yra įgyvendinama organizacijos strategija, kurie elementai reikalauja didesnio dėmesio, kokios yra galimos tolimesnės perspektyvos (Bieliūnas, 2000).

Antruoju atveju intelektualaus kapitalo informacija yra pateikiama išoriniams interesantams, šiuo atveju dažniausiai intelektualaus kapitalo ataskaita yra teikiama kaip dalis organizacijos metinės atskaitomybės, kartu su finansinėmis ataskaitomis. Intelektualaus kapitalo ataskaitos įtraukimas į organizacijos atskaitomybę pateikia detalesnę organizacijos paveikslą. Tokioje bendroje atskaitomybėje organizacijos partneriai, klientai ir investuotojai mato ne tik finansinius veiklos rodiklius, bet ir organizacijos potencialą, kuris slypi organizacijos darbuotojuose, ryšiuose su klientais, partneriais, kaip sėkmingai yra organizuojamas darbas organizacijos viduje. Papildoma informacija organizaciją daro patrauklesne investuotojams, būsimiems partneriams, klientams ir darbuotojams (Bieliūnas, 2000).

Intelektualaus kapitalo modelį sudaro trys pagrindinės dalys (jų gali būti ir daugiau, priklausomai nuo pasirinkto organizacijos kūrimo modelio):

- Žmogiškasis kapitalas – darbuotojai, jų patirtis, kvalifikacija, išsilavinimas, plėtra ir kt.;
- Ryšių (rinkos) kapitalas – klientai, partneriai, jų dinamika, bendradarbiavimas ir kt.;
- Struktūrinis (organizacinis) kapitalas – infrastruktūra, procesai, novatoriškumas ir kt. (Klein, 1998).

Neplanuoti veiksmai, neišreikštumas, dinamiškos aplinkos neprognozuojamumas taip pat gali būti valdomi - tai atsispindi net ir ekonominėse teorijose. Intelektualus kapitalas yra gana aiškiai apibrėžta ir išmatuojama sąvoka, jos paprasčiausias atspindys - kompanijos akcijų vertė. Organizacijos intelektualaus kapitalo dedamosios parodytos 1 pav. (Bieliūnas, 2000).



Šaltinis: BIELIŪNAS, Martynas. (2000) Žinių vadybos praktinis taikymas: pokyčiai, kuriuos lemia ekonomikos tendencijos.

2 pav. CRM vieta organizacijos intelektualaus kapitalo kūrime

Ryšių su klientais valdymas – tai ne technologija, o visa įmonę apimantis procesas. Tačiau technologijos čia taip pat vaidina ne mažiausią vaidmenį. Įmonę apimanti CRM transformacija ir CRM veiksmingumas didele dalimi priklauso nuo technologijų, ir pirmiausia – nuo pardavimų proceso automatizavimo. Pardavimų proceso automatizavimui skiriama tiek daug dėmesio todėl, kad būtent pardavimų metu geriausiai juntama įvairių į klientus orientuotų projektų rinkodaros ir klientų aptarnavimo bei globos poveikis. Duomenis apie klientus ir įmones ir klientų sąlyčio taškus sugebanti rinkti įmonė šią informaciją perduoda pardavimus vykdančiai komandai, taip sudarydama galimybes sutvirtinti ryšį su klientais ir patobulinti bendravimą. Pradžioje pardavimų automatizavimo sprendimai buvo skirti tam, kad padidintų pardavimų produktyvumą bei tam, kad paskatintų pardavimus vykdančius darbuotojus dokumentuoti ir perduoti informaciją apie pardavimų metu vykdomus veiksmus. Tačiau ilgainiui pardavimų proceso automatizavimo sprendimai vis daugiau dėmesio sutelkia ties ryšiu su klientais palaikymu ir tobulinimu, taip didinant klientu pasitenkinimą (Bielionas, 2000).

1.4. CRM APŽVALGA IR KLASIFIKACIJA

Ryšių su klientais valdymo sistemos (*angl. Customer Relationship Management - CRM*) paskutiniuosius dešimt metų atlieka svarbų vaidmenį kompanijų klientų pozicionavime, atspindi holizminis vaizdą į klientą (pirkėją). Ryšių su klientais valdymo sistemos apima

visapuses žinias apie klientus, pritaikydamos jas marketinge, produkcijos gamyboje, ar paslaugų teikime poreikiams analizuoti. Ryšių su klientais valdymo sistemos duoda kompanijoms galimybę koordinuoti informaciją apie klientus, padidinant galimybę planuoti verslo srities veiklą, atveriant perspektyvas tinkliniui verslui (Ramakrishnan, Agrawal, Freytag, 2005).

Šios sistemos, ar įrankio, esmė yra kliento pažinimas, kurio rezultatas akivaizdus kompanijos veikloje. Ryšių su klientais valdymo sistemos paskirtis yra maksimizuoti kliento išlaikymą, skatinti lojalumą, didinti kompanijos pelningumą. Kai kliento poreikiai, informacija įvertinama teisingai, numatomas potencialus klientas, kai klientai teisingai segmentuojami, tuomet didinamas kliento egzistavimas ryšių valdymo sistemoje (Ramakrishnan, Agrawal, Freytag, 2005).

Ryšių su klientais valdymo sistemos procesai ir funkcijos pavaizduotos 3 priede. Bendrine prasme ryšių su klientais valdymo sistemos įgyvendina šiuos procesus: informacijos analizė, ataskaitų rengimas, atgalinio ryšio su klientais apdorojimas, kampanijų valdymas, pardavimų, paslaugų fiksavimas. Pagrindiniai ryšių su klientais valdymo sistemų funkciniai vienetai: identifikavimas, pritaikymas esamai verslo struktūrai, sąsajos, interakcijos su klientais, nutolusiais padaliniais, informacijos išskaidymas, integravimas (Klein, 1998).

Daugelis klientų informacijos gavybos ir valdymo įrankių gali būti pritaikomi pagal konkrečios įmonės poreikius, jos pardavimų politiką ir procedūras. Tokių įrankių naudojimo įmonėje visuma vadinama klientų informacijos valdymu (Bieliūnas, 2000).

Grafiškai CRM klasifikacija pavaizduota 5 priede, 8 paveiksle – CRM skirstymas pagal valdymo pobūdį. Pagal savo pobūdį CRM skirstomas:

1. *Operatyvinis CRM*. Operatyvinis CRM apima sritis, kuriose įmonė tiesiogiai bendrauja su klientu. Pavadinkime šias vietas “sąlyčio taškais”. Sąlyčio tašku gali būti “įeinantis” kontaktas – pavyzdžiui, kliento skambutis į bendrovės “karštąją liniją” – arba “išeinantis” kontaktas – tarkime, el. paštu klientui išsiųstas reklaminis pranešimas. Dauguma šiandien rinkoje esančių CRM programinės įrangos produktų patenka būtent į operacinio CRM kategoriją. Pagrindiniai informacinės sistemos uždaviniai bei funkcijos:
 - a. klientų aptarnavimas;
 - b. pardavimų automatizavimas;
 - c. marketingo automatizavimas;
 - d. informacija apie klientus, kontaktus, konkurentus, pardavimo galimybes;

- e. prekybos tarpininkų renkamos įvairiapusės informacijos registravimas;
- f. sistemos priderinimas prie užsakovo naudojamo rinkos struktūrizavimo modelio;
- g. informacijos įvedimas tiesiogiai, programų eksportas bei importas
- h. užduotis prekybos tarpininkams bei jų vykdymo kontrolei;
- i. lanksčiai plečiamos duomenų struktūros;
- j. duomenų įvedimo kontrolės administravimas;
- k. generuojamos užklausos ir pateikiama informacija (Zavareh, 2007).

2. *Savitarnos (Bendradarbiavimo) CRM*. dar vadinamas “strateginiu”, leidžia suprasti kliento veiksmus. Analitinio CRM diegimui reikalingi adekvatūs IT sprendimai, leidžiantys surinkti ir apdoroti kalnus analizei reikalingos klientų informacijos. Jam taip pat reikalingi nauji verslo procesai, kuriais siekiama patobulinti klientų aptarnavimo praktiką, skatinant jų lojalumą ir didinant pelningumą. Ekspertų ir klientų spaudžiami, dauguma CRM programinės įrangos gamintojų šiandien skuba patys sukurti analitinio CRM produktus arba mėgina įtraukti šias galimybes į savo produktus sudarydami partnerystės susitarimus su analitinės verslo informacijos (BI – Business Intelligence) IT sprendimų tiekėjais. Verslo išvalga (Business Intelligence (BI)) yra didelių duomenų kiekių surinkimas, valdymas ir analizė, siekiant įgyvendinti strateginius verslo sprendimus ir papildyti esamus procesus naujomis funkcijomis. Taikomieji uždaviniai:

- a. informacijos kaupimas duomenų bazėse;
- b. informacijos apibendrinimas bei analizė;
- c. reikiamų duomenų apdorojimas ir paruošimas analizei;
- d. informacijos pateikimas pagal analitinius pjūvius, laikotarpius ir pan.;
- e. papildomas pardavimo galimybių identifikavimas (upselling);
- f. kryžminio pardavimo galimybių identifikavimas (cross-selling) (Zavareh, 2007).

3. *Analitinis CRM*. Tai verslas - verslui (B2B business to business) ir verslas - vartotojui (B2C business to consumer) internetinės savitarnos sistemos, kur klientas ar vartotojas didžiąją dalį procedūrų atlieka pats. Taikomieji uždaviniai:

- a. galimybė partneriams stebėti sandėlių būklę, vėliausias prekių kainas
- b. galimybė klientams patiems įvesti užsakymus

- c. galimybė klientams sekti užsakymų ir užsakytų prekių pristatymo būklę
- d. galimybė klientams matyti savo kredito ir likučio informaciją.
- e. galimybė klientams patiems atsispausdinti sąskaitas ir važtaraščius
- f. galimybė klientams turėti realaus laiko, susietą su verslo valdymo ir apskaitos sistema internetinę parduotuvę (Zavareh, 2007).

Pagal aptartą CRM klasifikaciją, duomenų gavybos metodai taikomi analitiniuose ryšių su klientais valdymo sistemose. Tai strateginio lygio informacijos analizės priemonė, kuri generuoja rezultatus padedančius priimti marketingo, organizacijos valdymo bei pardavimų sprendimus.

Toliau darbe pateikti duomenų gavybos metodai taikytini informacijos apdorojimui bei duomenų prognozei, vizualizuota, kurioje ryšiais su klientais valdymo sistemos lygmenyje naudojami šie metodai. Iš pristatytų duomenų rinkties bei metodų taikymo specifikos, tolimesnėje analizėje bus pasirinkta tyrimo koncepcija.

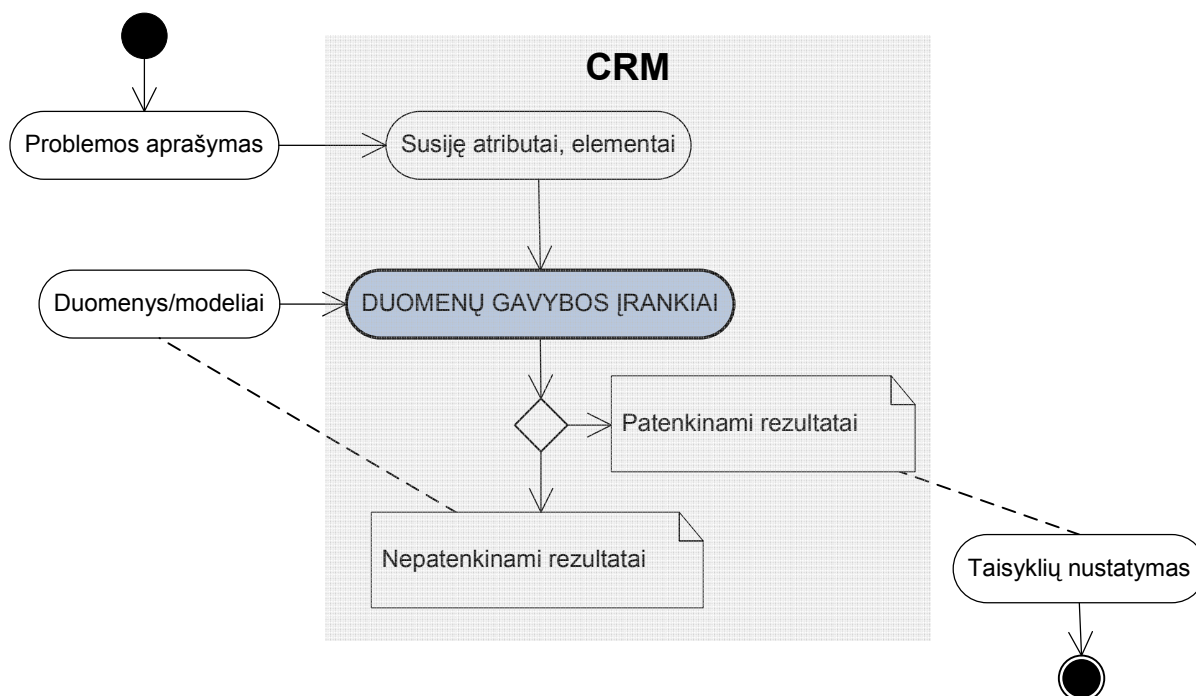
1.5. DUOMENŲ GAVYBOS METODŲ TAIKYMOS CRM VERTINIMAS

Duomenų gavybos technologijos apima kiekvieną kliento gyvavimo ciklo stadiją, įskaitant naujų klientų paiešką, naudos gavimą iš esamų klientų, bei klientų išlaikymą. Duomenų gavybos metodai gali padėti kompanijoms rasti klientų panašumų, taip juos grupuojant ir tiekiant jiems naudingiausių produktus ir paslaugas. Taip pat duomenų gavybos metodai gali būti pritaikomi sprendimų priėmimo, verslo valdyme.

Klientų valdymo sistema atspindi kompanijos veiklą, bei tarnauja daugeliui procesų, tokių kaip klientų paieška, poreikių tyrimas, reklamos sprendimai, marketingo uždaviniai, pardavimo analizės ir daugelis kitų. Verslo procesuose, o ypač tinklinio verslo, B2B aplinkos kompanijoms turi didelę reikšmę operuojamų duomenų analizė. Pjūviai, kuriais galima atlikti rinkos tyrimus, nustatyti, kiek vartotojų perka tam tikrą prekę, neužtenka vadovautis tik pardavimų duomenimis. Problema egzistuoja tame, kad vieno tipo duomenys dažnai duoda klaidingus analizės rezultatus, neatlieka funkcijų perspektyvoms numatyti. Uždavinio sprendimas yra statistinis modeliavimas, kurio dalykinė sritis duomenų gavybos metodai (Ying, Dezehen, 2003).

Analizuojant duomenų gavybos metodus, tinkamus klientų valdymo sistemoms, svarbu žinoti, koks yra modelių sudarymo principas.

Duomenų gavybos modelio procesas klientų valdymo sistemoje pavaizduotas 1 paveiksle.



Šaltinis: sudaryta autoriaus

3 pav. Duomenų gavybos metodo veikimo algoritmas klientų valdymo sistemoje

Duomenų gavybos modelių pagrindas yra problemos aprašymas t.y. nustatoma, kokių rezultatų tikimasi, kokie duomenys bus naudojami proceso metu. Nustačius probleminę sritį, įvertinami visi susiję duomenų atributai, surenkami elementai, lentelių sąrašai – paruošiami duomenys. Duomenų gavybos įrankiai – tai funkcijos ir algoritmai, kurie sudaro duomenų gavybos metodą. Duomenų gavybos įrankių rezultatas gali būti patenkinamas arba nepatenkinamas, priklausomai nuo to, kokia buvo probleminė sritis. Toliau, duomenų gavybos metodų plėtojimui sudaromi taisyklių nustatymai, į kuriuos atsižvelgiama toliau taikant duomenų gavybos metodus (Jaideep, 2005).

Kasdienis duomenų gavybos metodų naudojimas skatina metodus plėtoti taip, kad jie būtų kompaktiški bei jų apdorojimo laikas trumpas (Ramakrishnan, Agrawal, Freytag, 2005).

Naudojant duomenų gavybos metodus, kompanijos gali modeliuoti prarastus klientų kiekius, analizuoti sprendimus, kurie padėtų klientus išlaikyti. Sprendimo medžio analizės metodai gali padėti išspręsti šia problema.

Nauda, gaunama iš klientų gali būti skirtinga ir sunkiai prognozuojama. Duomenų gavybos metodai gali numatyti naudos gaunamos iš klientų pokyčius skirtingomis rinkos

sąlygomis. Minėti metodai analizuoja kompanijos istorinius duomenis, rinkos situaciją, randa sprendimus naujiems klientams įgyti (Ying, Dezheng, 2003).

Apžvelgus CRM sritis, kuriose taikomi duomenų gavybos metodai, prieiname išvados, jog iki šiol atliktuose moksliniuose tyrimuose duomenų analizė taikoma naudojant duomenis esančius ryšių su klientais valdymo sistemose. Atsižvelgiant į verslo srities specifiką siūloma keisti įvesties duomenų pobūdį nesiplečiant į kitas verslo sritis, bet giliau analizuojant į klientus orientuotą informaciją. Remiantis šia informacija svarbu atsižvelgti į duomenų analize, ne bendrine prasme, o atsižvelgiant į tiriamos srities verslo ypatumus – duomenų specifiką, klientų naudos analizės ypatybes ir kt.

2. DUOMENŲ GAVYBOS SPECIFIKA B2B APLINKOJE

Duomenų gavybos metodų taikymas tolesniame tyrime bus apspręstos B2B aplinkos specifikos. Šiame skyriuje pateikiamos įmonės veiklos funkcijos, kurioms aktualiausias aspektas – informacijos analizė.

2.1. PARDAVIMŲ KANALŲ OPTIMIZAVIMAS

Nesvarbu kokia yra verslo įmonės veikla, ar prekyba, ar statyba, ar gamyba, jos veiklos rezultatai priklauso nuo pardavimų funkcijos įgyvendinimo. Tobulinant šią funkciją arba kanalą, atsiveria naujos galimybės:

- Didėja įmonės pajamos, sukuriant elektroninį pardavimų kanalą,
- Mažėja pardavimų ir klientų aptarnavimo kaštai, pertvarkant ir automatizuojant procesus;
- Pardavimų kanalo tobulinimas gali būti panaudojamas, kaip įmonės atvirumo informacinėms technologijoms įvertinimas, t.y. įmonė gali suprasti, kaip reikės pakeisti savo elgseną, kad galėtų imtis sudėtingesnių elektroninio verslo santykių ir elektroninių ryšių su savo veiklos partneriais (LEPA, 2007).

Informacinių technologijų pasitelkimas atveria galimybes įmonėms mažesnėmis sąnaudomis pasiekti tolimas rinkas ir bendrauti su jose esančiais klientais. Išplėtotos elektroninės komunikacijos priemonės faktiškai sukūrė globalias rinkas. Daugelis įmonių naudojami elektroniniu verslu, kad pradėtų pardavinėti savo prekes tolimose šalyse ir regionuose, kuriuose turėti “gyvų” pardavėjų komandas būtų per brangu (dėl aukštų darbo kaštų arba nepakankamų apyvartų). Dirbant su tolimomis rinkomis pardavimo darbuotojų naudojami tradiciniai klientų paieškos ir pardavimo metodai būtų žymiai brangesni nei pasitelkiant elektroninius kanalus. Prekiauti B2B rinkoje siekiančios įmonės turi priimti daugybę sudėtingų verslo sprendimų, liečiančių užsienio šalyse gyvenančių klientų aptarnavimo ir prekių garantinio remonto ar keitimo kaštus.

Vartotojų informacijos valdymas tradicinėmis priemonėmis nesuteikia galimybių informaciją pasiekti greitai ir patogiai. Kartais informacijos trūkumas arba jos nepasiekiamumas tam tikru metu gali sąlygoti didesnius pardavimų kaštus ar galimų pajamų praradimą.

2.2. KLIENTŲ APTARNAVIMO TOBULINIMAS

Elektroninėmis priemonėmis savo klientus aptarnaujančios įmonės siekia kiek įmanoma labiau automatizuoti pagalbos klientams teikimo procesą.

Kad įmonės galėtų sėkmingai įdiegti elektroninio klientų aptarnavimo sprendimus, įmonėse turėtų būti naudojamos duomenų saugyklų ir duomenų analizės technologijos. Duomenų saugykla gali būti apibrėžiama, kaip specialiam tikslui sukurta duomenų bazė, kurioje laikomi suskirstyti ir apdoroti veiklos duomenys, surinkti iš įvairių įmonės operacinių duomenų bazių. Turėdamos gerai sutvarkytas duomenų saugyklas, įmonės gali analizuoti duomenis, ieškodamos universalių sąryšių ir pasikartojančių bruožų, kurie slypi įmonės duomenų bazėse, tačiau dėl didelio jų kiekio nėra lengvai pastebimi.

Duomenų analizė naudojama tiriant didžiulius kiekius istorinių įmonės veiklos duomenų, siekiant pastebėti tendencijas ir iki šiol neatrastas galimybes. Suprasdamos savo bendravimą su esamais klientais, įmonės gali gauti žinių, kaip projektuoti, gaminti ir aptarnauti savo produktus bei aptarnauti klientus, kad šie būtų labiau patenkinti. Augant elektroninio pardavimų kanalo populiarinimui, pardavimo ir vėlesnio klientų aptarnavimo kaštai pradeda sparčiai mažėti. Todėl ypač svarbu ieškoti tinkamų inovatyvių informacijos analizės ir duomenų apdorojimo būdų.

2.3. PIRKIMO (TIEKIMO) KANALŲ TOBULINIMAS

Efektyviai apdorota informacija ir jos analizė B2B versle leidžia kontroliuoti pirkimus.

B2B įmonėje įvedus informacijos analizės procedūras, matoma nauda mažinamų kaštų forma:

- Poreikio nustatymas. Atsakingi įmonės darbuotojai, remdamiesi sistemomis su duomenų analizės metodais, nustato, kokių priekių ar paslaugų reikia įsigyti.
- Pirkimus vykdančios įmonės darbuotojai turi laikytis iš anksto su tiekėjais sudarytų sutarčių. Tokiu būdu užtikrinama, kad nebus perkama iš atsitiktiniu šaltinių neretai padidintomis kainomis. Klientų duomenų analizė įgalina daryti prielaidas, koks klientas, kokią prekę pirks.
- Turėdama tikslius duomenis apie tai, kiek jos darbuotojai išleido pirkdami prekes ar paslaugas iš konkretaus tiekėjo, įmonė gali išsiderėti geresnes tiekimo sąlygas per sekančias derybas su tiekėjais.

- Suformuoti ir pateikti mažus užsakymus įmonei kainuoja tiek pat, kiek ir didelius. Įvedus elektroniniu pirkimu procedūras, atsiranda galimybė automatiškai sudėti pirkimus ir atitinkamai mažinti paties priekiu ar paslaugu užsakymo proceso kaštus. (LEPA, 2007).

Elektroninių pirkimų su paremtais klientų valdymo duomenų gavybos metodais sistemas įsidiegusių JAV ir Vakarų šalių bendrovių patirtis rodo, maždaug dvigubai mažesnės apimties ekonominę naudą teikia galimybė derėtis su tiekėjais dėl geresnių tiekimo sąlygų, ir maždaug dešimt kartų mažesnę efektą duoda paties pirkimų proceso kaštų sumažėjimas (LEPA, 2007).

Tiekimo grandinė (Supply chain) yra koncepcija apibūdinanti žaliavų, informacijos, pinigų ir paslaugų judėjimą nuo tiekėjo (per gamintojus) iki galutinių vartotojų. Tiekimo grandinės funkcijos yra planuoti, organizuoti ir koordinuoti visus tiekimo grandinės veiksmus.

Į tiekimo grandinę taip pat įeina organizacija ir procesai, kurie sukuria ir pristato produktą, informaciją ir paslaugas galutiniams vartotojams (OKSL, 2008).

Tiekimo grandinė gali būti išskaidoma į tris dalis:

1. „Prieš srovę“ (upstream) – apibūdinanti įmonės veiklas susijusias su jos pirmo lygio tiekėjais (gamintojais ar surinkėjais). Ši grandinės dalis gali būti pratęsta iki pat pirminių žaliavų lygio.
2. Vidinė tiekimo grandinės dalis – susijusi su visomis įmonės veikloje vykdomomis veiklomis, kuriuose įėjimus transformuoja į išėjimus.
 - Gamybos valdymas;
 - Gamyba;
 - Žaliavų valdymas (sandėlių valdymas).
3. „Pasroviui“ (downstream) – tiekimo grandinės dalis, susijusi su produkto pristatymu galutiniam vartotojui.
 - Sandėliavimas;
 - Platinimas (distributoriai);
 - Transportas ir logistika;
 - Klientų aptarnavimas.

Tiekimo grandinės srautai. **Materialūs srautai** yra fizinių produktų, naujų žaliavų ir medžiagų srautai tiekimo grandinėje. (Gražinami, perdirbami, atliekos). **Informacijos srautai** susiję su visais materialiniais srautais. (Poreikiai, gabenimai, užsakymai, gražinimai ir t.t.). Taip pat informacija, kuri naudojant vykdoma įmonės kontrolė. **Finansų srautai** apima pinigų pervedimus, apmokėjimus, kreditinių kortelių informaciją, apmokėjimo grafikus, elektroninius apmokėjimus ir su kreditais susijusius duomenis (OKSL, 2008).

IT tiekimo grandinės dalis:

1. Vidinė dalis - transakcijų apdorojimo sistemos (TPS) apimančios visas funkcinės sritis ir kitos įmonės vidines IS.
2. „Upstream“ – palengvinančios pirkimo procesus ir santykius su tiekėjais. Tai e-prekybas, CRM.
3. „Downstream“ – CRM, užsakymai ir pristatymai.

Tiekimo grandinės valdymo (Supply chain management - SCM) programinė įranga kuriama tam tikroms tiekimo dalims automatizuoti – gamybos, atsargų valdymo, tvarkaraščių sudarymo ir transporto. Tai padeda priimti sprendimus, optimizuoti, modeliuoti ir analizuoti problemas.

SCM padeda prognozuoti ir efektyviau valdyti žaliavų kiekius sandėlyje, verslo procesus ir vartotojų aptarnavimą. Šiuolaikinio SCM tikslas yra sumažinti netikrumą ir riziką tiekimo grandinėje. Tai veda prie padidėjusio pelningumo ir konkurentabilumo.

Norint efektyviai valdyti tiekimo grandinę būtina integruoti visas ją sudarančias sistemas:

1. Vidinis. Programų ir duomenų esančių įmonės viduje integravimas.
2. Programinis. Programų ir DB eksploatuojamų verslo partnerių įmonėse integravimas. Ypač svarbus B2B santykiams.
3. Integruota vertės grandinė (sąveika, collaboration). Keletas įmonių, dirbančių viename segmente, kartu planuoja, diegia ir valdo prekių, paslaugų ir informacijos srautus, visos tiekimo grandinės ribose, siekdamos padidinti galutinio vartotojo gaunamą naudą (OKSL, 2008).

Informacinių technologijų pasitelkimas tobulinant įmonės veiklos procesus kartais reikalauja nemažų investicijų, tiek materialiuju, tiek žmogiškuju. Tačiau efektyviai naudojant

šiuolaikines technologijas galima sėkmingai didinti gaunamas pajamas ir mažinti veiklos išlaidas, formuoti įmonės veiklos rezultatą – didėjanti pelną (LEPA, 2007).

2.4. VARTOTOJŲ ELGSENOS PROGNOZAVIMAS

Vartotojų elgsenos prognozavimas leidžia rinkodaros padaliniiui nustatyti, ką įmonės klientai bus linkę daryti ateityje. Prognozuojant klientų ateities elgseną, moderniu modeliavimo ir duomenų analizės technologijų pagalba tyrinėjamas vartotojų elgesys praeityje. Vartotojų elgsenos analizė apima keletą variacijų:

- Polinkio pirkti analizė. Atliekama siekiant suprasti, kokius produktus konkretus vartotojas apskritai būtų linkęs pirkti;
- Artimiausias pirkinys. Prognozuojama, kokį sekantį produktą klientas pasiruošęs pirkti;
- Produktų sąsajų analizė. Išsiaiškinama, kokie produktai bus perkami drauge. Tai tarytum mėginimas atrasti sąsajas tarp produktų tyrinėjant pirkinių krepšelį;

Žinant, kaip klientas bus linkęs elgtis ateityje, galima imtis daugelio tipinei situacijai pritaikytų rinkodaros veiksmų, pavyzdžiui:

- Iš anksto siūlyti nuolaidas ar naikinti dalį mokesčių klientams, artėjantiems link “pabėgimo rizikos zonos”;
- Tikslines rinkodaros kampanijas taikyti į mažesnius ir tikslesnius vartotojų segmentus;
- Pateikti tam tikrus produktus kartu ir siūlyti juos už fiksuotą kainą, taip parduodami daugiau ir pelningiau;
- “Kryžmai” pardavinėti tuos produktus, kuriuos vartotojai linkę įsigyti kartu.

Kaip ir kryžminiu pardavimu atveju, tokios analizės ir po jos planuojamų veiksmų sėkmė priklauso nuo to, ar įmonė aiškiai suvokia, kas yra jos geriausi klientai (LEPA, 2007).

2.5.DUOMENŲ GAVYBOS METODŲ TAIKYMO CRM APŽVALGOS IŠVADOS

1. Apžvelgus mokslinius tyrimus svarbios duomenų gavybos taikymo sąlygos:
 - 1.1. Duomenų gavybos metodo algoritmas neturi naudoti daug sistemos resursų, metodo algoritmas turi būti nesudėtingas;
 - 1.2. Gavybos rezultatas turi būti aktualus tinklinei sistemai ir universalus pritaikymo prasme;
 - 1.3. Gavybos metodų naudojimas turi turėti ilgalaikes perspektyvas, kadangi verslas verslui sistemos diegimas reikalauja finansinių investicijų;
 - 1.4. Taikant kompleksinius metodus, tampa sudėtingiau suprasti, kuris metodas turėjo didesnę įtaką duomenų teisingumui ar klaidingumui, todėl labai svarbu griežtai specifiuoti, kokie duomenys bus naudojami, kokio tikimasi rezultato.
2. Atlikus CRM klasifikacijos apžvalgą pastebėta, kad efektyviausi informacijos analizės rezultatai naudojant duomenų gavybos metodus gaunami analitinėse ryšių su klientais valdymo sistemose;
3. Išskirtos trys tiekimo grandinės dalys – klientų kreditingumas, prekių kiekio kontrolė, bei prekės pristatymo laiko įtaka ryšiams su klientais. Šiais trim aspektais tolimesnėje darbo dalyje bus tiriamas duomenų gavybos metodo taikymas ryšiais su klientais valdymo sistemoje.
4. Duomenų gavybos pritaikymo ryšių su klientais valdymo sistemose B2B aplinkoje tyrimo sėkme bus laikoma ryšiais su klientais sukūrimas, naudojantis pateikto koncepcinio modelio išeigos informacija, taikant vadybinius bei marketinginius veiksmus.

3. DUOMENŲ GAVYBOS TAIKYMO CRM B2B METODO SIŪLYMAS

Klientų valdymo sistemos dažnai veikia verslo aplinkoje, kuri glaudžiai kontaktuoja su kita verslo sritimi, atšaka ar partneriu. Tokios kompanijos vykdo plačią veiklą, kuri generuoja didelius duomenų kiekius. Informacijos apibendrinimui, o ir statistiniu atžvilgiu reikalingi analizei būtina turėti priemones. Kaip klientų valdymo sistemos, taip ir bendroje B2B aplinkoje galima taikyti duomenų gavybos metodus, pradinis duomenis – klientų informaciją, pakeičiant kitais, aktualiais duomenimis.

Dažnai vadybininkai abejoja duomenų gavybos metodų pritaikymu B2B verslo procesuose, teigdami, kad verslo sritis per daug unikali – taikomi unikalūs verslo modeliai.

Duomenų gavybos metodų aspektu svarbu ne kiek unikalus ir specifinis yra verslo modelis, bet kokie duomenys, kokia informacija yra naudojama kasdieninėje veikloje. Netgi duomenys sugeneruoti iš atsiskaitymo transakcijų yra tinkami duomenų gavybos metodams pritaikyti. Sąskaitos talpina duomenis apie klientą, apie įsigytas prekes/paslaugas, jų kieki ir t.t. Duomenų elementai, turintys ilgalaikę reikšmę, tokie kaip grynasis pelnas, parduotų prekių kiekis gali būti disponuojamas kaip duomenų gavybos metodų informacijos šaltinis (Powerin analytic desisions Inc., 2006).

Vystant geras analitinio modeliavimo priemones, didesnę įtaką rezultatų teisingumui turi duomenų kokybė nei kiekybė. Daugumoje duomenų gavybos modelių ir statistinio modeliavimo priemonių reikalauja tik 400 duomenų bazės eilučių informacijos, tam, kad modelis atliktų efektyvią funkciją.

Per eilę duomenų gavybos modelių vystymosi metų duomenų gavyba, kaip technologija pasiekė stabilios ir subrendusios struktūros statusą. Yra sukurtų daug gerą analitini B2B darbą atliekančių duomenų gavybos įrankių. Tačiau reiktų akcentuoti, kad skirtumas tarp sukurtų duomenų gavybos verslas-verslui įrankių glūdi tame, kiek gauti rezultatai atspindi realaus pasaulio vaizdą.

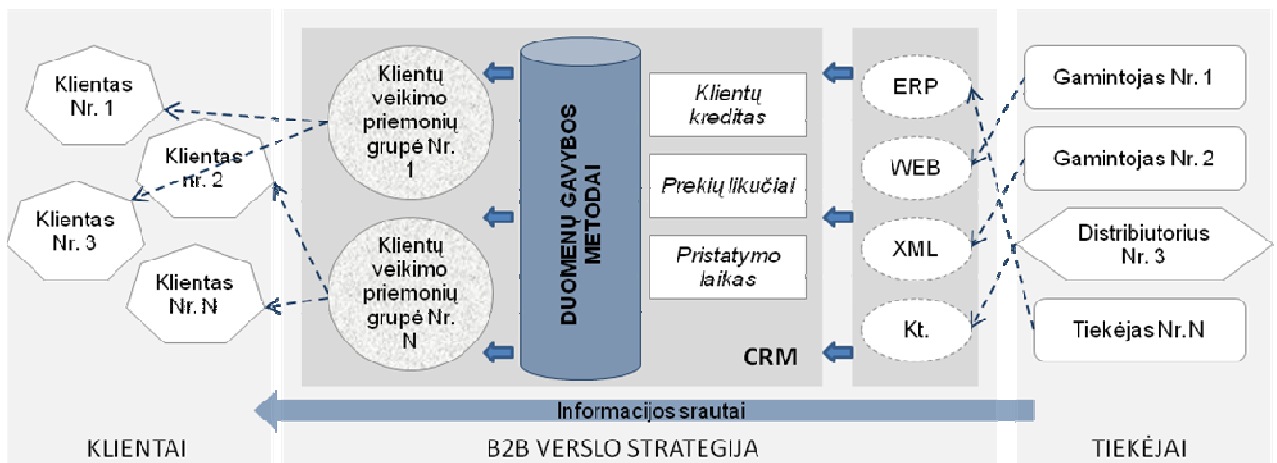
Koncepcinis modelis. Bendra B2B verslo aplinka susideda iš kompanijų, kurios susijusios tarpusavyje tam tikromis interakcijomis. Ryšiai tarp kompanijų grindžiami elektroninėmis transakcijomis. B2B verslo aplinkos esmė – automatizuoti ryšiai tarp kompanijų-partnerių, kompanijų-gamintojų, kompanijų-tiekėjų, kompanijų-vežėjų, kompanija-pirkėjų (Powerin analytic desisions Inc., 2006).

Pastebimas duomenų gavybos modelių paklausos marketingui ir vadybiniam darbui augimas. Išaugo paklausa tokiems modeliams, kurie gali apdoroti nepilnus duomenis, interpretuoti trūkstamą informaciją. Kitas aspektas duomenų gavybos modelių augimo tendencijos yra rankų darbo minimizavimas. Anksčiau informacijai analizuoti reikdavo statistikos žinių turinčių darbuotojų. Dabar siekiama automatizuoti šiuos procesus taikant įvairius statistinio modeliavimo principus. B2B verslo aplinkoje automatizavimas įmanomas atskiriant verslo procesus kurie bus statistiškai modeliuojami, bei pritaikomas jiems efektyviausias duomenų gavybos modelis (Powerin analytic desisions Inc., 2006).

Informacinių technologijų pasitelkimas atveria galimybes įmonėms mažesnėmis sąnaudomis pasiekti tolimas rinkas ir bendrauti su jose esančiais klientais. Tai svarbu B2B kompanijose, kadangi į masinį vartotoją orientuotos bendrovės dažniausiai kuria paprastus verslo pasiūlymus ir aiškiai formuluoja, kuo išskirtinis jų produktas. Dėl to joms paprasčiau pritraukti vartotoją ir paskatinti jį. Bendrovės, aptarnaujančios verslą, atvirkščiai, dažniausiai turi sudėtingesnius produktus ar paslaugas, o jų vertė nėra aiškiai matoma. Kad tam tikras prekių ženklas įsitvirtintų rinkoje, jį būtina atskirti nuo konkretaus produkto ar paslaugos. Ryšys su vartotoju turi būti užmegztas aukštesniame lygmenyje. O tai padaryti yra sudėtingiau.

Iki šiol B2B aplinkoje tiekimo grandinės valdymo programinė įranga kuriama tam tikroms tiekimo daliems automatizuoti - gamybos, atsargų valdymo, tvarkaraščių sudarymo ar transporto.

Tam, kad efektyviau valdyti prekių kiekius, verslo procesus ir vartotojų aptarnavimą, reiktų panaudoti priemones, galinčias analizuoti skirtingo tipo duomenis. Šios priemonės, kaip anksčiau buvo aptarta, yra duomenų gavybos metodai, bei jų vieta kompanijos informacinėje sistemoje - ryšių su klientais valdymo sistema.



Šaltinis: sudaryta autoriaus

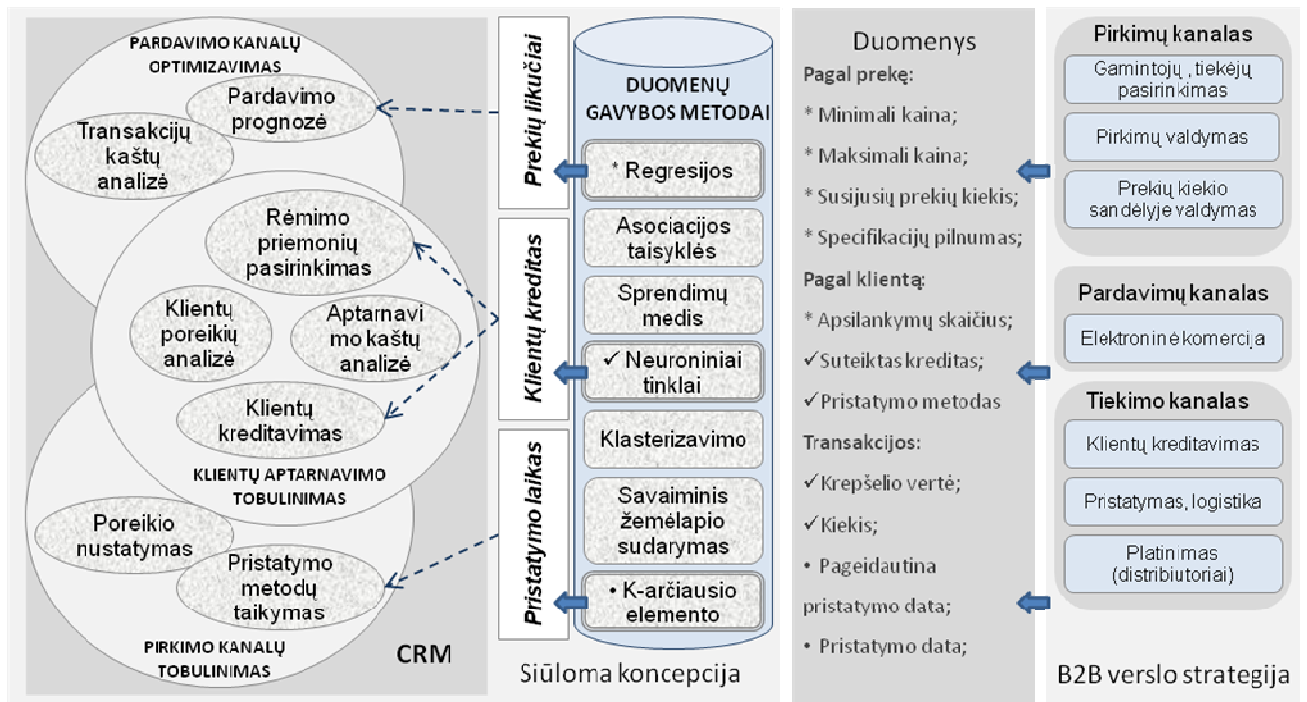
4 pav. Duomenų gavybos metodas B2B aplinkoje

Priešingai nei tiekimo grandinės valdymo sistemose, pasirenkame ne vieną o kelias sritis, kurios duomenys bus apdorojami duomenų gavybos metodais.

1. Pirkimų kreditavimas, nustatomas klientams. Tai klientų valdymo aspektas, leidžiantis ranguoti kompanijas – pirkėjus pagal galimybę bei poreikį įsigyti prekes. Tiekimo grandinės programinė įranga neapsprendžia, kokį optimalų kreditą suteikti kompanijai greitai besikeičiančiomis verslo sąlygomis. Taip pat ir standartinė klientų valdymo sistema neapsprendžia, kokius kriterijus taikyti teikiamo kredito dydžiui apskaičiuoti.
2. Prekių kiekiai. Tai tiekimo grandinės aspektas glaudžiai susijęs su pardavimų planavimu. Prognozuojant kiek bus parduota tam tikrų prekių, nustatoma, kiek jų reikia užsakyti iš gamintojų, o gamintojams jas pagaminti. Klientų valdymo sistemos požiūriu prekių kiekis matomas kaip pardavimo transakcijos bei kliento poreikio pirkti numatymas. Dabartinėje B2B aplinkoje prekių kiekį įtakoja daugiau faktorių – tai gamintojų pardavimų diktavimas, užbaigiant tam tikrą prekių liniją, papildomas tų prekių pardavimų skatinimas galutiniams vartotojams. Taip pat prekių kiekio valdymą įtakoja sandėlio prekių likučiai, gabenimo ypatybės (tam tikri gamintojai finansuoja prekės pristatymo išlaidas, patys nustato pristatymo būdus, o tai sąlygoja greitesnį tos prekės pristatymą taip sukurdamas jų gaminamų spausdintuvų privalumą) ir t.t. Visi šie faktoriai įtakoja marketingo sprendimus atliekamus ryšių su klientais valdymo sistemose.
3. Tvarkaraščių, gabenimo laiko valdymas. Tai vienas svarbiausių aspektų, sukuriančių pridėtinę vertę B2B verslo rinkoje. Pardavimai, fiksuojami ERP

sistemose, veikia kaip indikatorius – įvertina sąlygas, kurios klientui buvo palankios. Pagal tai, galima vertinti ar pristatymo laikas yra optimalus tam tikram klientui. Prekių pristatymo laiko kontrolė leidžia pritaikyti marketingo veiksmus atskiram klientui.

Apjungiant šiuos aspektus bei jų įgyvendinimą, pateiksiu koncepcinį modelį, kuriuo toliau vadovausiuosi atliekant eksperimentinę šio darbo dalį.



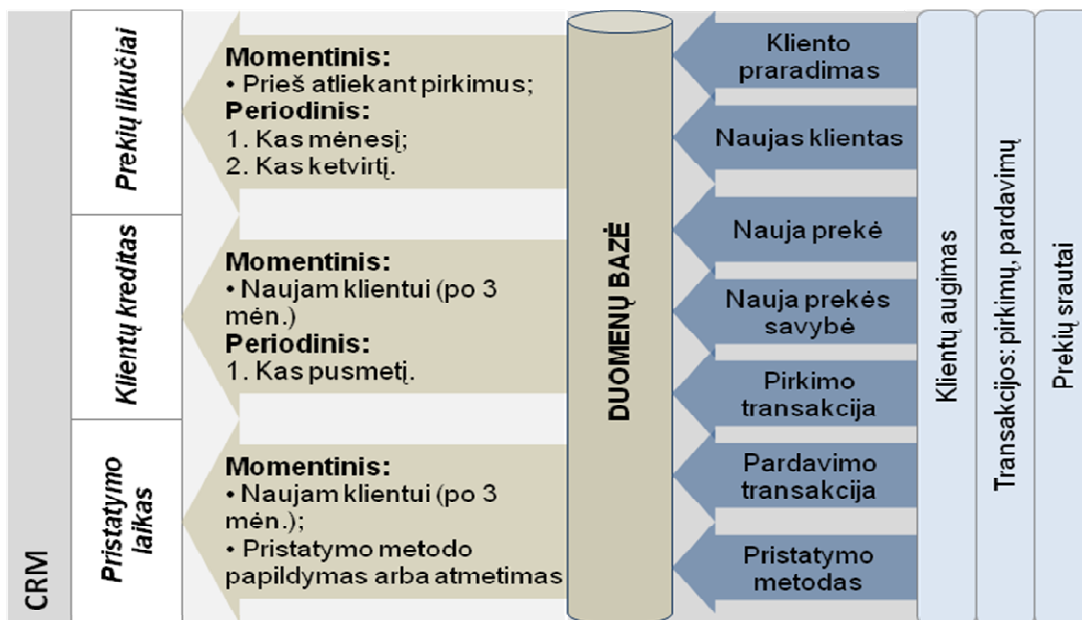
5 pav. Statinis modelio įgyvendinimas

Koncepciniame modelyje apibrėžta, kokius B2B verslo procesus apims tyrimas, kokie duomenų gavybos metodai gali būti naudojami tyrimui įgyvendinti, bei preliminariai nustatyti informacijos judėjimo srautai. Informacija gaunama iš tiekėjų per transakcijų priemones – el. komercijos, ERP sistemas. Ryšiais su klientais valdymo sistemoje atrenkami susiję duomenys su klientų kreditavimu, prekių likučiais bei pristatymo laiku. Turimi duomenys apdorojami pasirinkta analizės priemone – duomenų gavybos metodu. Gauta informacija taikoma atliekant vadybinius bei marketinginius veiksmus klientams.

Dinaminė modelio realizacija yra periodinė, momentinė bei nuolatinė. Įvedami duomenys atnaujinami kiekvieną kartą, kai atsiranda naujas klientas, prekė, pardavimo transakcija ar modifikuojami įrašai. Visi duomenys yra saugomi duomenų bazėje. Kaupiami yra tik tiek duomenys, kurie reikalingi analizei. Atlikus analizę įvesties duomenys yra ištrinami, tačiau saugomos analizės ataskaitos.

Pagal įvesties duomenis atliekama duomenų analizė pasiūlytu metodu pagal ataskaitinį laikotarpį ar momentiniu atveju. Visais atvejais analizė atliekama vadybininkų, gebančių pasinaudoti analitinės procedūros rezultatais bei taikyti reikiamus sprendimus marketingo, finansų ar tiekimo grandinės srityje.

Pateikiamas dinaminis modelio įgyvendinimas.



Šaltinis: sudaryta autoriaus

6 pav. Dinaminis modelio įgyvendinimas

Šis koncepcinis modelis parodo informacijos transformavimą į B2B kompanijos pridėtinės vertės – ryšiais su klientais, kūrimą.

Šioje darbo dalyje apžvelgėme, kaip informacijos analizė per ryšių su klientais valdymo sistemas gali įtakoti B2B įmonės pelningumą bei konkurencingumą. Remdamiesi šia informacija toliau bus rengiama eksperimentinio tyrimo metodika, kurios rezultatas – duomenų gavybos metodais apdorota informacija, bei jos palyginimas su lauktais duomenimis.

4. DUOMENŲ GAVYBOS METODŲ TAIKYMO B2B APLINKOJE EKSPERIMENTINIS TYRIMAS

Pagal CRM filosofiją verslo pagrindas yra klientas, su jo poreikiais derinama veikla bei kultūra, reikalinga efektyviai rinkodarai, pardavimams ir paslaugų teikimui. CRM prasideda nuo verslo strategijos, kuri daro įtaką pokyčiams organizacijoje bei jos veikloje ir yra susijusi su informacijos technologija. Šioje darbo dalyje bus aprašoma CRM duomenų gavybos dalies, naudojamos tyrime atlikti, technologiją, priemones ir duomenis.

Tam, kad efektyviau valdyti prekių kiekius, verslo procesus ir vartotojų aptarnavimą, reiktų panaudoti priemones, galinčias analizuoti skirtingo tipo duomenis. Tyrime bus naudojami duomenų gavybos metodai. Taip pat tyrimas vyks trijų B2B verslo aplinkos tiekimo grandinės sričių kryptimis siekiant kurti ryšius su klientais bei tokiu būdu padidinti B2B kompanijos rinkos vertę.

Tyrimo objektas – duomenų apdorojimas trimis B2B verslo operacijų kriterijais panaudojant duomenų gavybos metodus siekiant panaudoti gautus rezultatus verslo strategijai kurti, kuri orientuota į ryšių su klientais kūrimą.

Ryšius su klientais valdymo sistemose saugomi duomenys gali skirtis ypač skirtingo profilio kompanijose. Medikamentais prekiaujančioje kompanijoje CRM duomenų gavybos metodo įvesties duomenys, gali būti ne tik pardavimai, bet ir šalies specifikos – sergamumo sezonų ir pan. informacija. Standartinėmis priemonėmis gali nepavykti prognozuoti teisingus rezultatus. Todėl tolimesnėje tyrimo eigoje naudosime specifinės verslo srities informaciją bei analizuosime, kuriuo metodu rezultatai labiausiai atitinka tikrovę.

B2B verslo operacijų kriterijai, kuriais remiantis bus atliekamas tyrimas:

1. Pirkimų kreditavimas, nustatomas klientams;
2. Prekių kiekiai;
3. Tvarkaraščių, gabenimo laiko valdymas.

Iš ankstesniame skyriuje analizuotos literatūros bei mokslinių tyrimų, pastebėta, jog dažnai nėra aišku, kokį duomenų kiekį reikia pateikti kaip įvesties informaciją norint gauti arčiausiai prognozuojamų duomenų rezultato. Ypač duomenų kiekis aktualus B2B aplinkos kompanijoms, kurios disponuoja dideliais duomenų kiekiais. Atliekant duomenų apdorojimą duomenų gavybos metodais su nepakankamu kiekiu duomenų bei naudojantis daugiau nei vienu duomenų gavybos metodu didėja rizika gauti netikslius rezultatus. Ir atvirkščiai,

atliekant duomenų apdorojimą su milijono eilučių informacija, ženkliai padidėja apdorojimo laikas bei serverių apkrova, kuri dažnai realizuojama pardavimams, kasdieninėm ataskaitom bei testavimam.

Tyrimo dalyje aprašomi įvesties duomenys, jų specifika bei apdorojimo būdai. Tyrimas atliekamas, pasirinktomis programinėmis priemonėmis, aprašoma tyrimo eiga bei rezultatai. Tikrinamas tyrimo rezultatų panaudojimas ir teisingumas. Galiausiai tyrimo dalies pabaigoje sisteminami tyrimo rezultatai, atliekamos interpretacijos, teikiami siūlymai bei prognozės.

4.1. ĮVESTIES DUOMENŲ APRAŠYMAS

Pasirinkta B2B kompanija, kuri bendradarbiauja su 74-iais tarptautiniais periferinės, programinės kompiuterinės ir buities įrangos gamintojais. Kompanijos klientai – tarpininkai, mažmenininkai, kiti stambaus, vidutinio ir smulkaus verslo rinkos dalyviai. Tai tinklinė organizacija, kurios padaliniai išsidėstę Suomijoje, Švedijoje, Norvegijoje, Estijoje, Latvijoje, Lietuvoje, Lenkijoje, Šveicarijoje bei Vokietijoje. Tai sąlygoja duomenų srautus bei informacijos kompleksškumą. Iki šiol kompanijoje ryšiai su klientais valdymas vyksta dvejais lygiais:

1. Šalies lygiu - rinkodaros bei marketingo veiksmai taikomi konkrečiai šaliai, todėl duomenys analizuojami pagal kiekvienos šalies pardavimo rodiklius.

2. Kompanijos lygiu – vadybiniai sprendimai priimami atsižvelgiant į bendras pardavimų tendencijas, glaudžiai bendradarbiaujant su tiekėjais, analizuojant jų gamybos apimtį, prekių poreikį, klientų judėjimo srautus.

Kompanijos duomenys yra saugomi duomenų sandėliuose ir didelis jų kiekis sąlygoja duomenų panaudojimo, vadybiniam operaciniams sprendimams atlikti, sudėtingumą. Organizacijoje sprendimai atliekami pasitelkiant vadybininkus, atsakingus už tam tikrų prekių ar klientų grupę. Informacija vadybininkams pateikiama naudojant ataskaitas paruoštas SAP Crystal Report įrankiu. Atskiras informacinių technologijų ataskaitų padalinys priima užsakymus iš vadybininkų specifinėms ataskaitoms, todėl informacijos analizės priemonės naudojamos organizacijoje yra nevienodos, jautrios darbuotojų kaitai bei neuniversalios.

Kompanijoje naudojama ERP sistema, kurioje atliekamos šios operacijos:

1. fiksuojama pardavimo transakcija per XML, WEB arba atsakingą vadybininką, kuris suveda užsakymo informaciją.

2. Įvedami, keičiami duomenys apie klientą, susijusias kompanijas, galutinius vartotojus.

3. Atskirame modulyje atliekamos operacijos su prekėmis bei prekių kainomis. Prekių ir paslaugų kainoms didžiausią įtaką daro gamintojai, tačiau kompanijos vadybininkai sprendžia apie nuolaidas atskiroms prekių grupėms ar klientams, taip pat gali atlikti prekių grupavimą. Kainos atnaujinamos kartą per savaitę.

4. Įvedama ir redaguojama informacija apie vežėjus, pristatymų laikus bei įkainius.

Šio pobūdžio informacija saugoma mažiausiai vienerius metus.

Duomenys naudojami visuose IT sprendimuose. Duomenys yra valdymo ir prekybos operacijų pagrindas. Tačiau valdyti duomenis yra sunku dėl įvairių priežasčių.

- Duomenų kiekis auga eksponentiškai.
- Senus duomenis reikia saugoti, nuolat atsiranda nauji. Sunku rasti kiekvienam sprendimui reikalingus duomenis.
- Duomenys yra išbarstyti įmonėje, kaupiami skirtingais metodais ir prietaisais.
- Saugomi keliuose skirtinguose, nutolusiuose serveriuose ar apdorojimo sistemose.
- Naudojami skirtingi formatai ir kalbos (žmonių taip pat).

Renkant duomenis buvo susidurta su dideliu duomenų kiekiu – pusės metų kompanijos transakcijose užfiksuota milijonas MS Excel eilučių. Kiekviena pirkimo operacija pateikta dešimčia ir daugiau rodiklių. Dėl informacijos kiekio išauga duomenų apdorojimo laikas, todėl svarbu suskirstyti duomenis etapais ar dalimis, kurios bus apdorojamos pasirinktom priemonėm, tačiau nepraras galutinės informacijos teisingumo nepriklausomai nuo duomenų kiekio. Tyrime naudojami duomenys pateikti 3 lentelėje.

Įvesties laukų aprašymas

| Lauko pavadinimas | Lauko tipas |
|-------------------------------------|-------------|
| Kliento kodas | String |
| Suteiktas kreditas | Decimal |
| Prekės kodas | String |
| Vieneto kaina be PVM | Decimal |
| Pirkimo suma | Decimal |
| Kiekis | Integer |
| Pristatymo data | Date |
| Pageidaujama pristatymo data | Date |

Šaltinis: sudaryta autoriaus.

Duomenys naudoti eksperimentiniui tyrimui atlikti yra pirminiai. Tyrimo eigoje, atliekant eksperimentą minėtais trimis aspektais pridedami papildomi ar išvestiniai duomenys susiję su verslas verslui aplinkos specifika – generuojami iš elektroninės komercijos vartotojo sąsajos ir renkami ERP sistemoje, fiksuojančioje duomenis susijusius su tiekimo grandine.

4.2.ĮVESTIES DUOMENŲ APDOROJIMAS

Duomenų apdorojimo būdų pasirinkimas yra svarbus eksperimentinio tyrimo etapas, kadangi kiekvieno programinio įrankio specifika skiriasi, nors ir atliekamos tos pačios funkcijos bei duomenų gavybos metodo algoritmai.

Programinio įrankio pasirinkimo pagrindimas. Programinės priemonės apžvelgtos 4 priede. Ši programinė įranga naudojama verslo duomenų analizei, fiksuoja verslo transakcijas, generuoja ataskaitas ir t.t. Priede pateikta ryšių su klientais valdymo sistemos vieta bendroje informacinių sistemų rinkoje. Klientų valdymo sistema apima verslo valdymo sprendimų informaciją tiesiogiai susijusią su vadybiniais ir marketinginiais sprendimais. Atsižvelgiant į šią IS toliau pasirinksime programinį įrankį, kuriuo taikysim pasirinktiems B2B tiekimo grandinės aspektam duomenų gavybos metodus.

Analizei pasirinkta 11 programinių įrankių, kuriuose pagrindinis ar šalutinis modulis atlieka analize su vieno ar daugiau duomenų gavybos metodu. Palyginimas pateiktas 4 priede 4 lentelėje aprašant IS paskirtį, funkcionalumą bei išskiriant integruotus duomenų gavybos metodus. Iš lentelėje pateiktų programinės įrangos aprašymų galima teigti, kad aplikacijos, kurios apdoroja informaciją duomenų gavybos principu naudoja tam tikrus duomenų gavybos metodus, kurių pritaikymo ir naudojimo galimybės priklauso nuo IS, kuriuose tas modelis integruotas, savybių ir paskirties. Panaudojus šią informaciją, pasirinksime programinę įrangą, kurią vėliau apdorosime įvesties duomenis tyrimo eigoje.

Atlikus programinės įrangos analizę pasirinkta duomenų apdorojimo priemonė StatSoft STATISTICA 7, kadangi atitinka 3 lentelėje apibrėžtas savybes.

4 lentelė

STATISTICA 7 aprašymas

| Savybė | Aprašymas |
|--|---|
| Reikalavimai periferinei ir operatyvinei aplinkai | Minimalūs: Operacinė sistema: Windows 2000 ar vėlesnė. RAM: 256 MB. Procesoriaus greitis: 500 MHz. Rekomenduojami: Operacinė sistema: Windows XP ar vėlesnė. RAM: 512 MB. Procesoriaus greitis: 2.0 GHz. |
| Pritaikomumas | Java, Microsoft Excel, C#, VB.NET, taip pat kodas atviras trečios šalies procedūroms. WEB integracijos sąsaja. |
| Naudojamas modulis | Data Miner, specializuoti ir bendrieji duomenų gavybos moduliai |
| Aktualūs duomenų gavybos metodai | CHAID medžiai, sprendimų medžiai, regresiniai medžiai, klasterizavimas, neuroniniai tinklai, arčiausio K- elemento metodas, asociatyvios taisyklės. |
| Atvaizdavimo būdai | Grafikai, ataskaitos, duomenų grupės, išvestinės lentelės. |
| Prieinamumas | Naudojimas universitete |

Šaltinis: sudaryta autoriaus pagal STAT SOFT Data miner brochure, (2006).

Duomenys bus apdorojami skirtingais pjūviais kiekvienam tyrimo etapui. Prieš atliekant veiksmus su STATISTICA 7 duomenis reikia paruošti tyrimui, tai atlikta naudojantis MS Excel 2007 pagalba. Kadangi užtrunka apdoroti 1 milijoną eilučių atliekant grupavimo, rūšiavimo ir kitus veiksmus, buvo nuspręsta pakoreguoti įvesties duomenų imtį:

- Duomenų imtis buvo sutrumpinta iki apytikriai 150 tūkst. pardavimo transakcijų.
- Taip pat duomenys suskirstyti pagal laikotarpį – 2008 metų sausio, vasario ir kovo mėnesius.
- Analizei bus naudojami vasario mėnesio duomenys, nes sausio mėnesio pardavimų įtakai gali turėti naujametinės šventės. Tam, kad išvengti akivaizdžių duomenų iškraipymo pasirinktas vasario mėnesis.
- Atlikto tyrimo rezultatai bus lyginami su ateinančio, t.y. kovo mėnesio pardavimais. Atsižvelgiant į šį duomenų lyginimą bus sprendžiama apie atlikto tyrimo duomenų teisingumą, naudotų metodų pritaikomumą.

- Vasario mėnesio pardavimų duomenys bus suskirstomi į tris duomenų masyvus atsižvelgiant į tyrimo srities aspektus: pirkimų kreditavimą, nustatomą klientams, prekių kiekius bei tvarkaraščių, gabenimo laiko valdymo duomenų masyvą.
- Kiekvienam iš tyrimo srities aspektų duomenų masyvui bus priskiriamos reikšmės darančios įtaką reikšminiams tos srities duomenims.

Atliekant tyrimą apibrėžtais aspektais ir naudojant duomenų gavybos metodus tikimasi užtikrinti dviejų tipų galimybes:

1. Automatinę tendencijų ir elgsenos prognozę. Panaudojant anksčiau sukauptus duomenis atliekamos prognozės numatančio pristatymo laikų pasiskirstymą.
2. Automatinį anksčiau nežinomų ar paslėptų priklausomybių ir savybių aptikimą. Pirkimų kreditavimui nustatymui naudojamų dedamųjų aptikimas. Tokiu būdu, galima aptikti įtaką rezultatui daromus nežinomuosius bei klaidas, atsiradusias duomenų įvedimo metu.

Tikimasi, jog antrasis būdas padės atlikti pirkimo kreditavimo klientams kategorizavimą. Duomenų gavybos metodų panaudojimo galimybių pirmasis variantas tikėtina susies prekių pristatymą laike bei vaizduos prekių kiekių automatines tendencijas.

4.3.PIRKIMŲ KREDITINGUMO, NUSTATOMAS KLIENTAMS, TYRIMO EIGA

Atliekant tyrimą klientų kreditingumo srityje svarbu apibrėžti, kokios sudedamosios įtakoja suteikiamą kreditą kompanijos atliekamiems pirkimams. Nagrinėjamos kompanijos kreditų suteikimo valdymas paskiriamas finansinės ataskaitybos darbuotojui, kuris stebi kliento atsiskaitymus, sąskaitų apmokėjimo laiką. Kaip ir daugelyje kompanijų, šioje kompanijoje kreditingumas nustatomas remiantis finansinių įsipareigojimų valdymu. Dažniausiai kompanijų atstovai pasiūlo suteikti tam tikrą kredito dydį, o B2B kompanijos partnerės priima, atmeta arba pasiūlo kitokį sprendimą. B2B kompanijos nustato, koks gali būti minimalus ir maksimalus kredito dydis. Vėliau, sekant kliento finansines galimybes, suteiktas kreditas gali būti mažinamas arba didinamas. Kredito suteikimas pirkimams yra vienas B2B kompanijų bruožų. B2C, jei ir suteikiami kreditai, dažnai jie būna standartiniai ir nekeičiami atsižvelgiant į klientą. B2B aplinkos kompanijoms kreditų teikimas - tai vienas iš kompanijos pridėtinės vertės kūrimo aspektų.

Kadangi skirtingo laikotarpio duomenys leidžia naudoti tam pasirinktą metodiką ir ją vėliau patikrinti kito laikotarpio duomenimis, tyrimas bus atliekamas duomenų gavybos metodu – kategorizavimu.

Tyrimą galima suskirstyti etapais:

- (1) Nustatomos iš anksto žinomos grupės, kategorijos.
- (2) Priskyrimas kategorijoms vyksta per apmokymą su klasifikuotais duomenimis.
- (3) Sudaromas klasifikavimo modelis.
- (4) Modelis panaudojamas naujiems duomenims klasifikuoti.

Naudojami duomenys, duomenų tipai ir papildoma informacija pateikta lentelėje.

5 lentelė

Įvesties laukų aprašymas pirkimo kreditavimo tyrimui

| Lauko pavadinimas | Lauko tipas | Papildoma informacija |
|---------------------------------|-------------------------|---|
| Kliento kodas | String | Klientas identifikuojamas vietiniu identifikatoriumi |
| Krepšelių suma | Decimal | Įsigyta prekių už šią sumą per vasario mėnesį. |
| Vidutinė krepšelių vertė | Decimal | Vidutinė krepšelio vertė per vasario mėnesį. |
| Įsigyta prekių vienetų | Integer | Įsigyta prekių, vienetais per vasario mėnesį. |
| Apsilankymų skaičius | Integer | Apsilankymų skaičius per vasario mėnesį. |
| Suteiktas kredito dydis | Integer; categorical | Suteiktas mėnesinis kredito dydis pirkimams vasario mėnesiui. |

Šaltinis: sudaryta autoriaus

Duomenys suvesti i Statistica 7 darbo langą. Vasario mėnesio duomenų masyvas paruoštas pirkimų kreditavimui, jo fragmentas pateiktas žemiau lentelėje 5.

Iš visų suvestų duomenų pasirenkame du įtakingiausius įvesties parametrus:

1. Apsilankymų skaičius. Rodiklis, nurodantis, kiek kartų kompanija yra atlikus pirkimo transakciją. Tai vienas iš įtakojančių kreditavimą veiksnių. Tarp apsilankymo skaičiaus ir kreditavimo sumos tiesioginės priklausomybės nėra, kadangi galima atnešti kompanijai pelno atlikus mažiau pirkimo operacijų, perkant brangesnes prekes. Tam, kad kreditavimo suma padengtų ir šią priklausomybę, reikalinga įvesti vidutinę krepšelio sumą.
2. Vidutinė krepšelio suma. Šis duomuo rodo vidutines pajamas gautas iš vieno kliento per imties laikotarpį – vasario mėnesį.

Iš pateikto fragmento matosi, kokio tipo naudojami duomenys. Kreditavimo sumų kategorijos pasirinktoje kompanijoje yra septynios -

1000 000, 500 000, 200 000, 100 000, 50 000, 20 000 ir 10 000.

6 lentelė

Ivesties laukai pirkimo kreditavimo tyrimui

| 1 KlientoID | 2 KrepseliuSuma | 3 VidutineKrepSuma | 4 IsigytaVnt | 5 ApsilankymuSk | 6 Kreditas |
|----------------|--------------------|-----------------------|-----------------|--------------------|---------------|
| 2 | 195205.09 | 863.74 | 808 | 226 | 200000 |
| 12 | 4178.25 | 522.28 | 11 | 8 | 50000 |
| 15 | 30372.27 | 241.05 | 385 | 126 | 100000 |
| 23 | 22395.66 | 217.43 | 197 | 103 | 50000 |
| 30 | 1641.40 | 234.49 | 54 | 7 | 50000 |
| 62 | 35191.59 | 224.15 | 557 | 157 | 100000 |
| 72 | 15372.77 | 236.50 | 240 | 65 | 50000 |
| 73 | 33558.48 | 252.32 | 307 | 133 | 100000 |
| 85 | 1813.31 | 302.22 | 7 | 6 | 50000 |
| 91 | 17542.87 | 233.90 | 288 | 75 | 50000 |
| 92 | 18633.43 | 107.71 | 389 | 173 | 50000 |
| 98 | 3222.92 | 460.42 | 8 | 7 | 50000 |
| 100 | 7290.50 | 428.85 | 34 | 17 | 50000 |
| 101 | 26132.45 | 706.28 | 201 | 37 | 100000 |
| 132 | 2186.51 | 91.10 | 57 | 24 | 50000 |
| 144 | 20283.48 | 215.78 | 188 | 94 | 50000 |
| 148 | 12539.17 | 184.40 | 126 | 68 | 50000 |
| 156 | 5304.38 | 171.11 | 174 | 31 | 50000 |
| 165 | 6459.00 | 922.71 | 206 | 7 | 50000 |
| 167 | 3477.63 | 231.84 | 47 | 15 | 50000 |
| 170 | 5283.73 | 73.39 | 193 | 72 | 50000 |
| 197 | 15270.38 | 803.70 | 116 | 19 | 50000 |
| 206 | 10061.48 | 359.34 | 92 | 28 | 50000 |
| 212 | 3044.50 | 202.97 | 84 | 15 | 50000 |
| 237 | 2747.14 | 101.75 | 58 | 27 | 50000 |
| 242 | 96395.46 | 382.52 | 836 | 252 | 100000 |
| 251 | 28280.64 | 155.39 | 845 | 182 | 100000 |

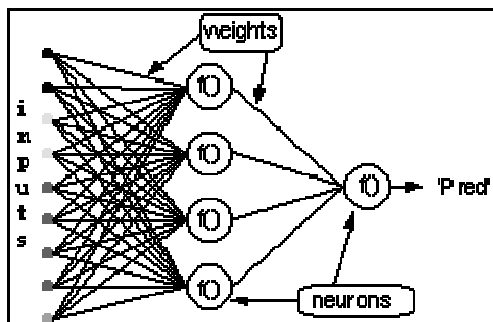
Šaltinis: sudaryta autoriaus

Šiam tyrimui naudojamas duomenų gavybos klasifikavimo metodas paremtas neuroninių tinklų principu.

STATISTICA 7 programos modulio pasirinkimo etapai: Statistics – Data Mining – Neural networks. Pasirinktas problemos tipas – Classification. Suvedami priklausomi kintamieji, kategorizavimo etapai, nurodomos papildomos vaizdavimo funkcijos.

Pagal tai, kokie yra algoritmo įvesties parametrai, klasifikavimo pradžioje yra priskiriami lyginamieji svoriai. Pirkimo kreditavimo tyrimo atveju, įvesties duomenys apima klientų apsilankymų skaičių B2B kompanijoje bei vidutinę krepšelių sumą per vasario mėnesį.

Priskyrimas kategorijoms vyksta per apmokymą su klasifikuotais duomenimis. Klasifikavimo algoritmą sudaro svorių suteikimas įvestoms reikšmėms. Suteikti svoriai perskirstomi taip, kad E reikšmė (iš formulės $Tinklo\ klaida = Prognozuojamas\ k - realus\ j = E$) būtų mažiausia.



Šaltinis: TIBERIUS Data Mining (2007), Back Propagation Weight Update Rule

7 pav. Teorinė tinklo svorių sudarymo iliustracija

Tam, kad minimizuoti klaidų aibę, nustatomas kiekvienas svorio jautrumas rezultatui. Suteikto svorio jautrumas apibrėžiamas keičiant pačius svorius bei fiksuojant pasikeitimo įtaką rodikliui E.

Tyrimo atveju įvesties duomenys yra dvejopi, todėl kiekvieno duomenų masyvo įtaka nustatytiems svoriams išreikšta rangavimu, pateiktu 6 lentelėje.

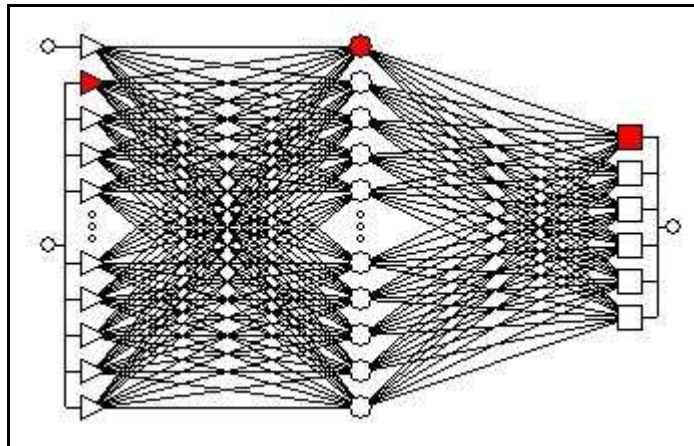
7 lentelė

Svorių rangavimas

| | Sensitivity Analysis - 2 (PagalKlienta in | |
|---------|---|---------------|
| | VidutineKrepSuma | ApsilankymuSk |
| Ratio.2 | 1.003824 | 1.015198 |
| Rank.2 | 2.000000 | 1.000000 |

Šaltinis: sudaryta autoriaus

Pasiskirsčius svoriams, apsilankymų skaičius daro didesnę įtaką galutiniam rezultatui – kreditavimo klasifikacijai. Vidutinė krepšelio sumos svertinė įtaka klasifikacijai lygi tik 1.003824. Toks įvesties svorių pasiskirstymas atvaizduojamas tinkle iliustracijoje, paveiksle 8.



Šaltinis: sudaryta autoriaus.

8 pav. Praktinė tinklo svorių sudarymo iliustracija

Neuroninio tinklas vizualiai iliustruoja neuroninius ryšius tarp kintamųjų ir jų svorių. Tinklo baigties taškai yra šeši, tai reiškia, kad kreditavimo sumos suskirstytos ne į septynias (kaip buvo pradžioje tyrimo), bet į šešias kategorijas.

Sudarytas kategorijų pasiskirstymo modelis – pateiktas paveikslas 7, stulpelis Kreditas.1. Pagal svertinius įverčius perskirstytos kredito sumos. Kliento identifikaciją atitinka 7 paveikslo pirmos stulpelio numeris. Tad, galutinis pritaikyto metodo rezultatas – duomenų masyvas su klientams išskirstytais kredito dydžiais.

| | Prediction (1) (Pagalklienta in Vasaris) | |
|----|---|------------|
| | Kreditas | Kreditas.1 |
| 1 | 200000 | 200000 |
| 2 | 50000 | 10000 |
| 3 | 100000 | 50000 |
| 4 | 50000 | 50000 |
| 5 | 50000 | 50000 |
| 6 | 100000 | 100000 |
| 7 | 50000 | 50000 |
| 8 | 100000 | 100000 |
| 9 | 50000 | 10000 |
| 10 | 50000 | 50000 |
| 11 | 50000 | 100000 |
| 12 | 50000 | 50000 |
| 13 | 50000 | 50000 |
| 14 | 100000 | 50000 |
| 15 | 50000 | 50000 |

Šaltinis: sudaryta autoriaus.

9 pav. Duomenų apdorojimo rezultato langas

Iš prognozuojamų kreditavimo kategorijų yra eliminuota 20 000 kredito suma. Priklausomai nuo vidutinės krepšelio vertės bei apsilankymo skaičiaus per vasario mėnesį priskirti atitinkami kredito dydžiai, kurie gali būti taikomi duomenų gavybos modulyje

integruotame ryšių su klientais sistemoje B2B verslo aplinkoje. Šiuos rezultatus iliustruoja 8 paveikslas.

8 lentelė

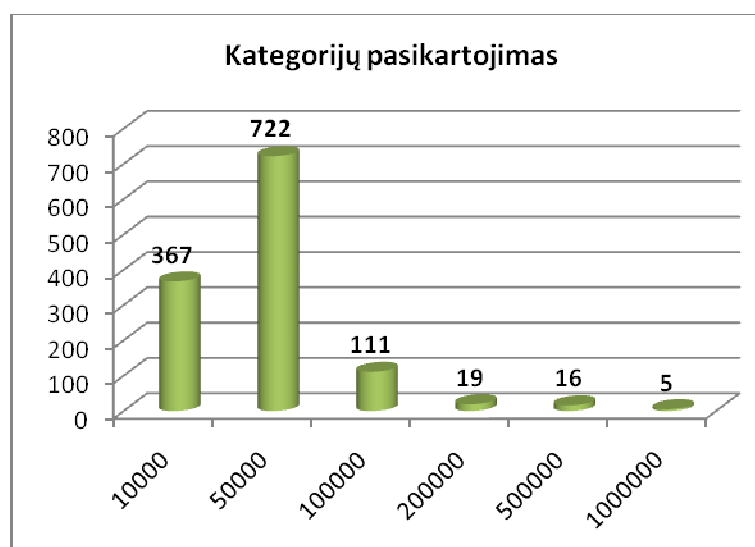
Prognozuojamo modelio duomenų palyginimas

| | Confusion Matrix - Kreditas(2) (Pagalklienta in Klasifikacija) | | | | | |
|-----------|--|-------|--------|-------|---------|--------|
| | 20000 | 50000 | 100000 | 10000 | 1000000 | 500000 |
| 200000.2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50000.2 | 16 | 662 | 107 | 281 | 5 | 15 |
| 100000.2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10000.2 | 3 | 60 | 4 | 86 | 0 | 1 |
| 1000000.2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 500000.2 | 0 | 0 | 0 | 0 | 0 | 0 |

Šaltinis: sudaryta autoriaus.

Klientų pirkimo kreditavimo kategorijos, kurių vietos pasikeitė po duomenų gavybos metodo panaudojimo aprašomos lentelėje. Daugiausiai keičiama pozicija buvo 50 tūkst. kredito suma. Tačiau mažiausiai kitusi pozicijos yra milijono kredito suma. Kitimų išsidėstymų matrica parodo, kurios kategorijos yra jautriausios svorinių koeficientų pasikeitimams. Tai rodo, kad 50 tūkst. kredito suma prognozuojamame modelyje bus suteikiama 662 skirtingom pozicijom – perskirstymų skaičius palyginus su įvesties duomenimis.

Apdorojant duomenų masyvą klasifikavimo metodu remiantis neuroniniais tinklais pastebėta, jog kategorijų priskyrimo skaičius nepakito lyginant su įvesties duomenimis. Sekančiame paveiksle išsiskiria 50 tūkst. kredito suma, kuri prieš ir po apdorojimo suteikta 722 klientams.



Šaltinis: sudaryta autoriaus.

10 pav. Kreditavimo kategorijų pasiskirstymo dažnis

Taikant šį metodą nesikeičiantis pasikartojimų skaičius leidžia kompanijai nenukrypti nuo užsibrėžtų finansinių tikslų. Kredito sumos yra perskirstomos atsižvelgiant į klientų apsilankymo dažnio bei vidutinės krepšelio sumos svertinius koeficientus nedidinant ar nemažinant bendro kreditavimo plano.

Atlikus pirkimams skirtu kreditu teikimo valdymo galimybių tyrimą paaiškėjo, jog duomenų gavybos klasifikavimo metodas yra tinkamas šiam uždaviniui spręsti. Gauti rezultatai parodė, jog 0,89 proc. didesnę įtaką kredito kategorijos priskyrimui daro apsilankymų skaičius nei vidutinė krepšelio suma. Pagal rezultatus, kurių tikėtasi, panaudojus duomenų gavybos metodą, aptiktas ir pagrįstas parametras, kuris įtakoja prikimo kredito sumą – klientų apsilankymo kompanijoje arba pirkimo transakcijų skaičius.

Ryšiu su klientais valdymo sistemose įdiegtas klasifikavimo naudojantis neuroniniais tinklais duomenų gavybos metodas padėtų tiksliau atlikti klientų segmentavimą pagal suteikiamus kreditus, o tai leistų adaptuoti paslaugas klientų poreikiams, kuris sąlygotų didesnę klientų pasitenkinimą. Atskiroms, pagal kredito kategorijas, priskirtoms klientų grupėms galima būtų taikyti skirtingas kainodaros strategijas. Šie veiksmai skatina ryšius su klientais, o tai padeda geriau suprasti konkurencinę situaciją rinkoje, kuria B2B kompanijos pridėtinę vertę, skatina inovacijas.

4.4. PREKIŲ KIEKIŲ PIRKIMAMS TYRIMO EIGA

Tiekimo grandinės valdymas glaudžiai susijęs su pardavimų planavimu. Prognozuojant, kiek bus parduota tam tikrų prekių, nustatoma, kiek jų reikia užsakyti iš gamintojų, o gamintojams jas pagaminti.

Klientų valdymo sistemos požiūriu prekių kiekis matomas kaip pardavimo transakcijos bei kliento poreikio pirkti numatymas.

Dabartinėje B2B aplinkoje prekių kiekį įtakoja daugiau faktorių – tai gamintojų pardavimų diktavimas, užbaigiant tam tikrą prekių liniją, papildomas tų prekių pardavimų skatinimas galutiniams vartotojams.

Taip pat prekių kiekio valdymą įtakoja sandėlio prekių likučiai, gabenimo ypatybės. Visi šie faktoriai įtakoja marketingo sprendimus, atliekamus ryšiu su klientais valdymo sistemose.

Ryšiu su klientais valdymo sistemose kaupiama informacija gali būti panaudota sprendžiant tiekimo grandinės uždavinius. Toks sistemos integralumas leistų pritaikyti

duomenis, kurie nebūtinai tiesiogiai priklauso pirkimo prekių kiekio planavimui, tačiau gali būti būdas papildantis pardavimų prognozavimą bei padedantis už pirkimus atsakingiems vadybininkams nuspręsti, kiek ir kokių prekių reiktų įsigyti sekantį periodą.

Šiam uždaviniui įgyvendinti siūloma atlikti tyrimą, remiantis realiais duomenimis bei pasirinktu duomenų gavybos metodu. Gautos prognozės rezultatus palyginę su ateinančio laikotarpio duomenimis, nuspręstumėm, ar tikslinga spręsti prekių kiekio uždavinius tiekimo grandinėje pasiūlytais metodais.

Tyrimo eiga:

- (1) Surikiuoti duomenis pagal prekių kodą, pateikiant papildomą suminę informaciją;
- (2) Pasirinkti duomenų gavybos metodą, galintį atlikti prognozę naudojant nepriklausomus kintamuosius bei iš anksto nežinomus ryšius tarp kintamųjų;
- (3) Pagrįsti modelio pasirinkimo aktualumą keliamam uždaviniui;
- (4) Atlikti tyrimą su eksperimentiniais duomenimis – gauti prognozuojamų reikšmių sąrašą;
- (5) Empirinis rezultatų patikrinimas. Palyginti gautą prognozę su sekančio mėnesio praduotų prekių kiekiu.
- (6) Pateikti išvadas.

Eksperimento metu tikimasi gauti šiuos rezultatus:

1. Prekių kiekiui įtakos turi bent du loginės sąsajos neturintys kintamieji.
2. Atlikus prekių kiekio prognozę sekančiam laikotarpiui bei empiriškai patikrinus rezultatus priimsime išvadą, jog pasirinktas duomenų gavybos modelis gali būti rekomenduojamas taikyti tiekimo grandinės uždaviniams spręsti.

Lentelėje pateikti kintamieji, kurie bus naudojami tyrimui atlikti. Siekiant išvengti didžiųjų žiemos švenčių įtakos, tyrimui naudojamas vasario mėnesio duomenų masyvas.

9 lentelė

Įvesties laukų aprašymas prekių kiekių pirkimams tyrimui

| Lauko pavadinimas | Lauko tipas | Papildoma informacija |
|-----------------------|-------------|--|
| Produkto_kodas | String | Produktas identifikuojamas unikaliu kodu |
| Kiekis | Integer | Suma prekės vienetų, įsigytų už laikotarpį |
| Parduota | Decimal | Parduota prekių už tam tikrą sumą tam laikotarpiui |
| Vid_kaina | Decimal | Vidutinė prekės kaina. |

9 lentelės tęsinys

| Lauko pavadinimas | Lauko tipas | Papildoma informacija |
|--------------------------|-------------|---|
| Min_kaina | Decimal | Minimali prekės kaina. |
| Max_kaina | Decimal | Maksimali prekės kaina. |
| Klientų_sk | Integer | Klientų kiekis, įsigijusių tą prekę per laikotarpį |
| Sand_kiekis | Integer | Kiekis prekių sandėlyje sumuojant kartu su atvykstančiomis . |
| Web_click | Decimal | Vidutinis „prekės aktyvumas“ * per dieną už tam tikrą laikotarpį |
| Susijusios_kiekis | Integer | Susijusių prekių kiekis (Pvz. fotoaparato susijusios prekės yra dėklas, atminties korta ir t.t.) skaičiuojant pagal prekės kodą, ne pagal prekės grupę. |
| Specifikacijos | Integer | Produkto specifikacijos pilnumas procentais. Visos privalomos savybės bei 50 proc. neprivalomų produkto savybių. |

Šaltinis: sudaryta autoriaus

*- prekės pasirinkimas per internetinę vartotojo sąsają

Lentelėje pateiktos ne visos reikšmės akivaizdžiai yra susijusios su prognozuojamu kiekiu. Todėl tolimesnėje tyrimo eigoje bus analizuojama kaip priklausomi bei nepriklausomi kintamieji įtakos prekių kiekį.

Išanalizavus pardavimo operacijas, nustatyta, jog per vasario mėnesį buvo parduota 14093 produktų. Duomenys buvo grupuojami pagal produkto kodą, skaičiuojant įsigytą kiekį, sumuojant kainą bei pateikiant bendrines produkto savybes. Eksperimentinių duomenų imtis nekoreguojama, kadangi tokį duomenų kiekį galima apdoroti STATISTICA 7.

Suvestų duomenų fragmentas pateiktas sekančioje lentelėje.

10 lentelė

Įvesties duomenų fragmentas

| 1 Produkto_kodas | 2 Kiekis | 3 Parduota | 4 Vid_kaina | 5 Min_kaina | 6 Max_kaina | 7 Klientų_sk | 8 Sand_kiekis | 9 Web_click | 10 Susijusios_Kiekis | 11 Specifikacijos |
|---------------------|-------------|---------------|----------------|----------------|----------------|-----------------|------------------|----------------|-------------------------|----------------------|
| 152 | 14 | 4878.88 | 419.46 | 277.52 | 505.24 | 7 | 15 | 6.78927485 | 42 | 72% |
| 42157 | 15 | 335.64 | 22.15 | 0.00 | 26.50 | 10 | 31 | 0.881911975 | 53 | 64% |
| 42160 | 21 | 524.05 | 24.77 | 19.75 | 26.80 | 10 | 40 | 1.65932471 | 87 | 71% |
| 42161 | 1286 | 24498.40 | 21.91 | 0.00 | 26.92 | 121 | 160 | 2.00761402 | 100 | 95% |
| 42163 | 85 | 4041.08 | 48.52 | 39.26 | 53.20 | 33 | 115 | 5.39593082 | 73 | 54% |
| 42164 | 16 | 1135.03 | 72.16 | 58.69 | 79.70 | 13 | 33 | 9.57550214 | 30 | 60% |
| 42166 | 51 | 3553.43 | 70.26 | 60.82 | 82.40 | 11 | 13 | 7.93990031 | 83 | 65% |
| 42169 | 4 | 152.00 | 38.10 | 37.90 | 38.30 | 2 | 128 | 2.22358935 | 47 | 68% |
| 42333 | 6 | 161.80 | 25.93 | 21.80 | 28.00 | 3 | 249 | 8.10602276 | 5 | 10% |
| 44113 | 206 | 494.00 | 2.36 | 1.55 | 2.60 | 25 | 94 | 9.8682351 | 74 | 46% |
| 44214 | 20 | 82.00 | 4.10 | 4.10 | 4.10 | 1 | 90 | 7.02446464 | 46 | 83% |
| 47864 | 16 | 2179.87 | 132.21 | 110.98 | 150.40 | 6 | 7 | 3.86223272 | 31 | 98% |
| 65703 | 97 | 5696.66 | 63.00 | 51.89 | 70.30 | 31 | 43 | 2.18057728 | 28 | 75% |
| 65875 | 56 | 1078.73 | 19.36 | 17.62 | 22.50 | 16 | 229 | 6.26220725 | 30 | 19% |
| 65897 | 3 | 63.80 | 21.27 | 17.70 | 23.30 | 3 | 172 | 4.25732359 | 98 | 30% |
| 65927 | 26 | 1885.70 | 73.09 | 71.50 | 77.90 | 15 | 182 | 7.02197842 | 51 | 81% |
| 65998 | 16 | 1295.00 | 81.32 | 80.30 | 83.70 | 5 | 14 | 0.781204336 | 79 | 24% |
| 66064 | 48 | 458.40 | 9.57 | 9.50 | 9.90 | 9 | 248 | 3.03095444 | 79 | 53% |
| 66085 | 9 | 355.92 | 38.56 | 31.72 | 41.70 | 4 | 143 | 7.12418607 | 40 | 38% |
| 66240 | 55 | 1565.93 | 28.53 | 24.51 | 33.90 | 8 | 84 | 3.33989017 | 46 | 82% |

Šaltinis: sudaryta autoriaus.

Prognozuojamas kintamasis – Kiekis, toliau apdorojimo procese bus vadinamas priklausomu kintamuoju.

Susiję kintamieji – Parduota, Vid_kaina, Min_kaina, Max_kaina, Klientų_sk, Sand_kiekis, Web_click, Susijusios_kiekis, Specifikacijos. Šie kintamieji duomenų apdorojimo rezultatuose pateikiami kaip nepriklausomi kintamieji.

Duomenų apdorojimo modeliui bei prognozės algoritmui keliami šie uždaviniai:

1. Duomenų analizė turi būti atliekama išvengiant prielaidos apie tiesinę priklausomybę tarp kintamųjų. Prognozės algoritmas turi būti įgyvendinamas remiantis atskirais duomenų masyvais ir ryšiais tarp jų. Ši sąlyga užtikrintų, jog kintamieji nebus lyginami įprastais algoritmais, kadangi duomenys neatspindi prognozuojamo kintamojo loginės reikšmės (nėra tiesioginio ryšio tarp prekės kiekio sandėlyje ir specifikacijos pilnumo reikšmės);

2. Pasirinktas metodas užtikrintų duomenų vientisumą ir negeneruotų papildomos informacijos, skatinančios duomenų kompleksiskumo augimą. Priešingu atveju analizės procesas neatitiktų turimų techninių resursų.

3. Rezultatai, pritaikius pasirinktą metodą, turi būti suprantami ir palyginami. Rezultatas bus laikomas tinkamais, jei konkrečiai prekei bus gautas sekančio laikotarpio prognozuojamas kiekis. Metodo išvesties lange turi pateikiamas klaidos tikimybės rodiklis, statistiniai aprašomieji rodikliai.

Uždaviniui buvo atmesti netiesinės regresijos, regresijos medžio metodai kadangi rezultatai šiais metodais gaunami atliekant prielaidas apie ryšius tarp kintamųjų arba jų svertinių svorių. Duomenų masyvams taip pat tinkamas sprendimo medžio metodas.

11 lentelė

Metodų tyrimui pasirinkimo apibendrinimas

| | CHAID | Neuroniniai tinklai | Regresiniai medžiai | MARSplines |
|--|----------------------------|----------------------|----------------------------|----------------------|
| Prognozuoja | + | + | + | + |
| Priklausomiems kintamiems | + | +/- | + | + |
| Tiesinė regresija (ryšiai tarp kintamųjų) | + | +/- | + | - |
| Rezultatas | Prognozės kategorijų medis | Prognozė pareikšmiui | Prognozės kategorijų medis | Prognozė pareikšmiui |
| Grupinė analizė | + | - | - | - |

Šaltinis: sudaryta autoriaus remiantis STATISTICA 7

Pateiktoje lentelėje yra apibendrinti pagrindiniai duomenų analizės metodai uždaviniui spręsti pagal aktualius aspektus. Atmetami variantai, kuriuose prognozė pateikiama suskirstant duomenis į kategorijas. Dėl tinkamo rezultatų pateikimo ir algoritmo bazinėmis funkcijomis, uždaviniui spręsti pasirinktas MARSplines modelis.

Daugiavariantis prisitaikantis regresijos modelis (angl. *Multivariate Adaptive Regression Spline - MARSplines*) sprendžia problemas susijusias su reikšmių prognozavimu, remiantis nepriklausoma įvesties informacija.

MARSplines yra neparametrinė regresijos procedūra, kuri nedaro prielaidos apie netiesioginius funkcionalius ryšius tarp priklausomų ar nepriklausomų kintamųjų.

Regresijos tipo problemas spręsti yra įprasta apibrėžiant ryšius ar įtakas tarp kintamųjų, tokie uždaviniai remiasi parametrinėmis, tiesinėmis funkcijomis, siekiant atrasti pasikartojimų skaičių, pažingsniui besikeičiančių įtakos koeficientų.

- MARSplines modeliu galima tirti kokybinius ir kiekybinius kintamuosius;
- Šis duomenų gavybos modelis remiasi ryšiais ne tarp kintamųjų, bet tarp įtakos koeficientų grupių, vadinamų bazinėmis funkcijomis, kurios sudaromos iš analizės duomenų.
- MARSplines modelio lygtis yra ryšys tarp įvesties (X) ir išvesties (Y) kintamųjų remiantis bazinių funkcijų kombinacijomis (M).

$$y = f(X) = \beta_0 + \sum_{n=1}^M \beta_n \cdot h_n(X), \quad (1)$$

Čia h_n – bazinės funkcijos;

β_n – bazinių funkcijų svertiniai svoriai;

β_0 – pertraukiamas parametras.

Iš įvesties duomenų bazinių funkcijų, kurios patvirtina išvesties duomenis arba prognozę, modelis išrenka bazinių funkcijų svarių sumas.

Bazinės funkcijos yra apibrėžiamos kaip sutrumpintos funkcijos:

$$(x-t)_+ = \begin{cases} x-t & x > t \\ 0 & x \leq t \end{cases}$$

$$(t-x)_+ = \begin{cases} t-x & x < t \\ 0 & x \geq t \end{cases}$$

(2)

Čia, parametras t yra bazinės funkcijos mazgas (apibrėžiantis dalį tiesinės regresijos toje duomenų dalyje). Jie yra ieškomi visoje įvesties duomenų aibėje bei ryšiuose tarp duomenų. Atliekant šią mazgų paiešką bazinių funkcijų skaičius auga eksponentiškai, tačiau mažinama klaidų tikimybė. Sekantis modelio algoritmas atlieka kompleksškumo sumažinimą. Lieka tik tos bazinės funkcijos, kurios daro didžiausią arba reikšminę įtaką prognozuojamiems duomenis.

MARSplines:

1. Sąlygoja mažesnę resursų apkrovą – atrenka eksperimentinius duomenis, paliekant tik reikšmingiausius rezultatui (kompleksiškumo mažinimo algoritmas);
2. Įvesties duomenys gali būti nepriklausomi ir neturėti suprantamo ryšio su prognozuojamais duomenimis;
3. Modelis identifikuoja reikšmines iteracijas tarp duomenų;
4. MARSplines rezultatas – prognozės koeficientai kiekvienai priklausomai reikšmei.

Duomenų apdorojimui atlikti pasirenkam šiuos veiksmus STATISTICA 7 programos pagrindiniame lange: Statistics – Data Mining – MARSplines, atrenkame anksčiau apibrėžtus nepriklausomus ir vieną priklausomą kintamąjį bei pasirenkame modelio parametrus.

Parametrų pasirinkimas MARSplines modeliui reikalauja empirinio situacijos patikrinimo. Atlikus duomenų analizę gautas apibendrintas rezultatų langas, kuriame parodoma panaudoto modelio specifika.

MARSpline parametrai analizei

| Model specifications | Model Summary (Pa | |
|---------------------------|-------------------|--|
| | Value | |
| Independents | 9 | |
| Dependents | 1 | |
| Number of terms | 14 | |
| Number of basis functions | 13 | |
| Order of interactions | 1 | |
| Penalty | 2.000000 | |
| Threshold | 0.000500 | |
| GCV error | 19669.17 | |
| Prune | Yes | |

Šaltinis: sudaryta autoriaus.

Išanalizuoti 9 nepriklausomi kintamieji, analizės procese buvo naudojama 13 bazinių funkcijų, reiškiančių atskiras priklausomybes kiekvienam duomenų masyvui. Iteracijų skaičius parodo, kiek kartų duomenys buvo analizuoti nuo pradinio įvesties taško iki rezultatų pateikimo. Prieš atliekant duomenų analizę, tikslinga pasirinkti kuo daugiau bazinių funkcijų ir iteracijų. Didinant iteracijų skaičių siekiama padidinti prognozuojamo modelio tikslumą, o bazinių funkcijų gausa lemia didesnę ryšių tarp masyvų santykį. Esant optimaliam bazinių funkcijų ir iteracijų skaičiui, modelio apmokymo procesas pateikia mažiau abstraktų rezultatą.

Renkantis bazinių funkcijų ir iteracijų kiekį atsižvelgta į GCV klaidos (*angl. generalized cross validation error*) rodiklį. Didinant bazinių funkcijų skaičių didėja klaidų tikimybė modelyje t.y. į rezultatus įtraukiami mažiau reikšmingi ryšiai ir įtakos masyvai. Analizuojamų duomenų specifika ir atvejų gausa lėmė mažesnę iteracijų skaičių. Pastebėjus, kad rezultatų pokytis nereikšminis, o apdorojimo laikas sumažėjo trečdaliu buvo pasirinktas vienos iteracijos variantas.

Pasirinkus reikiamus parametrus, pradedama duomenų analizė. Modelio apmokymas pradedamas ieškant duomenų grupių, kurioms galima pritaikyti sugeneruotas bazines funkcijas. Bazinės funkcijos yra sudaromos remiantis tik eksperimentiniais duomenimis, bet ne ryšiais tarp jų. Modelio apmokymas šiam tyrimui realizuotas per 13 atvejų, kurių kiekvieno analizės metu nustatomi reikšminiai koeficientai iš tų kintamųjų grupių, kurios turi didžiausią įtaką prognozuojamam kintamajam. Paveiksle pateikti modelio apmokymo rezultatai.

Modelio koeficientai

| Model coefficients (PagalPreke) NOTE: Highlighted cells indicate basis functions of type max(0, independent-knot), otherwise max(0, knot-independent) | | | | | | | | | | |
|---|------------------------|-------------------|--------------------|--------------------|--------------------|---------------------|-----------------|--------------------|--------------------|-------------------------|
| Coeff. | Coefficients Kiekis | Knots Parduota | Knots Vid_kaina | Knots Min_kaina | Knots Max_kaina | Knots Klientų_sk | Knots id_kie | Knots web_click | Knots Susijos_k | Knots Specifikacijos |
| Intercept | -49.5928 | | | | | | | | | |
| Term.1 | 6.5589 | | | | | 96.00 | | | | |
| Term.2 | -1.6199 | | | | | 96.00 | | | | |
| Term.3 | -0.0141 | 61239.60 | | | | | | | | |
| Term.4 | 0.1645 | | | | 584.50 | | | | | |
| Term.5 | -0.0068 | 13283.04 | | | | | | | | |
| Term.6 | -6.2314 | | | | | 217.00 | | | | |
| Term.7 | 0.3183 | | | 656.37 | | | | | | |
| Term.8 | -0.3051 | | | 656.37 | | | | | | |
| Term.9 | -0.4919 | | 2663.15 | | | | | | | |
| Term.10 | 0.4743 | | 2663.15 | | | | | | | |
| Term.11 | 1.3523 | | | | | | | | | 16.09 |
| Term.12 | 0.0098 | 81809.00 | | | | | | | | |
| Term.13 | 4.0052 | | | | | 164.00 | | | | |

Šaltinis: sudaryta autoriaus.

MARSplines modelis konstruoja regresines funkcijas remiantis svertiniais koeficientais kartu su bazinių funkcijų rezultatais. Raudonai pažymėtos tos nepriklausomų kintamųjų reikšmės, kurios dalyvaus rezultatų modelio formavime. Iš pateiktos lentelės matome, kurie kintamieji turi įtakos prekių kiekio prognozei ir kuriuos kintamuosius galime eliminuoti.

Įtakų pasiskirstymas

| Number of References to Each Predictor (PagalPreke) Number of times each predictor is referenced (used) | |
|--|---------------------------------------|
| Dependents | References (to Basis Functions) |
| Parduota | 3 |
| Vid_kaina | 2 |
| Min_kaina | 2 |
| Max_kaina | 1 |
| Klientų_sk | 4 |
| Sand_kiekis | 0 |
| Web_click | 0 |
| Susijusios_Kiekis | 0 |
| Specifikacijos | 1 |

Šaltinis: sudaryta autoriaus.

Pagal pateiktą paveikslą matome, jog didžiausią įtaką daro klientų skaičius. Taip pat pastebime, kad nevienodą įtaką rezultatui turi minimali ir maksimali prekės kaina, pastaroji vidutiniškai du kartus mažiau įtakoja prekių kiekį.

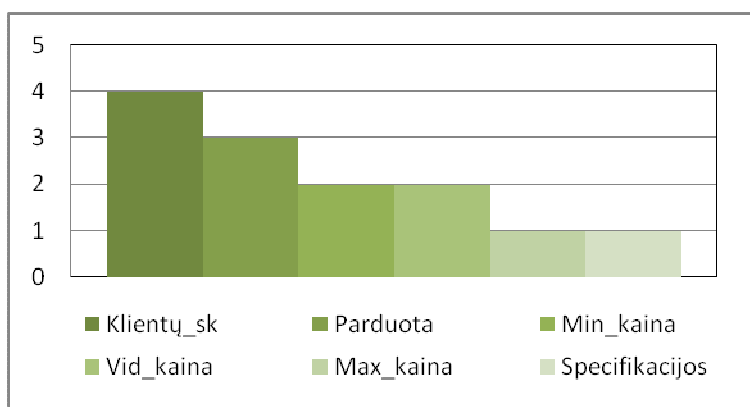
Pastebime, jog loginės sąsajos neturintis kintamasis – Specifikacijos, galutiniam prognozės modeliui turi tiek pat įtakos, kiek maksimali prekės kaina.

Apibendrinant šį tyrimo etapą, galime teigti, jog eksperimento pradžioje apsibrėžtus tikėtinius rezultatus galime patvirtinti tik iš dalies. Tikėtasi, jog iš loginės sąsajos neturinčių kintamųjų reikšmingi galutiniam rezultatui bus du. Tačiau pritaikius pasirinktą modelį, gauname, jog iš loginės sąsajos neturinčių kintamųjų (susijusios prekės, prekių kiekis sandėlyje, prekės domėjimasis per internetinę sąsają bei specifikacijos pilnumas), prognozuojamam modeliui įtakos turi tik specifikacijos pilnumas. Šio kintamojo įtaka yra lygi maksimaliai prekės kainai vasario mėnesį. Susiejus šių kintamųjų svertinius koeficientus, galime teigti, jog brangesnės prekės yra perkamos kai prekių specifikacijos yra pilnesnės.

Taigi, įtakos rezultatų modeliui nedaro:

1. Sandėlyje esantis tam tikros prekės kiekis – **Sand_kiekis**;
2. Skaičius, parodantis, kiek kartų vartotojas peržiūrėjo tos prekės detales - **Web_click**;
3. Susijusių prekių su analizuojama preke kiekis – **Susijusios_Kiekis**.

Minėti kintamieji neįtraukiami į galutinį prognozės modelį. Pateiktoje diagramoje pavaizduoti kintamieji pagal jų įtaką prognozuojamam modeliui.



Šaltinis: sudaryta autoriaus.

12 pav. Įtaką darančių kintamųjų pasiskirstymas

Įtraukus į dedamuosius kintamuosius tas duomenų grupes, kurios pavaizduotos 12 paveiksle, gauname GVC klaidos koeficientą lygų 12499674. Lyginant su modelio sudarymo klaidos koeficientu įtraukus visus kintamuosius, kaip GVC lygus 19669,17

MARSpline parametrai analizei

| Model specifications | Model Summary (PagalPreke) | | |
|---------------------------|----------------------------|--|--|
| | Value | | |
| Independents | 1 | | |
| Dependents | 6 | | |
| Number of terms | 5 | | |
| Number of basis functions | 4 | | |
| Order of interactions | 1 | | |
| Penalty | 2.000000 | | |
| Threshold | 0.000500 | | |
| GCV error | 12499674 | | |
| Prune | Yes | | |

Šaltinis: sudaryta autoriaus.

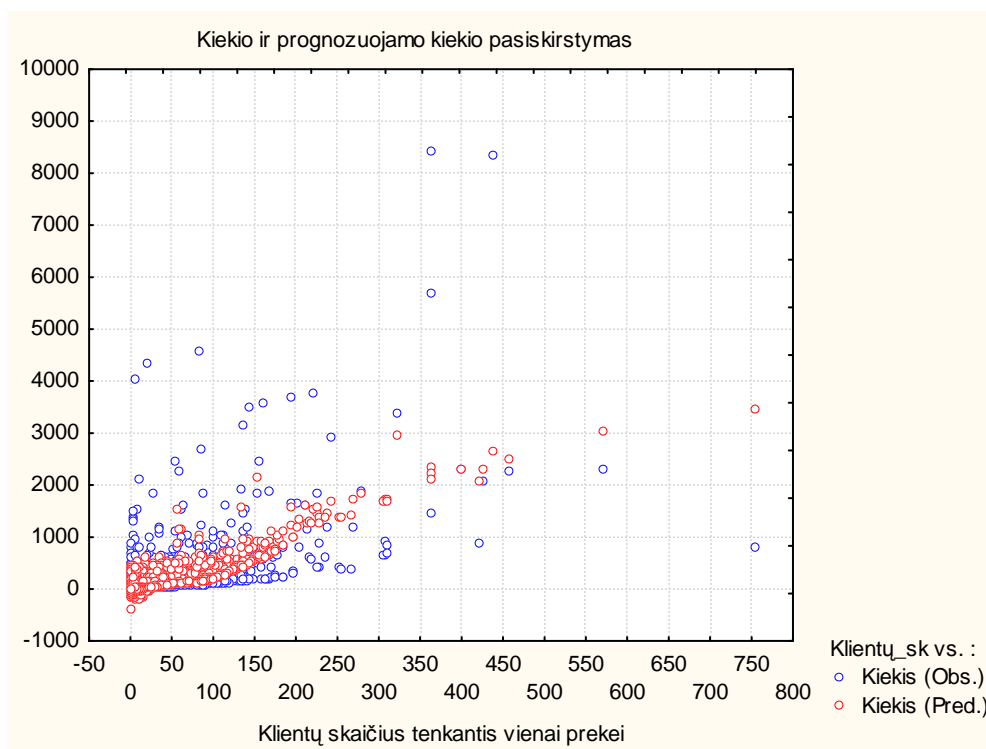
Pavaizduoti į modelį netraukti kintamieji bei jų pasiskirstymas pagal prognozei daromą įtaką. Darome išvadą, kas atliekant prekių kiekio prognozę netikslinga įvesti sandėliuojamą prekių kiekį, tos prekės aktyvavimo skaičių per kliento elektroninę sąsają bei susijusių prekių kiekį. Todėl taikant pateiktą modelį minėtų duomenų galima nekaupiti.

Tęsiant tyrimą, iš koeficientų, kurie įtakoja prognozės rezultatus sudaroma prognozės lygtis:

$$\begin{aligned}
 \text{Kiekis} = & -4.95928461575798e+001 + 6.55888371126071e+000 * \max(0, \\
 & \text{Klientu_sk} - 9.60000000000000e+001) - 1.61988687609117e+000 * \max(0, \\
 & 9.60000000000000e+001 - \text{Klientu_sk}) - 1.41116357930488e-002 * \max(0, \\
 & 6.12396000860000e+004 - \text{Parduota}) + 1.64460498171012e-001 * \max(0, \text{Max_kaina} - \\
 & 5.84499999000000e+002) - 6.75777016474354e-003 * \max(0, \text{Parduota} - \\
 & 1.32830398720000e+004) - 6.23141401735387e+000 * \max(0, \text{Klientu_sk} - \\
 & 2.17000000000000e+002) + 3.18306972531470e-001 * \max(0, \text{Min_kaina} - \\
 & 6.56370004000000e+002) - 3.05101856375438e-001 * \max(0, \\
 & 6.56370004000000e+002 - \text{Min_kaina}) - 4.91915875921864e-001 * \max(0, \text{Vid_kaina} - \\
 & 2.66314999850000e+003) + 4.74295087465882e-001 * \max(0, \\
 & 2.66314999850000e+003 - \text{Vid_kaina}) + 1.35228701779473e+000 * \max(0, \\
 & 1.60940282829070e+001 - \text{Specifikacijos}) + 9.83471938027821e-003 * \max(0, \\
 & \text{Parduota} - 8.18090004550000e+004) + 4.00522058106124e+000 * \max(0, \text{Klientu_sk} - \\
 & 1.64000000000000e+002)
 \end{aligned}$$

Pagal gautą lygtį sudaroma kiekio prognozė kiekvienai prekei ateinančiam laikotarpiui. Modelio išvesties rezultatas – prognozuojamos prekių kiekio reikšmės ateinančiam laikotarpiui.

Paveiksle pateiktas prognozės reikšmių pasiskirstymas pagal didžiausią įtaką turintį rodiklį – klientų skaičių.



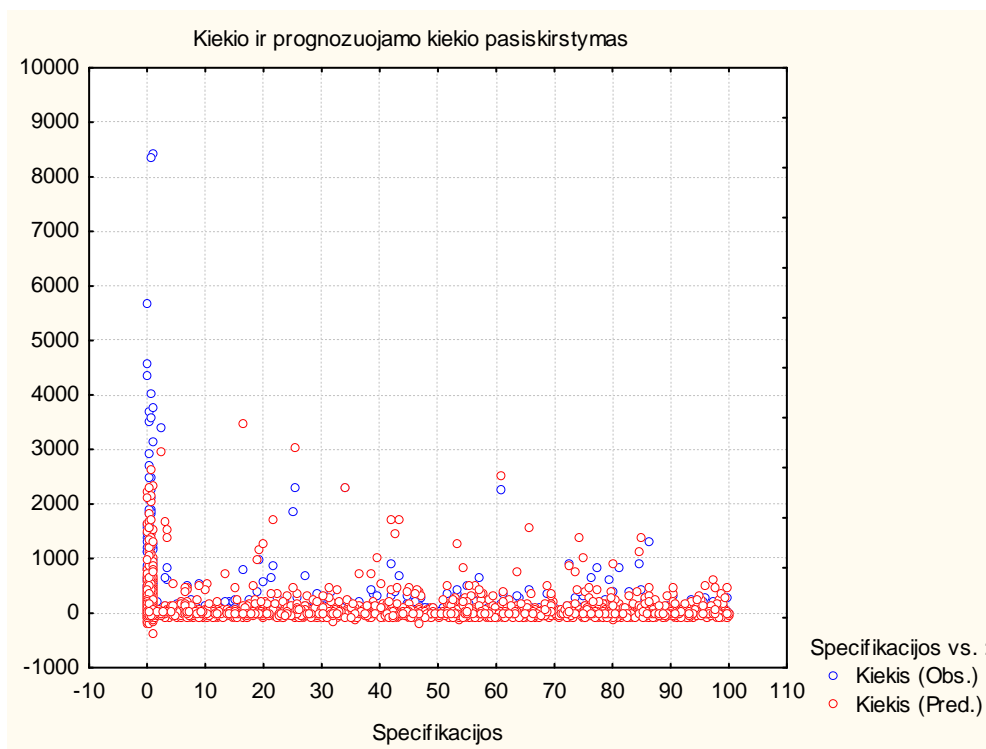
Šaltinis: sudaryta autoriaus.

13 pav. Kiekio ir prognozuojamo kiekio pasiskirstymas pagal klientų skaičių

Iš grafiko matyti, jog prognozės prekių kiekio pasiskirstymas yra mažesnės amplitudės, nei vasario mėnesio duomenų. Klientų nusipirkusių tam tikrą prekę kiekis yra prognozuojamos reikšmės vienas iš kintamųjų. Pagal klientų kiekį, prognozę ir stebėjimo rezultatus galime spręsti, jog didžiausia prekių kiekio sklaida yra nuo 50 iki 100 klientų tenkančių vienai prekei.

Iš čia darome išvadą, jog daugiausiai prekių nuperkama, kai prekę perka apie 50 klientų. Šį rodiklį galima panaudoti marketinginiams sprendimams, susijusiems su rėmimais orientuotais į ryšius su klientais. Taip pat šis rodiklis naudingas prekių pirkimo prognozavimui. Žinant, jog tam tikros prekės pardavimas išaugo 50 klientų, galime daryti prielaidą jog to mėnesio prekės pirkimas daugiau nebedidės, arba didės nežymiai.

Loginės sąsajos neturintis kintamasis – specifikacijos daro keturis kartus mažesnę įtaką prognozuojamai reikšmei. Tad specifikacijų pilnumo ir prognozuojamų reikšmių pasiskirstymo grafikas pateiktas žemiau.



Šaltinis: sudaryta autoriaus.

14 pav. Kiekio ir prognozuojamo kiekio pasiskirstymas pagal specifikacijų pilnumą

Prognozės reikšmių pasiskirstymas pagal specifikacijos rodiklį parodo, jog prekių paklausa yra tolygiai pasiskirsčiusi tarp prekių, kurių specifikacijos yra pilnos ir tų, kurių specifikacijos neatitinka reikalavimų. Galima būtų išskirti tas prekes, kurių specifikacijos yra nuo 0 iki 10 proc., jų paklausa išaugusi atskirom prekėm. Peržiūrėjus šias prekes, paaiškėjo, jog šis pokytis išaugo dėl to, jog klientų, kurių pirkimai neperiodiški, o perkamos prekės išsiskiria savo specifika.

Pagal prognozės išsidėstymą ir įtakos koeficientus galime spręsti, jog atliekant prekių kiekio prognozę pirkimams, reiktų atsižvelgti labiausiai į klientų skaičių rodiklį, kuris rodo prekės populiarumą. Prekių kiekio prognozę įtakoja tos prekės, kurias įsigijo daugiau klientų, nei tos, kurių mažiau, nepaisant to, jog vieno kliento įsigytų prekių kiekis yra didesnis nei kelių klientų įsigytų prekių kiekis.

4.4.1. Prognozės patikrinimo tyrimas.

Pritaikius MARSplines modelį gautą rezultatą – prekių kiekio prognozę ateinančiam laikotarpiui, patikrinsime su kovo mėnesio prekių kiekio duomenimis. Tolimesniam tyrimui atrinksime tik tas prekes, kurios buvo parduotos vasario mėnesį, t.y. į kovo mėnesio prekių

kiekio imtį neįtrauksime tų prekių, kurios nebuvo įsigytos vasario mėnesį ir tų prekių. Atrinkus duomenis, gaunama imtis su tiek pat duomenų, kiek ir vasario mėnesio prekių kiekio masyvas – 14093 prekės. Taip pat prieš atlikdami tolimesnę analizę prognozės duomenim įvedame sąlyga:

=iif(Prognozė<0,Prognozė=0,Prognozė)

Ši sąlyga išrenka neigiamas prognozuojamos prekių kiekio reikšmes. Taip pat koreguojame prognozuojamų duomenų formatą – suapvaliname reikšmes iki sveikos skaičiaus dalies.

Žemiau pateiktas tyrimo įvesties duomenų fragmentas.

16 lentelė

Įvesties duomenų fragmentas

| | 1 Kiekis02 | 2 Prognozė | 3 Kiekis03 |
|----|---------------|---------------|---------------|
| 1 | 14 | 0 | 9 |
| 2 | 15 | 25 | 9 |
| 3 | 21 | 32 | 13 |
| 4 | 1286 | 593 | 830 |
| 5 | 85 | 114 | 53 |
| 6 | 16 | 35 | 10 |
| 7 | 51 | 68 | 34 |
| 8 | 4 | 13 | 2 |
| 9 | 6 | 17 | 3 |
| 10 | 206 | 62 | 148 |
| 11 | 20 | 16 | 13 |
| 12 | 16 | 26 | 10 |
| 13 | 97 | 131 | 64 |
| 14 | 56 | 52 | 36 |
| 15 | 3 | 16 | 1 |
| 16 | 26 | 52 | 17 |
| 17 | 16 | 27 | 10 |
| 18 | 48 | 34 | 32 |
| 19 | 9 | 18 | 5 |
| 20 | 55 | 43 | 36 |
| 21 | 1 | 9 | 0 |
| 22 | 69 | 125 | 44 |
| 23 | 9 | 19 | 5 |
| 24 | 3 | 14 | 1 |
| 25 | 11 | 21 | 7 |
| 26 | 8 | 19 | 5 |
| 27 | 79 | 48 | 49 |
| 28 | 44 | 83 | 29 |
| 29 | 138 | 52 | 95 |
| 30 | 90 | 41 | 57 |
| 31 | 5 | 7 | 3 |

Šaltinis: sudaryta autoriaus.

Remiantis kelemečiais stebėjimais, vasario mėnesio prekių pirkimai būna sumažėję lyginant su kovo mėnesio pirkimais, kadangi gruodžio ir sausio mėnesio pirkimus sąlygoja didžiosios metų šventės. Kompanija prekiauja informacinių technologijų įranga, elektroninėmis buities prekėmis, todėl pirkimai laibiausiai išauga prieš naujuosius metus bei rugėsio mėnesį (mokslo metų pradžia). Po šių mėnesių, sekantys laikotarpiai dažnai yra

mažiau sėkmingi pardavimų atžvilgiu. Panašios tendencijos vyrauja ir pirkimų srityje. Prieš šventes gamintojai siūlo daugiau akcijų bei vertingų pasiūlymų distributoriams, taip skatindami parduoti siūlomas prekes. Pirkimo vadybininkai sudarinėja produktų grupes, kurias galima būtų parduoti kartu, todėl šventinių mėnesių metu gali išaugti pardavimai ir tų prekių, kurios prieš šį laikotarpį buvo parduodamos vangiai.

Po švenčių, vasario mėnesį pirkimų skaičius dažniausiai krenta ir normalizuojasi balandžio mėnesį. Todėl lyginant eksperimentinių duomenų prognozę su ateinančio laikotarpio duomenimis, svarbu įvertinti minėtus aspektus.

Vertinimui pateikiama trijų duomenų masyvų - vasario mėnesio stebėjimo, prognozės bei kovo mėnesio prekių kiekio koreliacinė matrica.

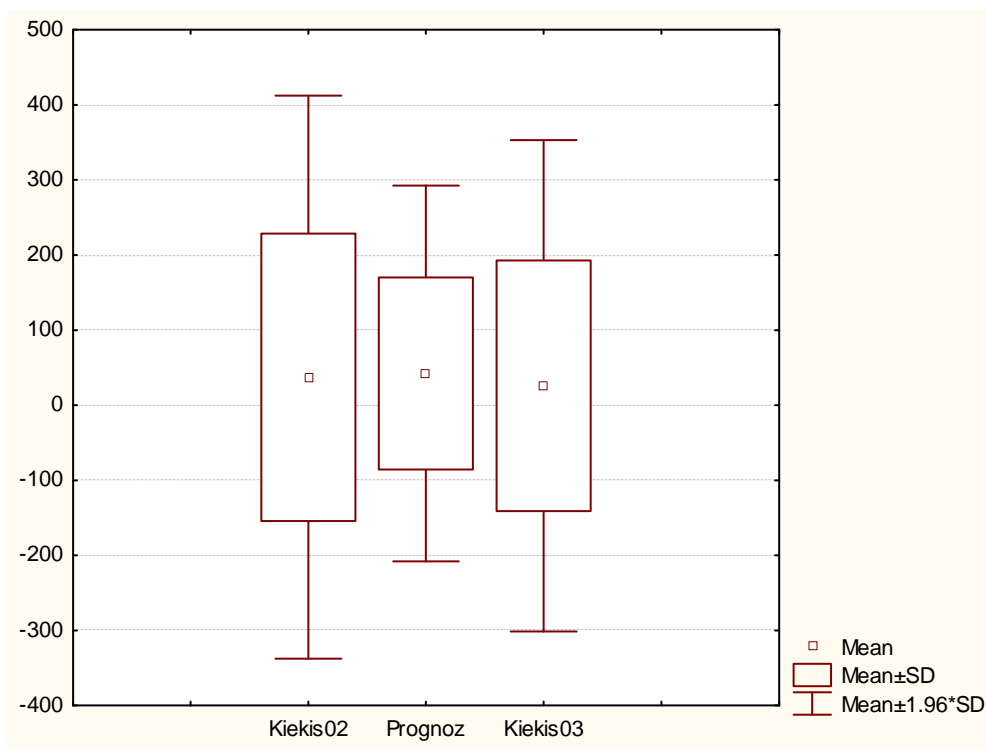
17 lentelė

Kiekio tyrimo koreliacinė matrica

| Correlations (Prognozes) | | | | |
|---|----------|----------|----------|--|
| Marked correlations are significant at p < .05000 | | | | |
| N=14093 (Casewise deletion of missing data) | | | | |
| Variable | Kiekis02 | Prognozė | Kiekis03 | |
| Kiekis02 | 1.00 | 0.68 | 0.97 | |
| Prognozė | 0.68 | 1.00 | 0.59 | |
| Kiekis03 | 0.97 | 0.59 | 1.00 | |

Šaltinis: sudaryta autoriaus.

Iš pateiktos lentelės sprendžiame, jog kovo mėnesio duomenys koreliuoja su vasario mėnesio koeficientu 0,97, o su prognozės – 0,59. Vertinant prognozuojamus duomenis, svarbus koreliacijos koeficientas, kuris artėjant vienetui rodo didesnę prognozuojamo modelio patikimumą su tikimybe mažesne už 0,05.



Šaltinis: sudaryta autoriaus.

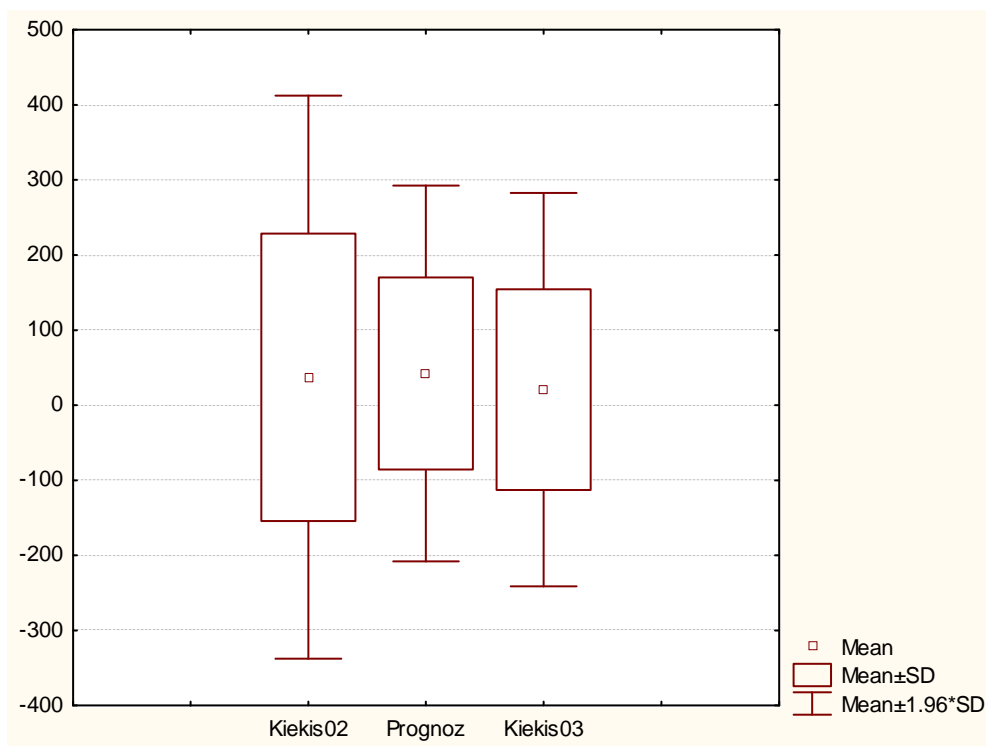
15 pav. Vidurkio ir standartinio nuokrypio pasiskirstymas

Iš grafiko matyti, jog prognozė pateikia duomenis mažesniu intervalu, nei buvo vasario mėnesio prekių kiekio pasiskirstymas. Lyginant su kovo mėnesio prekių kiekiu, prognozės prekių kiekio standartinis nuokrypis pasiskirstęs siauriau. Prognozuojamų duomenų vidurkis yra 42,26 vienetai vienai prekei, o kovo mėnesio prekių kiekis yra mažesnis – 25,93 vienetai. Šio parametro skirtumą lemia tai, jog prognozė sudaryta pagal vasario mėnesio pardavimus. Kadangi po švenčių, vasario mėnesį pirkimų skaičius dažniausiai, todėl lyginant eksperimentinių duomenų prognozę su ateinančio laikotarpio duomenimis, svarbu įvesti sezoniškumo koeficientą.

Sezoniškumo koeficiento tikslas yra paruošti skirtingų laikotarpių duomenis palyginimui. Šiuo atveju verslo specifika sąlygoja, jog kovo mėnesio pardavimai yra mažesni lyginant su vasario mėnesio pardavimais. Remiantis šia sąlyga įvedame lygtį:

$$(\text{Kovo mėnesio pardavimai}) = \text{Kiekis03} * 0.6$$

Įvedus sezoniškumo koeficientą, gauname palyginimo rezultatus, pateiktus grafike 16.



Šaltinis: sudaryta autoriaus.

16 pav. Vidurkio ir standartinio nuokrypio pasiskirstymas

Vidurkio ir standartinio nuokrypio pasiskirstymo grafikas pasikeitė. Vizualiai matomas prognozuojamų ir kovo mėnesio duomenų panašumas – standartinis nuokrypis prognozės apytikriai lygus 127 vnt., o kovo mėnesio duomenų apytikriai 134 vnt.

19 lentelė

Prognozės tinkamumo analizė statistiniai rezultatai

| Variable | Descriptive Statistics (Prognozes) | | | | |
|----------|------------------------------------|----------|----------|----------|----------|
| | Valid N | Mean | Minimum | Maximum | Std.Dev. |
| Kiekis02 | 14093 | 37.22096 | 1.000000 | 8424.000 | 191.2591 |
| Prognozė | 14093 | 42.26505 | 0.000000 | 3481.632 | 127.7138 |
| Kiekis03 | 14093 | 20.74460 | 0.000000 | 7132.000 | 133.5492 |

Šaltinis: sudaryta autoriaus.

Iš sudaryto modelio ir eksperimentinių rezultatų palyginimo galime spręsti, jog modelis tinka prekių kiekio prognozei, kadangi standartinis prognozuojamų duomenų nuokrypis yra 7 vienetais mažesnis už kovo mėnesio duomenų standartinį nuokrypį. Kadangi prognozė atliekama kiekvienai prekei atskirai, todėl aktualesnis rodiklis yra standartinis nuokrypis. Siūlomo modelio prognozė prekių pirkimams pagal standartinį nuokrypį yra 4 proc. mažesnė lyginant su kovo mėnesio duomenimis.

MARSplines duomenų gavybos metodas taip pat turi galimybę įvertinti sudarytos prognozės atitikimo sekančiam laikotarpiui įvertinimo galimybę. Pateikiamas įvertinimo langas, kuriame svarbiausias rodiklis yra koreliacijos koeficientas.

20 lentelė

Prognozės tinkamumo analizė

| | |
|--------------------------|--|
| | Summary Goodness of Fit (Prognozes) Observed variable: Kiekis03 |
| | Prognozė |
| Mean square error | 18846.27 |
| Mean absolute error | 30.33 |
| Correlation coefficient | 0.60 |

Šaltinis: sudaryta autoriaus.

Tam, kad įvertinti siūlomo modelio tinkamumą uždaviniui spręsti, atliksime prognozę tik pagal vieną loginę sąsają su prognozuojamu rodikliu turintį kintamąjį – Parduota. Šis rodiklis rodo už, kokią sumą buvo parduota prekių per ataskaitinį laikotarpį. Rezultatų palyginimui naudosime tą patį - MARSpline duomenų gavybos modelį.

Duomenų įvestį šiuo atveju sudaro prognozė pagal Parduota bei kovo mėnesio prekių kiekis.

21 lentelė

Prognozės tinkamumo analizė

| | |
|--------------------------|--|
| | Summary Goodness of Fit (Prognozes) Observed variable: Kiekis03 |
| | Prog_parduota |
| Mean square error | 24488.03 |
| Mean absolute error | 35.81 |

Šaltinis: sudaryta autoriaus.

Lyginant su siūlomo modelio prognoze gauti rezultatai rodo, jog prognozė pagal vieną rodiklį turi 15 proc. didesnę vidutinę kvadratinę klaidą (MSE). Remiantis šiuo įvertinimu galime teigti, kad pasiūlytas modelis tinka spręsti tiekimo grandinės uždaviniui – prekių kiekio prognozei.

4.5. TVARKARAŠČIŲ, GABENIMO LAIKO VALDYMO TYRIMO EIGA

Prekių gabenimo laiko valdymas tai vienas svarbiausių aspektų, sukuriančių pridėtinę vertę B2B verslo rinkoje. Pardavimai, fiksuojami ERP sistemose, veikia kaip indikatorius – įvertina sąlygas, kurios klientui buvo palankios. Pagal tai, galima vertinti ar pristatymo laikas yra optimalus tam tikram klientui. Prekių pristatymo laiko kontrolė leidžia pritaikyti marketingo veiksmus atskiram klientui.

Gabenimo laiko valdymas tiesiogiai susijęs su gabenimo laiko optimizavimu bei logistika. Nepaisant tų uždavinių, kurie sprendžiami tiekimo grandinės aspektu, naudojantis pardavimų duomenimis bei klientų aktyvumo statistika, toliau tyrime sieksime papildyti gabenimo laiko valdymo sprendimus naudojantis duomenų gavybos metodu.

Šiam uždaviniui įgyvendinti siūloma atlikti tyrimą, remiantis realiais duomenimis bei pasirinktu duomenų gavybos metodu. Gautus rezultatus, patikrinę statistiniais rodikliai, pateiktume kompanijos logistikos skyriui. Klientams, suklasifikuotiems pagal pateiktus kintamuosius, būtų galima taikyti vienodus logistinius sprendimus, rengti prekių gabenimo tvarkaraščius bei kontroliuoti prekių pristatymą tais pačiais aspektais.

Tyrimą galima suskirstyti etapais:

- (1) Nustatomos iš anksto žinomos grupės, kategorijos.
- (2) Priskyrimas kategorijoms vyksta per apmokymą su klasifikuotais duomenimis.
- (3) Sudaromas klasifikavimo modelis.
- (4) Modelis panaudojamas naujiems duomenims klasifikuoti.

Eksperimentinių duomenų populiaciją šiam uždaviniui sudaro visi susiję ir nesusiję su prekių pristatymo laikų kintamieji. Tačiau imtį sudarome iš lentelėje pateiktų vasario mėnesio duomenų.

22 lentelė

Įvesties laukų aprašymas gabenimo laiko valdymo tyrimui

| Lauko pavadinimas | Lauko tipas | Papildoma informacija |
|---------------------------------|-------------|--|
| Kliento kodas | String | Klientas identifikuojamas vietiniu identifikatoriumi |
| Vėlavimas | Decimal | Vidutinis prekių pristatymo vėlavimas dienomis |
| Dienu_visu | Integer | Iš viso vėluota pristatyti dienų |
| Max_dienų | Integer | Maksimalus dienų skaičius, kurių reikėjo laukti, kol prekės buvo pristatytos |
| Krepšelių suma | Decimal | Įsigyta prekių už šią sumą per vasario mėnesį. |
| Vidutinė krepšelių vertė | Decimal | Vidutinė krepšelio vertė per vasario mėnesį. |
| Įsigyta prekių vienetų | Integer | Įsigyta prekių, vienetais per vasario mėnesį. |
| Apsilankymų skaičius | Integer | Apsilankymų skaičius per vasario mėnesį. |

Šaltinis: sudaryta autoriaus

Kintamieji – vėlavimas, *Dienų_viso*, *Max_dienų* yra išvestiniai duomenys, gauti lyginant pageidautą prekių pristatymo laiką su realiu prekės pristatymo laiku. Likę duomenys yra susiję su pardavimais.

Šio tyrimo tikslas yra parinkti tokias klientų grupes, kurioms būtų būdingos to paties tipo reikšmės. Kadangi kintamieji yra kategorijos ir skaičiuojamojo tipo įprastų klasifikavimo priemonių tyrimui taikyti negalime.

Duomenų gavybos metodai STATISTICA 7 programoje gali atlikti skaičiavimus įtraukiant tiek kategorijas, tiek skaičiuojamuosius duomenų masyvus.

Duomenys suvedami į programos langą. Vasario duomenų masyvas paruošas klientų klasifikacijai pagal parinktus kintamuosius, duomenų fragmentas pateiktas priede 7.

Tyrimui atlikti naudojamas apibendrintas K- arčiausio nario klasterizavimo metodas – *Generalized K-means cluster analysis*. Pažymime kategorijos ir skaičiuojamuosius kintamuosius bei pasirenkame klasterių skaičių.

Renkantis klasterių skaičių buvo išbandyti trys variantai:

50 – suklasifikavus duomenų masyvą į penkiasdešimt grupių, pastebėta, kad apdoravimo ir metodo pritaikymo galimybės yra ribotos. Grupių parametrų skirtumai sumažėja, todėl taikyti skirtingus prekių pristatymo metodus tampa sudėtinga.

10 – optimalus klasterių skaičius, leidžiantis lengvai valdyti grupėms būdingą informaciją. Grupių skirtumai tokia klasterių skaičiuje išryškėja, todėl vadybininkams parinkti tinkamus pristatymo metodus yra lengviau.

5 – atlikus analizę su penkiais klasteriais, pastebėta, jog grupės sudaro klientai, kuriems taikyti vienodus pristatymo metodus rizikinga, dėl tikimybės nepatenkinti vartotojų poreikių bei prarasti dalį klientų, kuriems pristatymo laiko vėlavimas netoleruotinas.

Sekantis klasifikavimo parametro pasirinkimas – atstumo matavimo būdas. Iš sekančių pasirinkimų:

Euklidinis atstumas (Euclidean distance). Matuoja geometrinį atstumą tarp daugiadimensinėje erdvėje. Pasirinkus šį būdą, atstumo tarp kintamųjų reikšmių neįtakoja kiti kintamieji. Tačiau reikšmės gali įtakoti skirtingi kintamųjų reikšmių dydžiai. Šiuo atveju tyrime turime kintamuosius dienų skaičiaus pavidale, kurių maksimalus reikšmių skaičius nesiekia 50, taip pat tyrimo imtį sudaro ir kiti kintamieji – vidutinė krepšelio vertė, apsilankymų skaičius. Šių kintamųjų reikšmių intervalai skiriasi.

Kvadratinis Euklidinis atstumas (Squared Euclidean distance). Šiuo būdu skaičiuojant atstumus tarp kintamųjų reikšmių, reikšmių skalės skirtumai mažiau įtakoja klasifikavimo rezultatų nei skaičiuojant pagal euklidinį atstumą.

Svoriniai įvertinimai (Power distance). Klasifikavimas šiuo būdu atliekamas, kai norima įvesti svorinius koeficientus tam tikriems rodikliams.

Procentiniai nesutapimai (Percent disagreement). Taikant šį atstumų tarp reikšmių skaičiavimo metodą, įvertinami kategorijos kintamieji. Remiantis STATISTICA 7 rekomendacijomis, klasifikavimas šiuo būdu labiausiai tinka neišvestinėms kategorijoms.

Kadangi įvesties duomenų kintamųjų reikšmių intervalai skiriasi, pasirinktas kvadratinis Euklidinis atstumas duomenų klasifikacijai. Atstumai tarp kintamųjų reikšmių apskaičiuojami pagal pateiktą formulę:

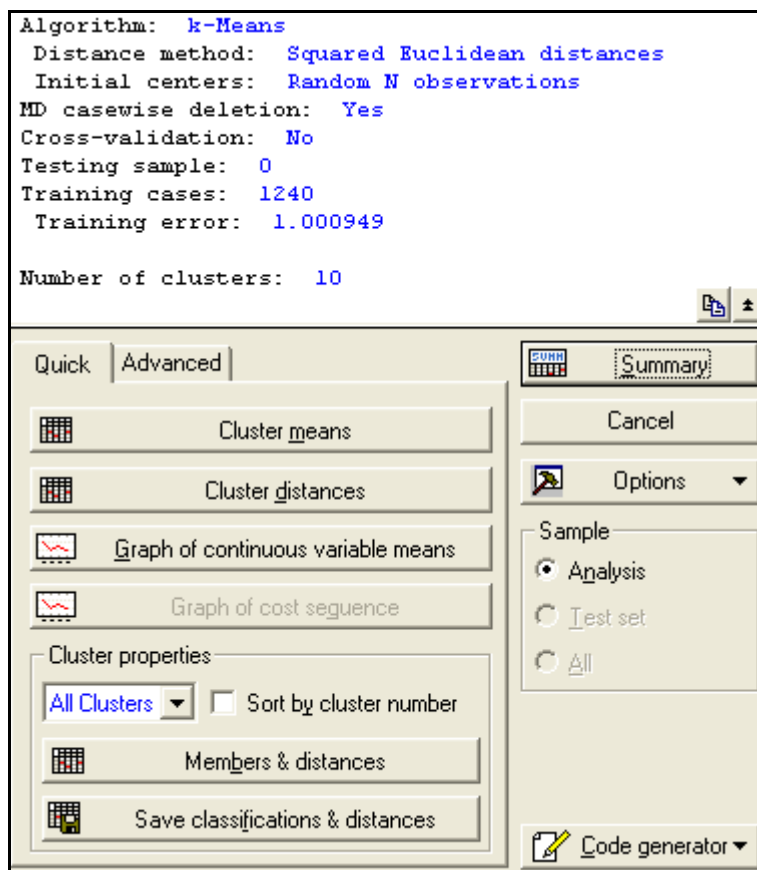
$$\text{Atstumas}(x,y) = \sum_i (x_i - y_i)^2 \quad (3)$$

Čia, x – pradinis taškas;

y – galutinis taškas;

i – kitimo žingsnis.

Pasirinkus dešimties klasterių variantą bei atstumo tarp reikšmių skaičiavimo metodą, gauname klasterių skaičiavimo bendrinį langą.



Šaltinis: sudaryta autoriaus

17 pav. Duomenų apdorojimo rezultatų langas

Gaunamas k-arčiausio nario klasifikavimo metodo rezultatus, kurie pateikti 15 lentelėje. Gautas duomenų masyvas sudarytas iš klientų kodų, grupės, kuriai priklauso tas klientas, įvesties kintamųjų reikšmės bei atstumas nuo centrinio parametro.

Pagal pastarąjį parametą galime spręsti apie grupės viduje esančių klientų priskyrimo tai grupei koeficientą – atstumą nuo centrinio nario. Gautas metodo rezultatas – klientų suskirstymas į 10 grupių (klasterių). Pateikta lentelė parodo kiekvienos gautos grupės statistinius duomenis – minimalią, maksimalią reikšmes, vidurkį bei standartinį nuokrypį.

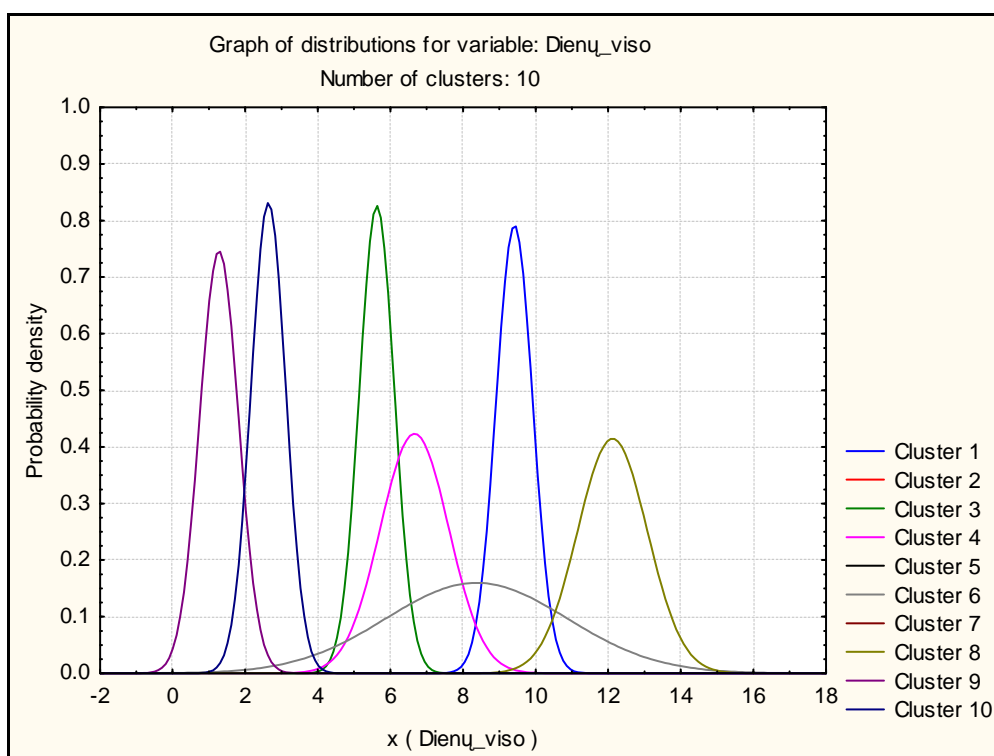
23 lentelė

Bendra statistinė informacija kiekvienai grupei

| | Statistics for continuous variable: Vėlavimas (PagalKlienta_gabenimo_laiko) | | | | | | | | | | |
|--------------------|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|----------|
| | Number of clusters: 10 | | | | | | | | | | |
| | Total number of training cases: 1240 | | | | | | | | | | |
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 | Cluster 8 | Cluster 9 | Cluster 10 | Overall |
| Minimum | 0.730000 | 0.570000 | 0.420000 | 0.470000 | 0.310000 | 0.310000 | 0.680000 | 0.990000 | 0.000000 | 0.160000 | 0.000000 |
| Maximum | 0.930000 | 0.620000 | 0.520000 | 0.730000 | 0.360000 | 1.090000 | 0.730000 | 1.300000 | 0.100000 | 0.260000 | 1.300000 |
| Mean | 0.847112 | 0.593437 | 0.465033 | 0.573333 | 0.335000 | 0.727500 | 0.707723 | 1.129719 | 0.054138 | 0.207941 | 0.613935 |
| Standard deviation | 0.056093 | 0.025082 | 0.039898 | 0.137961 | 0.025134 | 0.273274 | 0.024975 | 0.100904 | 0.037502 | 0.041194 | 0.142125 |

Šaltinis: sudaryta autoriaus

Lentelėje paryškintos maksimalios reikšmės kiekvienam statistiniam rodikliui. Pagal šiuos rodiklius galime atskirti ar įsivaizduoti, kokio tipo duomenys buvo priskirti vienai ar kitai grupei. Grupė 8 sudaryta iš duomenų, kuriose vyrauja didžiausi rodikliai – pavėluotų pristatyti dienų vidurkis, maksimalus ir minimalus dienų skaičius. Sekanti grupė 7 susideda iš klientų, kurių statistiniai pavėluotų pristatyti prekių rodikliai mažiausi. Tačiau bendros išvados apie grupių skirstymą priimti negalime, kadangi grupių sudarymas buvo paremtas 1240 skaičiuotu algoritmu, kuris periodiškai, esant sekančiai iteracijai keičia klasifikavimo būdus pagal ankstesnio algoritmo rezultatų duomenis.

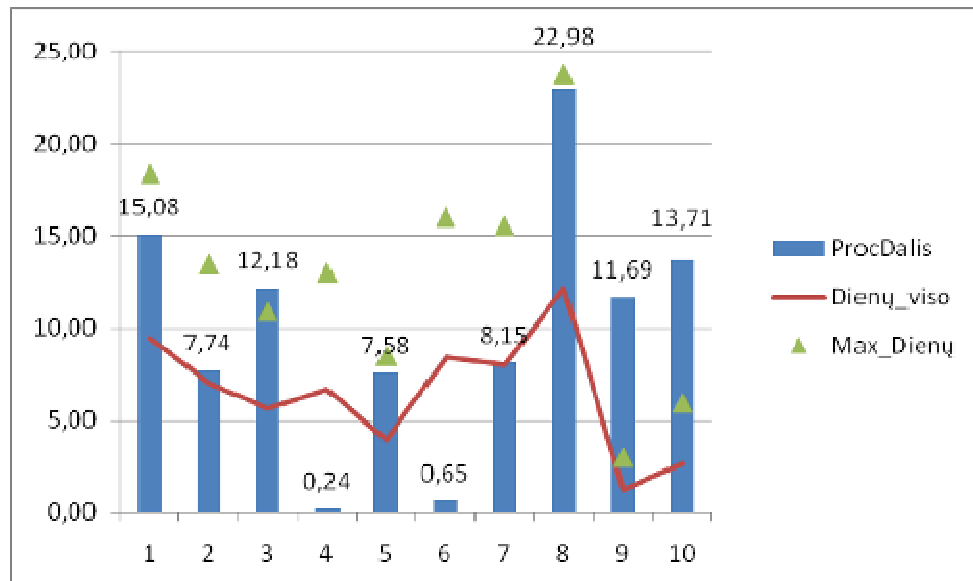


Šaltinis: sudaryta autoriaus

18 pav. Pavėluotų pristatyti dienų pasiskirstymas pagal grupes

Tyrimo rezultatai pateikiami pagal dienų skaičiaus kintamąjį. Grafiškai didžiausią tikimybę klientams būti priskirtiems vienoje grupėje turi 1, 3, 10 ir 9 klasteriai. Šių grupių nulinės hipotezės tikimybė, rodo, jog grupavimo tikslumas yra nuo 0,75 iki 0,83.

Pateikiamas grafikas, rodantis grupių pasiskirstymą procentaliai.



Šaltinis: sudaryta autoriaus

19 pav. Grupių pasiskirstymas pagal maksimalų pristatyti pavėluotų dienų skaičių

Gautos 10 grupių, kurios yra skirtingai jautrios prekių pristatymo laikui. Šis skirstymas aktualus vadybininkams užsakantiems prekes, priimantiems marketinginius sprendimus. Atlikus klasifikaciją pagal pristatymo laiko vėlavimą, galima taikyti skirtingus pristatymo metodus vienai grupei, taip optimizuojant tiekimo grandinės gabenimo ir logistikos uždavinius.

IŠVADOS

1. Darbe apžvelgti moksliniai tyrimai duomenų gavybos metodų taikymo verslo srityse bei pateiktas įvertinimas, rodantis, jog duomenų analizė, remiantis tik ryšių su klientais valdymo sistemų duomenimis, yra nepilna. Atsižvelgiant į verslas verslui problematiką, reikalinga papildyti analizę klientų aptarnavimo, pardavimo kanalų valdymo ir tiekimo grandinės kintamaisiais.
2. Darbe pateiktas modelis, apimantis klientų kreditavimą, prekių kiekio prognozę bei pristatymo laiko jautrumo analizę klientams, ir atliktas siūlomo modelio eksperimentinis tyrimas.
3. Atlikus kreditų teikimo valdymo galimybių tyrimą, paaiškėjo, jog duomenų gavybos klasifikavimo metodas yra tinkamas šiam uždaviniui spręsti. Rezultatai parodė, jog kreditavimo grupes tikslinga perskirstyti į 6 vietoje 7. Neuroninio tinklo analizės pagalba nustatyta, kad 20 tūkst. kreditavimo suma netenkina klientų poreikio. Tiems klientams, kuriems anksčiau buvo teikiamas 20 tūkst. kredito limitas, suteikiami 10 tūkst. arba 50 tūkst. kreditai remiantis svoriniais koeficientais.
4. Taip pat gauti rezultatai parodė, jog 0,89 proc. didesnę įtaką kredito kategorijos priskyrimui daro apsilankymų skaičius nei vidutinė krepšelio suma.
5. Atlikus prekių kiekio prognozės eksperimentinį tyrią remiantis MARSplines modeliu, gavome, jog iš loginės sąsajos neturinčių kintamųjų (susijusios prekės, prekių kiekis sandėlyje, prekės domėjimasis per internetinę sąsają bei specifikacijos pilnumas), prognozuojamam modeliui įtakos turi specifikacijos pilnumas. Šio kintamojo įtaka yra lygi maksimaliai prekės kainai vasario mėnesį. Taip pat pastebėta, jog prekės, kurių specifikacijų pilnumas yra didesnis nei 50 proc. pardavimai yra 2,3 proc. didesni nei tų, kurių specifikacijos yra nepilnos. Siekiant didesnių pardavimų rekomenduotina užtikrinti kuo didesnę specifikacijų pilnumą.
6. Pagal bazinių funkcijų skaičių, įtakos prekių kiekio prognozės modeliui nedaro šie kintamieji: sandėlyje esantis tam tikros prekės kiekis – **Sand_kiekis**; skaičius, parodantis, kiek kartų vartotojas peržiūrėjo tos prekės detales - **Web_click**; susijusių prekių su analizuojama preke kiekis – **Susijusios_Kiekis**. Todėl taikant pateiktą modelį minėtų duomenų galima nekaupiti.
7. Prekių kiekio prognozės gauti rezultatai parodė, jog didžiausią įtaką prognozei daro klientų, nusipirkusių tą prekę, skaičius. Pagal šį kintamąjį modelis buvo sudarytas remiantis 4 bazinėmis funkcijomis iš 13 galimų. Taip pat pastebėta, jog šis rodiklis naudingas

momentinei prekių pirkimo prognozei. Žinant, jog tam tikros prekės pardavimas išaugo 50 klientų, pirkimų vadybininkas gali daryti išvadą, jog to mėnesio prekės pirkimas daugiau nebedidės, arba didės nežymiai.

8. Lyginant su siūlomo prekių kiekio prognozės modeliu gauti rezultatai rodo, jog prognozė pagal vieną rodiklį turi 15 proc. didesnę vidutinę kvadratinę klaidą (MSE). Remiantis šiuo įvertinimu galime teigti, kad pasiūlytas modelis tinka spręsti tiekimo grandinės uždaviniui – prekių kiekio prognozei.
9. Klientų klasifikacijos pagal pristatymo laiką eksperimentinio tyrimo metu nustatyta, kad 10 yra optimalus klusterių skaičius, leidžiantis lengvai valdyti grupėms būdingą informaciją. Grupių skirtumai tokia klasterių skaičiuje išryškėja, todėl vadybininkams parinkti tinkamus pristatymo metodus yra lengviau.
10. Atlikus klasifikaciją pagal pristatymo laiko vėlavimą, galima taikyti skirtingus pristatymo metodus vienai grupei, taip optimizuojant tiekimo grandinės gabenimo ir logistikos uždavinius. Gautos 10 grupių, kurios yra skirtingai jautrios prekių pristatymo laikui. Šis skirstymas aktualus vadybininkams užsakantiems prekes, priimantiems marketinginius sprendimus.

PASIŪLYMAI

1. Atlikus CRM klasifikacijos apžvalgą pastebėta, kad efektyviausi informacijos analizės rezultatai naudojant duomenų gavybos metodus gaunami analitinėse ryšių su klientais valdymo sistemose, todėl rekomenduojama siūloma modelį integruoti į analitinę ryšių su klientais valdymo sistemą;
2. Išskirtos trys tiekimo grandinės dalys – klientų kreditingumas, prekių kiekio kontrolė, bei prekės pristatymo laiko įtaka ryšiams su klientais gali būti papildytos kompleksiniais duomenų gavybos metodais. Tolimesnėje mokslinėje veikloje galima būtų atlikti duomenų analize remiantis tiek fundamentaliais, tiek išvestiniais duomenų gavybos metodais.
3. Ryšių su klientais valdymo sistemose įdiegtas klasifikavimo naudojantis neuroniniais tinklais duomenų gavybos metodas padėtų tiksliau atlikti klientų segmentavimą pagal suteikiamus kreditus, o tai leistų adaptuoti paslaugas klientų poreikiams, kuris sąlygotų didesnę klientų pasitenkinimą. Atskiroms, pagal kredito kategorijas, priskirtoms klientų grupėms galima būtų taikyti skirtingas kainodaros strategijas. Šie veiksmai skatina ryšius su klientais, o tai padeda geriau suprasti konkurencinę situaciją rinkoje, kuria B2B kompanijos pridėtinę vertę, skatina inovacijas.
4. Būtina atsižvelgti į tai, jog taikant kompleksinius metodus, tampa sudėtingiau suprasti, kuris metodas turėjo didesnę įtaką duomenų teisingumui ar klaidingumui, todėl labai svarbu griežtai specifiuoti, kokie duomenys bus naudojami, kokio tikimasi rezultato.
5. Gavybos metodų naudojimas turi turėti ilgalaikes perspektyvas, kadangi verslas verslui sistemos diegimas reikalauja finansinių investicijų;

LITERATŪRA

1. BERRY, Michael J.A. (1997) *Data mining techniques : for marketing, sales, and customer support*. New York: J. Wiley. ISBN: 0-471-17980-9
2. LAROSE, Daniel T., (2006) *Data mining methods and models*. Wiley-Interscience J. Wiley. ISBN: 0-471-66656-4
3. TANG, ZhaoHui. (2005) *Data mining with SQL server 2005*. Indianapolis: Wiley Publishing. ISBN: 978-0-471-46261-3
4. ARNOLD & PORTER UPDATE. (2000) *B2B Marketplaces & Common Sense Antitrust Precautions*. Prieiga prie internetu: <http://www.arnoldporter.com/pubs/files/b2bclientupdate7192000.pdf> [žiūrėta: 2007-12-03].
5. BIELIŪNAS, Martynas. (2000) *Žinių vadybos praktinis taikymas: pokyčiai, kuriuos lemia ekonomikos tendencijos*. Prieiga prie internetu: <http://www.leidykla.vu.lt/inetleid/inf-mok/14/str4.html> [žiūrėta 2008-05-22];
6. DR. WEGMAN. (2005) *IT 871/CSI 709 – Statistical Data Mining*. Prieiga prie internetu: <http://www.galaxy.gmu.edu/stats/syllabi/IT871/IT871InternetSurveyReport.pdf> [žiūrėta 2008-09-22]
7. HEDLUND Johan, INGMAN Ana. (2008) *Unlocking the value of CRM – from the perspective of a global IT company*. Prieiga prie internetu: <http://epubl.luth.se/1402-1617/2008/082/LTU-EX-08082-SE.pdf> [žiūrėta 2008-09-23]
8. JAMES EDWARD, Richard. (2008) *The Impact of Customer Relationship Management (CRM) Technology on Business-to-Business Customer Relationships*. Prieiga prie internetu: <http://researcharchive.vuw.ac.nz/bitstream/handle/10063/376/thesis.pdf?sequence=1> [žiūrėta 2008-09-23]
9. JAIDEEP, Srivastava. (2005) *Data Mining for Customer Relationship Management (CRM)*. Prieiga prie internetu: <http://www.dtc.umn.edu/ddmc/resources/crm.pdf> [žiūrėta: 2007-11-26].
10. JOUNGHAЕ, Bang, NIKHILESH, Dholakia, LUTZ, Hamel, RUBY ROY, Dholakia. (2004) *Data mining of CRM knowledge bases for effective market segmentation - A Conceptual Framework*. Prieiga prie internetu: <http://www-lih.univ-lehavre.fr/Intranet/proceedings/ICEIS2004/ICEIS%202004/Area%20%20-%20Artificial%20Intelligence%20and%20Decision%20Support%20Systems/Oral%20Presentations/Full%20Papers/C2318Bang.pdf> [žiūrėta 2008-09-22]

11. KENNETH, Elliott, SCIONTI, Richard, PAGE, Mike Ph.D. (2003) *The Confluence of Data Mining and Market Research for Smarter CRM*. Kenneth Research Inc. Prieiga prie interneteto: <http://www.spss.gr/whitepapers/tworivers.pdf> [žiūrėta 2008-09-23]
12. KLEIN, David. (1998) *The strategic management of intellectual capital*. Business management. Butterworth Heinemann press. 32-48, p. ISBN 0-7506-9850-01.
13. Lietuvos Ekonominės plėtros agentūra, (2007) *Praktinio informacinių technologijų taikymo ir elektroninio verslo sprendimų įmonėse metodinis vadovas: Veiklos kaštų mažinimas informacinių technologijų pagalba*. Prieiga prie interneteto: <http://www.verslovertai.lt/files/File/vadovas/6-VeiklosKastuMazinimas.pdf> [žiūrėta 2008-11-29]
14. Lietuvos Ekonominės plėtros agentūra, (2007) *Praktinio informacinių technologijų taikymo ir elektroninio verslo sprendimų įmonėse metodinis vadovas: Santykiu su klientais ir tiekėjais* va <http://www.verslovertai.lt/files/File/vadovas/3-SantykiuSuKlientaisValdymas.pdf> [žiūrėta 2008-11-29];
15. NISBET, Robert. (2006) *Data Mining Tools: Which One is Best for CRM? Part 3*. Prieiga prie interneteto: <http://www.dmreview.com/specialreports/20060321/1049954-1.html> [žiūrėta 2008-05-30];
16. OKSL KTU studijos, (2008) *Tiekimo grandinės valdymas*. Prieiga prie interneteto: www.oksl.ktu.lt/studijos/T120B120/slide/12.pdf [žiūrėta 2009-01-04]
17. OKSL KTU studijos, (2008). Duomenų valdymas, duomenų sandėliai. Prieiga prie interneteto: <http://www.oksl.ktu.lt/studijos/T120B120/slide/06.pdf> [žiūrėta 2008-06-13].
18. POWERIN ANALYTIC DECISIONS. (2006) *Five Common Misconceptions About B2B Data Mining* Prieiga prie interneteto; <http://www.asacorp.com/docs/publicdocs/ASAcatalogeb2b.pdf> [žiūrėta: 2007-12-03];
19. RAMAKRISHNAN, Raghu, AGRAWAL, Rakesh, FREYTAG, Johann-Christoph ir kiti. (2005) *Data Mining: The Next Generation*. Prieiga prie interneteto: <http://drops.dagstuhl.de/opus/volltexte/2005/270/pdf/04292.DagstuhlManifesto.270.pdf> [žiūrėta: 2007-11-23].
20. RAHMAN IJAS, Dar, YONGCHAO, Hu (2005) *How banks manage CRM – B2B perspective*. Prieiga prie interneteto: <http://epubl.luth.se/1404-5508/2005/065/LTU-SHU-EX-05065-SE.pdf> [žiūrėta 2008-09-23]
21. STATSOFT. (2008). *Data Mining Techniques* Prieiga prie interneteto: <http://www.statsoft.com/textbook/stdatmin.html#mining> [žiūrėta 2008-06-13].

22. STAT SOFT, (2006). *STATISTICA Data miner brochure*. Prieiga prie interneto: http://www.statsoft.co.uk/STATISTICA_Data_Miner.pdf [žiūrėta 2009-01-05]
23. SUNIL, Arte, CHIEN-CHIH, Lin, PUTTIKAN, Prapai.(2005) *Mining Internet user survey data Improving Customer Relationship Management (CRM)* Prieiga prie interneto: <http://www.galaxy.gmu.edu/stats/syllabi/IT871/IT871InternetSurveyReport.pdf> [žiūrėta: 2007-11-23].
24. TIBERIUS Data Mining (2007), *Back Propagation Weight Update Rule*. Prieiga prie interneto <http://www.tiberius.biz/bpproof.html> [žiūrėta 2009-01-04]
25. XIAOSHAN, Du.(2006) *Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship*. Prieiga prie interneto: <http://www.diva-portal.org/diva/getDocument?urn=bnbnesevxeudiva-865-2eeefulltext.pdf> [žiūrėta: 2007-12-01].
26. YING, Gao, DEZHEN, Feng.(2003) *Research on Application of Data Mining Technology in CRM* Prieiga prie interneto: <http://www.cs.swufe.edu.cn/xsjl/sixwuhan/Paper/EB324.pdf>. [žiūrėta: 2007-11-26].
27. ZAVAREH, Javad. (2007) *The role of Analytical CRM in Maximizing Customer Profitability in private banking*. Prieiga prie interneto: <http://epubl.luth.se/1402-1552/2007/065/LTU-DUPP-07065-SE.pdf> [žiūrėta 2008-06-13].

PRIEDAI

| | |
|--|----|
| 1 PRIEDAS Literatūros apžvalga..... | 79 |
| 2 PRIEDAS | 82 |
| 3 PRIEDAS CRM procesai ir funkcijos | 84 |
| 4 PRIEDAS Programinių paketų rinka..... | 85 |
| 5 PRIEDAS Programinės įrangos su DGM palyginimas | 86 |
| 6 PRIEDAS STATISTICA 7 rezultatai..... | 88 |

LITERATŪROS APŽVALGA

24 lentelė

Literatūros šaltiniai rasti naudojantis paprastomis paieškos sistemomis

| Puslapio adresas | Kalba | Aprašymas |
|---|------------|---|
| http://www.statsoft.com/textbook/stdatmin.html#mining | Anglų klb. | Duomenų gavybos apibrėžimai bei samprata. Duomenų gavybos metodų konstravimas, modeliai. Atkreipiamas dėmesys į prognozuojamų duomenų gavyba bei jai įgyvendinti skirtų algoritmų sudarymą. <i>Data Mining: Model building and validation, Deployment. Crucial Concepts in Data Mining: Bagging (Voting, Averaging), Boosting, Models for Data Mining, Models of Data Mining: Predictive Data Mining, SEMMA, Stacked Generalization</i> |
| http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm | Anglų klb. | Šaltinyje analizuojama, kas yra duomenų gavyba, kokiomis technologijomis paremta. Plačiau apžvelgiama, kokios infrastruktūros reikalauja duomenų gavybos modelių įgyvendinimas. <i>Data Mining: What is Data Mining?, Example, How does data mining work? What technological infrastructure is required? Query complexity:</i> |
| http://www.spss.com/datamining/index.htm?source=homepage&hpzone=tech | Anglų klb. | Duomenų gavyba spėdimų priėmimo, SPSS, duomenų gavybos priemonės, teksto gavyba, WEB gavyba. Tiriamas duomenų gavybos efektyvumas ir panaudojimo galimybės realiuose verslo procesuose. <i>Data Mining Improves Decision Making, Data mining helps SPSS customers solve business problems; Using data mining tools, Clementine data mining transforms data into actionable results; The latest data mining advances—text mining and Web mining</i> |
| http://www.the-datamine.com/bin/view/Software/AI/DataMiningSoftware | Anglų klb. | Šaltinyje apžvelgtas programų sąrašas, išskirti jų ypatumai, pateiktos rekomendacijos. Pagrindiniai paminėti programiniai paketai: <i>Angoss KnowledgeACCESS, ARMiner, Clementine, Compumine Rule Discovery System, DMax Chemistry Assistant</i> |
| http://www.datamentors.com/content/view/65/58/ | Anglų klb. | Duomenų gavyba, analizė ir duomenų bazių realizavimas rinkoje. Šaltinyje nagrinėjami duomenų gavybos metodai duomenų bazėse, apžvelgiami gauti rezultatai. Atskleidžiama sąsaja tarp duomenų gavybos metodų ir marketingo funkcijų. <i>Data Mining, Analysis and Database Marketing</i> |
| www.siam.org/meetings/sdm02/proceedings/sdm02-34.pdf | Anglų klb. | Duomenų gavybos technologijos, kokybės realizavimas. Pateiktas analitiškas požiūris į duomenų gavybos metodų ir kokybinių rezultatų koreliaciją. Išvardinami duomenų gavybos metodai labiausiai įtakojantis duomenų atrinkimo procesą. <i>Data Mining Techniques to Quantitative Analysis of Conspiracy Theory: In the current analysis, data mining was applied to the USDA RMA "national book of insurance business" for the year 2000 to quantify conspiracy.</i> |
| http://www.thearling.com/dmintro/dmintroiframe.htm | Anglų klb. | Duomenų gavyba ir CRM. Analizuojama sąsaja tarp duomenų gavybos ir klientų valdymo sistemų. Pateikiamos sistemų sritys, kuriose duomenų gavybos metodai gali būti realizuoti. |

1 PRIEDAS (TĘSINYS)

24 lentelės tęsinys

| Puslapio adresas | Kalba | Aprašymas |
|---|------------|---|
| http://www.dmreview.com/article/sub.cfm?articleId=1046025 | Anglų klb. | Pagrindiniai duomenų gavybos uždaviniai, modeliavimo įgyvendinimas ir įrankiai, procesų modeliai. Šaltinyje nagrinėjamas bendras duomenų gavybos modelių įgyvendinimo procesas. Išvardinta modelių kūrimo ir veikimo darbų atlikimo seka. <i>Major Data Mining Tasks, Data Preparation Tasks, Model Evaluation Tools, The Data Mining Process, Process Models</i> |
| http://globalcardservices.com/products.asp?vid=14&nid=26 | Anglų klb. | Duomenų gavyba ir CRM, klientų analizė, lojalumo analizė ir kt. Šaltinyje pateikiama duomenų gavybos analizė valdymo sistemų aspektu. Analizuojami verslo procesai bei jų pokyčiai pritaikius duomenų gavybos technologijas. Pateikta apibendrinamoji klientų lojalumo, ryšių tarp klientų, pardavimų, klientų profilio analizė. <i>Data Mining, CRM and Business Intelligence, Customer Relationship Analytics, Campaign Analysis, Customer Profiling, Loyalty Analysis, Sales Analysis, Customer Care Analysis</i> |
| http://www.greaterchinacrm.org:8080/eng/content/details.jsp?contentid=1037&subjectid=19 | Anglų klb. | Olap modeliai, apibendrinančioji analizė. Pagrindinė aiškinamoji medžiaga sistemų modeliavimui. Aprašomi esminiai analizės principai. Plačiau tiriamas prognozavimo modeliavimas. <i>OLAP (On-Line Analytical Processing), Descriptive Analysis, Predictive Modeling</i> |
| http://www.cpm.com/providers/solutions/crmemining.cfm | Anglų klb. | Duomenų gavyba, modeliai, analizė, duomenų struktūros. Šiame šaltinyje tiriami analogiški duomenų gavybos metodams veiklos stebėjimo metodai. Pateikiamos duomenų analizės alternatyvos. Analizuojama duomenų bazės struktūra. <i>Data Model, Data Analysis, Campaign Management, Database Construction, Performance Monitoring Dashboard</i> |
| http://databases.about.com/od/databasing/DataMiningandDataWarehousing.htm | Anglų klb. | Šaltinyje pateikiamas duomenų valdymo metodų klasifikavimas. Struktūriškai vaizduojama dalykinė sritis, pateikiama duomenų gavybos ir duomenų talpinimo teorija. Pristatomi metodų evoliucijos etapai. <i>Classification, Data Mining and Data Warehousing</i> |
| http://www.tdwi.org/MarketPlace/subcat.aspx?catId=83 | Anglų klb. | Šaltinyje pateikiamas duomenų gavybos metodų programinių paketų sąrašas. |
| http://drops.dagstuhl.de/opus/volltexte/2005/270/pdf/04292.DagstuhlManifesto.270.pdf | Anglų klb. | Duomenų gavybos metodų panaudojimas CRM ir LS (angl. Life science). Apražiamos duomenų charakteristikos. Antrame šaltinio skyriuje pateikiamas CRM ciklas. Aptariami duomenų gavybos rezultatai. Analizuojami specifiniai duomenų gavybos metodų klientų valdymo sistemose faktoriai. RAMAKRISHNAN, Raghu, AGRAWAL, Rakesh, FREYTAG, Johann-Christoph ir kiti.(2005) <i>Data Mining: The Next Generation.</i> |
| http://www.galaxy.gmu.edu/stat/syllabi/IT871/IT871InternetSurveyReport.pdf | Anglų klb. | Straipsnyje pateikiami pagrindiniai duomenų gavybos metodai bei jų įtaka klientų valdymo sistemų kūrimo procesui. Pateikiami eksperimentiniai rezultatai. Plačiau analizuojami duomenų gavybos internetu metodai. SUNIL, Arte, CHIEN-CHIH, Lin, PUTTIKAN, Prapai.(2005) <i>Mining Internet user survey data Improving Customer Relationship Management (CRM)</i> |

1 PRIEDAS (TĘSINYS)

24 lentelės tęsinys

| Puslapio adresas | Kalba | Aprašymas |
|---|------------|--|
| http://www.cs.swufe.edu.cn/xsjl/sixwuhan/Paper/EB324.pdf | Anglų klb. | Šaltinyje grindžiami duomenų gavybos metodo naudojimo aspektai klientų valdymo sistemose. Detaliai pateikiama klientų poreikių analizė. Su ja lyginami duomenų gavybos metodų rezultatai pateikiami struktūriškai. YING, Gao, DEZHEN, Feng. <i>Research on Application of Data Mining Technology in CRM</i> |
| http://www.dtc.umn.edu/ddmc/resources/crm.pdf | Anglų klb. | Apžvelgiama duomenų gavybos metodų panaudojimo sritis. Pateikiamas požiūris: duomenų gavybos metodų įrankiais atliekama <i>GAP analysis</i> . Išvardinami ir aprašomi duomenų gavybos pagrindiniai ir išvestiniai metodai. JAIDEEP, Srivastava. <i>Data Mining for Customer Relationship Management (CRM)</i> . |
| http://www.diva-portal.org/diva/getDocument?url=bnbnesevxxudiva-865-2efulltext.pdf | Anglų klb. | Straipsnyje akcentuojamas duomenų paruošimas duomenų gavybos metodams. Struktūriškai pateikiami fundamentalūs duomenų gavybos metodai. XIAOSHAN, Du.(2006) <i>Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship</i> . |

Šaltinis: sudaryta autoriaus

25 lentelė

Informacijos paieška populiariose mokslinės informacijos duomenų bazėse

| DB pavadinimas | Adresas Knygų, straipsnių bibliografinis aprašymas |
|--|--|
| Kembridžo universiteto duomenų bazė: http://www.lib.cam.ac.uk/electronicresources/databases.html | Pateikiamas metodas – teksto duomenų gavyba. Analizuojamas jo taikymas klientų valdymo sistemose, žinių valdymo ir analizės modeliuose. Text mining and its applications to intelligence, CRM, and knowledge management / editor, A. Zanasi. http://catalog.loc.gov/cgi-bin/Pwebrecon.cgi?v1=1&ti=1,1&SAB1=data%20mining&BOOL1=all%20of%20these&FLD1=Keyword%20An |
| Lietuvos integrali bibliotekų informacijos sistema: http://www.libis.lt/ | Building <i>data mining</i> applications for CRM / Alex Berson, Stephen Smith, Kurt Thearling, New York : McGraw-Hill, c2000 http://catalog.loc.gov/cgi-bin/Pwebrecon.cgi?v1=2&ti=1,2&SAB1=data%20mining&BOOL1=all%20of%20these&FLD1=Keyword%20An |
| KTU bibliotekos duomenų bazė: http://ktu.library.lt:83/ALEPH | Data mining techniques : for marketing, sales and customer support / M.J.A.Berry, G.Linoff, - New York...[et al.] : Wiley, 1997. - x, 454 p. http://www.libis.lt/owalnb/nbi.bpewww2.FromList?pèpg=1&pèqid=4&pèsch=nbi&pèlan=lt&pètime=071007194233&pèaction=1&pè |
| | Data management strategies : the monthly newsletter on data warehousing, data mining and decision support. - Arlington : Cutter Information Corp., 1997-1999 http://www.libis.lt/owalnb/nbi.bpewww2.FromList?pèpg=1&pèqid=4&pèsch=nbi&pèlan=lt&pètime=071007194233&pèaction=1&pè |
| | Greenberg, Paul. CRM at the speed of light : essential customer strategies for the 21st century / Paul Greenberg http://aleph.library.lt/F/4RJYI9LB3AKMCNSX1P52LBQMKEM81NVR84AUTADSANLXU4PFJL-00439?func=item-global&docèlibrary=K |
| | Principles of data mining / David Hand, Heikki Mannila, Padhraic Smyth Adaptive computation and machine learning, Hand, David. http://aleph.library.lt/F/4RJYI9LB3AKMCNSX1P52LBQMKEM81NVR84AUTADSANLXU4PFJL-01180?func=item-global&docèlibrary=K |

Šaltinis: sudaryta autoriaus

**VILNIAUS UNIVERSITETAS
KAUNO HUMANITARINIS FAKULTETAS
INFORMATIKOS KATEDRA**

**VERSLO INFORMACINIŲ SISTEMŲ MAGISTRANTŪROS PROGRAMOS
MOKSLO TIRIAMOJO DARBO PLANAS**

Magistrantė..... Tel.....

Magistrantūros trukmė nuo 2007 m. iki 2009 m.

TEMA.....

Vadovas.....

(vardas, pavardė, mokslinis laipsnis, pedagoginis vardas, darbovietė, telefono nr.)

Darbo anotacija:

Tikslas:.....

Uždaviniai:.....

Metodai, kuriuos ketinama panaudoti darbe.....

.....Laukiami rezultatai.....

Mokslo-tiriamąjo darbo planas

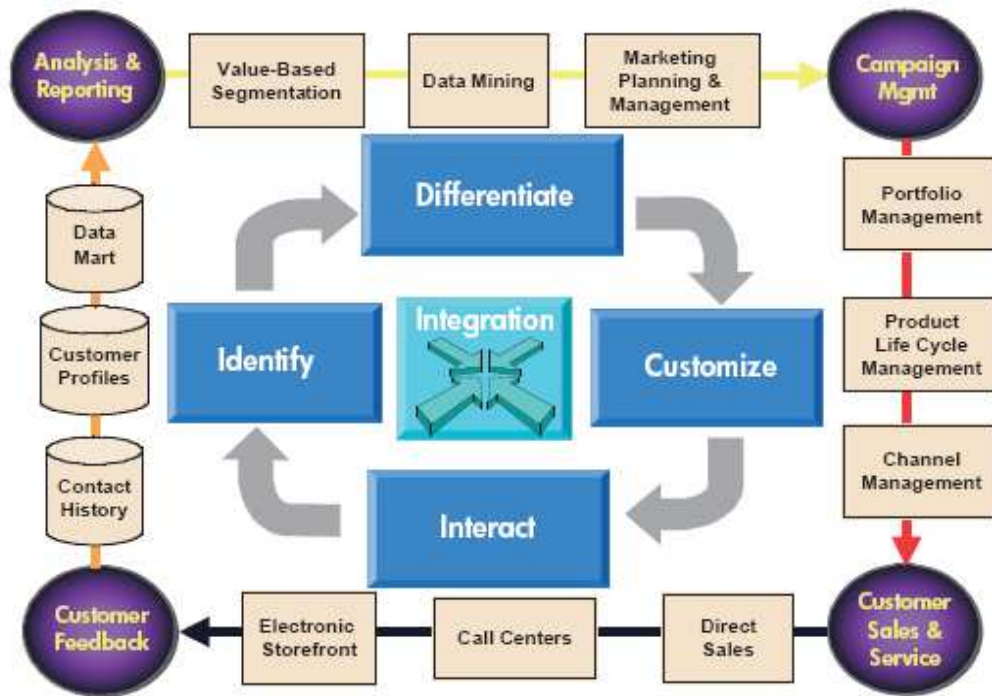
| Semestras | Užduotys | Galutinė atlikimo data |
|-----------|---|--|
| S1 | Magistrinio tema pasirinkimas | 2007 09 13 – 2007 10 04 |
| S1 | Literatūros analizė | 2007 10 05 – 2007 10 19 |
| S1 | Informacijos sisteminimas | 2007 10 20 – 2007 12 23 |
| S1 | Darbo preliminarų plano parengimas; Pirmojo darbo etapo ataskaitos paruošimas; Ataskaitos pateikimas vadovui. | 2007 12 24 – 2008 01 17 |
| S2 | Tinkamiausių metodų atrinkimas; Hipotezių iškėlimas | 2008 01 18 – 2008 03 03 |
| S2 | Teorijos analizė: 1. teorinės medžiagos ruošimas; 2. išsami metodų analizė; 3. jų lyginamoji analizė; 4. preliminarus siūlomo sprendimo skyriaus parengimas; 5. Ataskaitos už antrą etapą paruošimas | 2008 03 04 – 2008 06 29 2008 03 05 – 2008 04 05 2008 04 06 – 2008 04 16 2008 04 17 – 2008 04 28 2008 04 29 – 2008 05 20 2008 05 21 - 2008 06 30 |

| Semestras | Užduotys | Galutinė atlikimo data |
|-----------|--|--|
| S2-S3 | Analitinės dalies rašymas: 1. B2B CRM aspektų tyrimas; 2. CRM modulių analizė; 3. Eksperimentinių tyrimų atlikimas; | 2008 06 30 – 2008 08 12 2008 08 13 – 2008 09 11 2008 09 12 – 2008 10 10 2008 10 11 – 2008 11 10 |
| S3 | Eksperimento projektavimas ir architektūra | 2008 11 11 – 2008 11 26 |
| S3 | Eksperimento rezultatų aprašymas | 2008 11 27 – 2008 12 20 |
| S3-S4 | Dalies išvados | 2008 12 21 – 2009 02 09 |
| S4 | Darbo redagavimas | 2009 02 10 – 2009 04 30 |
| S4 | Darbo pristatymas ir gynimas | 2009 05 01 |

Magistrantė: S.Klikūnaitė Vadovė: doc.D.Krikščiūnienė

CRM PROCESAI IR FUNKCIJOS

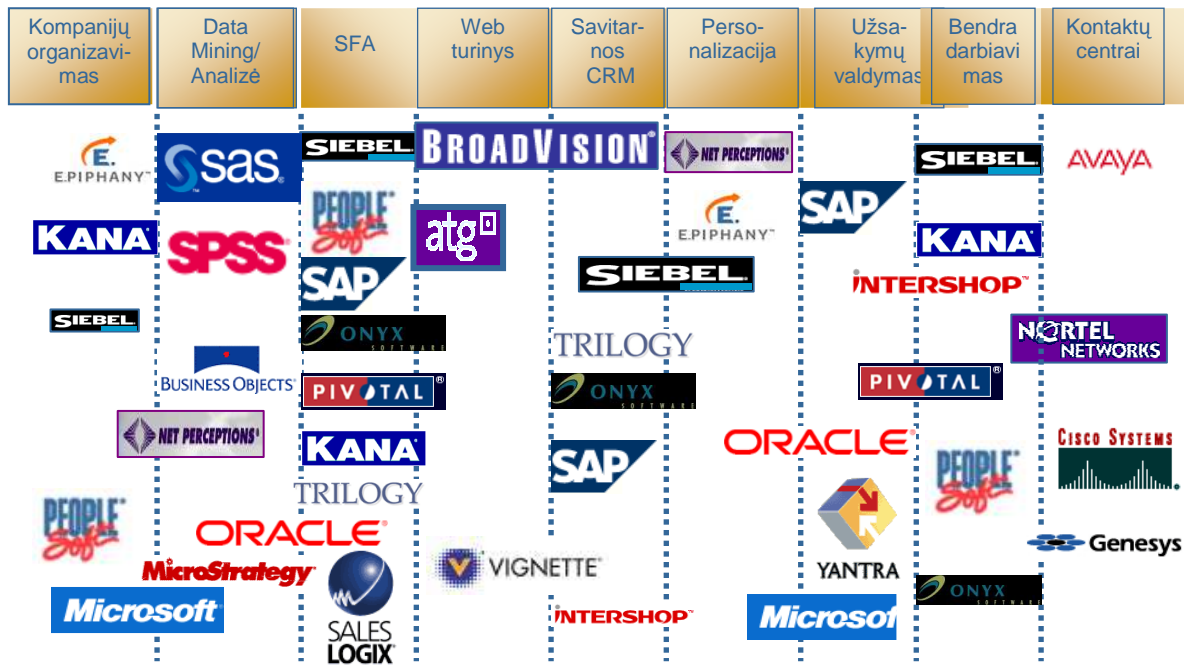
CRM processes and functions



Šaltinis:

20 pav. CRM procesai ir funkcijos

PROGRAMINIŲ PAKETŲ RINKA



Šaltinis:

21 pav. Programinių paketų rinka pagal valdymo lygmenius bei stipriausios kompanijos gamintojos

PROGRAMINĖS ĮRANGOS SU DGM PALYGINIMAS

26 lentelė

Programinės įrangos su DGM palyginimas

| Programinė įranga | Duomenų gavybos taikymas | Naudojamas duomenų gavybos metodas |
|---------------------------------|--|--|
| AutonomyVoiceSuite | Analizuoja didelės informacijos sklaidos turinį, perrašo jį kaip tekstą, jame identifikuoja ir klasifikuoja pagrindines sąvokas, automatizuoja informacijos gavimą per internetą ar kitus skaitmeninius kanalus, pavyzdžiui, mobiliuosius telefonus ir kt. | Klasifikavimo ir indeksavimo priemonė. <u>Priežastinių ryšių ir nuosekliojo eiliškumo metodai</u> |
| Autonomy KnowledgeSTUDIO | Vizualizuota aplinka, teikianti daug galimybių duomenų ir žinių gavybos darbuotojams ir modeliotojams, palaiko visas modelių kūrimo fazes: profiliavimą, tyrimą, modeliavimą, įgyvendinimą, rezultatų įvertinimą, patvirtinimą ir stebėjimą. | |
| AngossKnowledgeSEEKER | Teikia duomenų profiliavimo ir vizualizacijos, sprendimų medžio tipo analizės priemones, leidžia nustatyti kintamųjų ryšius. | <u>Sprendimų medžio metodas</u> |
| Convera RetrievalWare | ŽV sprendimas organizacijoms, kurios intensyviai dirba su informacija. RetrievalWare padeda surasti informaciją bet kokiuose šaltiniuose (portaluose, saugyklose ar failų sistemose), taip pat paversti žiniomis duomenis. RetrievalWare gali naudoti įvairius indeksavimo, sąvokų ir esybių išgavimo, turinio filtravimo metodus. RetrievalWare teikia unikalių ŽV paslaugų – profiliavimo, sąvokų paieškos, automatinės ir dinaminės klasifikacijos, personalizacijos ir bendradarbiavimo, jungia žinių struktūras įvairiomis kalbomis ir pagal dalykinę sritį ir kt. Convera TrueKnowledge for Discovery sistema skirta modernioms verslo organizacijoms sudėtingai paieškai atlikti. Convera RetrievalWare technologija, turi visas profesionalios paieškos priemones. Convera TrueKnowledge for Web skirta informacijos kaupėjams (angl. aggregators), norintiems pateikti savo klientams autoritetinę Web turinį, taip pat B2B – verslininkams, norintiems plėsti savo prijungtinį verslą. | <u>Vaizduojamasis asociacijos metodas</u> |
| Converaba- seRelevance | Dinaminė, integruota semantinių technologijų ir intelektinių priemonių visuma, kuri teikia išsamesnį turinio prasmės supratimą, profesionalią paiešką. BaseRelevance semantinės paieškos komponentai yra: lingvistika, morfologija, tokenizacija (simbolizacija), taksonomija (sistematika), skirstymas į kategorijas, dinaminė klasifikacija. | <u>Metodas nepateiktas.</u> |

5 PRIEDAS (TĘSINYS)

26 lentelės tęsinys

| Programinė įranga | Duomenų gavybos taikymas | Naudojamas DGM |
|---|---|---|
| Oracle Data Mining (ODM) | Įeinantis į Oracle Database 10g Enterprise Edition, suteikia organizacijoms galimybę efektyviai gauti informaciją iš labai didelių DB. ODM teikia duomenų gavybos priemones: klasifikaciją, prognozavimą, klasterizaciją, asociacijas ir kt. Megaputer Intelligence Inc. PolyAnalyst – platus ir universalus duomenų gavybos priemonių rinkinys. PolyAnalyst atlieka visas žinių gavybos operacijas – nuo duomenų importavimo iki ataskaitų pateikimo. | Galimybė naudotis <u>sprendimų medžio, klasterizavimo, priežastinius bei vektorinius duomenų gavybos metodus.</u> |
| Tiberius data mining Tiberius (versija 5.0.1 2007 02 05) | Duomenų modeliavimo ir pavertimo regimą priemonė, veikianti Windows aplinkoje. | <u>K-arčiausio elemento metodas.</u> |
| California Scientific BrainMaker (versija 3.7 Windows XP, 2000 ir Me aplinkai) | Neuroninių tinklų programinė įranga, kuri naudojama versle ir rinkodaros prognozavimui, pavyzdžių atpažinimui, medicinoje diagnozėms nustatyti ir kt. | <u>Neuroninių tinklų metodas.</u> |
| ClearForest Analytics | Gali išgauti aktualią informaciją iš bet kokių organizacijos duomenų ar tekstų. Informacijos pavertimo regimą priemonės taip pat įvairios. | <u>Metodas nepateiktas.</u> |
| Spotfire | Vizualizacijos priemonė DecisionSite teikia supaprastintą prieigą prie duomenų jos informacinėje bibliotekoje. Spotfire Analytics Server leidžia organizacijoms integruoti įvairias verslo analizės priemones: duomenų bazes (Oracle, IBM DB2, Microsoft SQL Server, Microsoft Access, MySQL, SAS/ SHARE, PostgreSQL, Sybase, Informix), vartotojų katalogus (Microsoft Windows Domain, LDAP, Active Directory) ir kt. Spotfire DXP Professional verslo analitikams teikia prieigą prie duomenų ir patogią, interaktyvią ir vizualią analitinio darbo aplinką. SpotfireDXP Enterprise Player papildomai dar teikia darbo eigos (Workflow) įrangą. | <u>Metodas nepateiktas.</u> |
| Advizor Solutions Inc. | ADVIZOR Server teikia interaktyvią verslo analizės aplinką organizacijos ir ekstraneto vartotojams naudotis tik jų naršykle. Recommind, Inc.MindServer lengvai integruojamas į portalus, dokumentų valdymo sistemas, duomenų bazes, darbo eigos programas. Teikia paieškos galimybę, klasifikacijos priemones, prieigą prie archyvų ir kt | <u>Metodas nepateiktas.</u> |
| STATISTICA 7 | Duomenų analizės, vizualizavimo ir modeliavimo priemonė. STATISTICA integruoti matematiniai ir statistiniai algoritmai. | <u>Neuroninių tinklų metodas, sprendimų medžio metodas, analizės gylyn ir kiti metodai.</u> |

Šaltinis: sudaryta autoriaus.

STATISTICA 7 REZULTATAI

27 lentelė

Galutinė klasifikacija

| Cluster members (PagalKlienta_gabenimo_laiko) | | | | | | | | | |
|---|----------------------|-----------|-----------|------------|-----------|------------------|------------|---------------|----------------------|
| Number of clusters: 10 | | | | | | | | | |
| Total number of training cases: 1240 | | | | | | | | | |
| Case No. | Final classification | KlientoID | Vėlavimas | Dienų_viso | Max_dienų | VidutineKrepSuma | IsigytaVnt | ApsilankymuSk | Distance to centroid |
| 1 | 2 | 2 | 0.620000 | 7.00000 | 14.00000 | 863.74 | 808.00 | 226.00 | 0.001082 |
| 2 | 1 | 12 | 0.730000 | 8.00000 | 16.00000 | 522.28 | 11.00 | 8.00 | 0.025939 |
| 3 | 9 | 15 | 0.050000 | 1.00000 | 3.00000 | 241.05 | 385.00 | 126.00 | 0.000443 |
| 4 | 7 | 23 | 0.680000 | 8.00000 | 15.00000 | 217.43 | 197.00 | 103.00 | 0.000912 |
| 5 | 10 | 30 | 0.260000 | 3.00000 | 7.00000 | 234.49 | 54.00 | 7.00 | 0.003771 |
| 6 | 5 | 62 | 0.360000 | 4.00000 | 9.00000 | 224.15 | 557.00 | 157.00 | 0.000832 |
| 7 | 9 | 72 | 0.050000 | 1.00000 | 3.00000 | 236.50 | 240.00 | 65.00 | 1.000438 |
| 8 | 5 | 73 | 0.360000 | 4.00000 | 9.00000 | 252.32 | 307.00 | 133.00 | 1.000823 |
| 9 | 8 | 85 | 0.990000 | 11.00000 | 21.00000 | 302.22 | 7.00 | 6.00 | 0.028302 |
| 10 | 7 | 91 | 0.730000 | 8.00000 | 16.00000 | 233.90 | 288.00 | 75.00 | 1.000598 |
| 11 | 5 | 92 | 0.310000 | 4.00000 | 8.00000 | 107.71 | 389.00 | 173.00 | 1.000938 |
| 12 | 8 | 98 | 0.990000 | 11.00000 | 21.00000 | 460.42 | 8.00 | 7.00 | 1.028314 |
| 13 | 1 | 100 | 0.880000 | 10.00000 | 19.00000 | 428.85 | 34.00 | 17.00 | 1.003005 |
| 14 | 3 | 101 | 0.470000 | 6.00000 | 11.00000 | 706.28 | 201.00 | 37.00 | 0.000899 |
| 15 | 8 | 132 | 1.250000 | 13.00000 | 26.00000 | 91.10 | 57.00 | 24.00 | 1.019594 |
| 16 | 9 | 144 | 0.100000 | 2.00000 | 4.00000 | 215.78 | 188.00 | 94.00 | 1.005269 |
| 17 | 1 | 148 | 0.830000 | 9.00000 | 18.00000 | 184.40 | 126.00 | 68.00 | 1.001311 |
| 18 | 7 | 156 | 0.730000 | 8.00000 | 16.00000 | 171.11 | 174.00 | 31.00 | 1.000621 |
| 19 | 2 | 165 | 0.620000 | 7.00000 | 14.00000 | 922.71 | 206.00 | 7.00 | 1.001025 |
| 20 | 8 | 167 | 1.250000 | 13.00000 | 26.00000 | 231.84 | 47.00 | 15.00 | 1.019538 |

Šaltinis: sudaryta autoriaus

28 lentelė

Analizės rezultatų langas

| Summary for k-means clustering (PagalKlienta_gabenimo_laiko) | |
|--|-----------------------------|
| Number of clusters: 10 | |
| Total number of training cases: 1240 | |
| Algorithm | k-Means |
| Distance method | Squared Euclidean distances |
| Initial centers | Random N observations |
| MD casewise deletion | Yes |
| Cross-validation | No |
| Testing sample | 0 |
| Training cases | 1240 |
| Training error | 1.000949 |
| Number of clusters | 10 |

Šaltinis: sudaryta autoriaus

6 PRIEDAS (TĘSINYS)

29 lentelė

Atstumai tarp sudarytų grupių

| | | Distance between centroids of k-means clustering (PagalKlienta_gabenimo_laiko) | | | | | | | | |
|------------|-----------|--|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| | | Number of clusters: 10 | | | | | | | | |
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 | Cluster 8 | Cluster 9 | Cluster 10 |
| Cluster 1 | 0.000000 | 1.100488 | 1.235584 | 1.668219 | 1.437800 | 1.470665 | 1.032425 | 1.124547 | 2.032566 | 1.686371 |
| Cluster 2 | 1.100488 | 0.000000 | 1.028421 | 1.553578 | 1.119347 | 1.476221 | 1.018806 | 1.448743 | 1.489219 | 1.262279 |
| Cluster 3 | 1.235584 | 1.028421 | 0.000000 | 1.570071 | 1.031542 | 1.569338 | 1.093245 | 1.702609 | 1.281914 | 1.118167 |
| Cluster 4 | 1.668219 | 1.553578 | 1.570071 | 0.000000 | 1.642391 | 2.050631 | 1.580750 | 2.041354 | 1.994629 | 1.778057 |
| Cluster 5 | 1.437800 | 1.119347 | 1.031542 | 1.642391 | 0.000000 | 1.710137 | 1.232162 | 2.028868 | 1.125896 | 1.027912 |
| Cluster 6 | 1.470665 | 1.476221 | 1.569338 | 2.050631 | 1.710137 | 0.000000 | 1.452283 | 1.702598 | 2.208820 | 1.921132 |
| Cluster 7 | 1.032425 | 1.018806 | 1.093245 | 1.580750 | 1.232162 | 1.452283 | 0.000000 | 1.283958 | 1.699185 | 1.420637 |
| Cluster 8 | 1.124547 | 1.448743 | 1.702609 | 2.041354 | 2.028868 | 1.702598 | 1.283958 | 0.000000 | 2.873952 | 2.395003 |
| Cluster 9 | 2.032566 | 1.489219 | 1.281914 | 1.994629 | 1.125896 | 2.208820 | 1.699185 | 2.873952 | 0.000000 | 1.035428 |
| Cluster 10 | 1.686371 | 1.262279 | 1.118167 | 1.778057 | 1.027912 | 1.921132 | 1.420637 | 2.395003 | 1.035428 | 0.000000 |

Šaltinis: sudaryta autoriaus

30 lentelė

Grupių parametrai

| | | Centroids for k-means clustering (PagalKlienta_gabenimo_laiko) | | | | | | | | |
|---------|-----------|--|-----------|------------------|------------|---------------|-----------------|---------------|--|--|
| | | Number of clusters: 10 | | | | | | | | |
| | | Total number of training cases: 1240 | | | | | | | | |
| Cluster | Vėlavimas | Dienų viso | Max dienų | VidutineKrepSuma | IsigytaVnt | ApsilankymuSk | Number of cases | Percentage(%) | | |
| 1 | 0.847112 | 9.41711 | 18.34225 | 434.34 | 439.11 | 103.57 | 187 | 15.08065 | | |
| 2 | 0.593437 | 7.00000 | 13.46875 | 479.12 | 143.70 | 41.41 | 96 | 7.74194 | | |
| 3 | 0.465033 | 5.62914 | 10.90066 | 316.28 | 166.70 | 57.42 | 151 | 12.17742 | | |
| 4 | 0.573333 | 6.66667 | 13.00000 | 160.20 | 39936.00 | 17803.00 | 3 | 0.24194 | | |
| 5 | 0.335000 | 4.00000 | 8.50000 | 545.86 | 562.90 | 86.23 | 94 | 7.58065 | | |
| 6 | 0.727500 | 8.37500 | 16.00000 | 20539.19 | 489.63 | 5.00 | 8 | 0.64516 | | |
| 7 | 0.707723 | 8.00000 | 15.55446 | 391.60 | 145.90 | 54.12 | 101 | 8.14516 | | |
| 8 | 1.129719 | 12.11930 | 23.74386 | 346.11 | 373.36 | 64.63 | 285 | 22.98387 | | |
| 9 | 0.054138 | 1.28276 | 3.00000 | 368.84 | 254.67 | 63.56 | 145 | 11.69355 | | |
| 10 | 0.207941 | 2.64118 | 5.95882 | 295.53 | 321.16 | 86.45 | 170 | 13.70968 | | |

Šaltinis: sudaryta autoriaus

31 lentelė

Kiekio tyrimo statistinė informacija

| | | Descriptive Statistics (Proгноzes) | | | | | | | |
|----------|---------|------------------------------------|----------|----------|-------------------|----------|----------|----------|----------|
| Variable | Valid N | Mean | Median | Mode | Frequency of Mode | Sum | Minimum | Maximum | Std.Dev. |
| Kiekis02 | 14093 | 37.22096 | 5.00000 | 1.000000 | 2887 | 524555.0 | 1.000000 | 8424.000 | 191.2591 |
| Prognose | 14093 | 42.26505 | 16.50893 | 0.000000 | 2583 | 595641.3 | 0.000000 | 3481.632 | 127.7138 |
| Kiekis03 | 14093 | 25.93075 | 2.00000 | 1.000000 | 3790 | 365442.0 | 0.000000 | 8915.000 | 166.9365 |

Šaltinis: sudaryta autoriaus.