

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
KOMPIUTERIŲ KATEDRA

Laimonas Selenis

Vartotojų sąryšio informacijos valdymo sistema

Magistro darbas

Darbo vadovas:
doc. dr. E. Kazanavičius

Kaunas, 2004

KAUNO TECHNOLOGIJOS UNIVERSITETAS

INFORMATIKOS FAKULTETAS

KOMPIUTERIŲ KATEDRA

TVIRTINU

Katedros vedėjas

doc. dr. E. Kazanavičius

2004 05

Vartotojų sąryšio informacijos valdymo sistema

Informatikos mokslo magistro baigiamasis darbas

Kalbos konsultantė

Lietuvių kalbos katedros lektorė

dr. J. Mikelionienė

2004 05

Recenzentas

doc. S. Maciulevičius

2004 05

Vadovas

doc. dr. E. Kazanavičius

2004 05

Atliko

IFM-8/1 gr. stud.

Laimonas Selenis

2004 05

Kaunas, 2004

Turinys

Įvadas	4
1. Problemos analizė	5
1.1. Tyrimų metodologija	6
1.2. Informacijos valdymas ir vartojamumo strategija	9
1.3. Veiklos planavimas	10
1.3.1. Vartotojų informacijos valdymo verslo plano kūrimo procesas	11
1.3.2. Išteklių planavimo procesas	11
1.3.3. Duomenų kokybės įvertinimo procesas	11
1.3.4. Vartotojo poreikių nustatymo procesas	13
1.3.5. Praeities įvykių kaupimo procesas	13
1.3.6. Kriterijų įvertinimo procesas	13
1.3.7. Vartotojų atpažinimo procesas	14
1.3.8. Tarpininkavimo įvertinimo procesas	15
1.3.9. Privatumo įvertinimo procesas	15
1.3.10. Darbų perdavimo galimybių nustatymo procesas	15
1.4. Įgyvendinimo priemonių analizė	16
1.4.1. Pagrindiniai veiklos automatizavimo įrankiai	17
2. Teorinis problemos sprendimas	19
2.1. Duomenų gavybos metodai	19
2.1.1. Artimiausių kaimynų metodas	21
2.1.2. Sprendimų medžiai	23
2.1.3. Asociatyvių taisyklių metodas	29
2.1.4. Duomenų gavybos metodų įvertinimas	32
2.2. Vartotojų atpažinimo sistemos modelis	33
2.2.1. Vartotojų profilių sudarymas	33
2.2.2. Konceptualusis vartotojų atpažinimo sistemos modelio aprašymas	35
2.2.3. Duomenų perkėlimo į saugyklą procesas	38
2.2.4. Vartotojų profilių kūrimo procesas	39

3.	Vartotojų atpažinimo sistemos modelio eksperimentinis tyrimas	42
3.1.	Tiriamųjų vartotojų elgsenos duomenų struktūra	42
3.2.	Eksperimento atlikimo eiga ir rezultatai	45
	Išvados	49
	Summary	50
	Literatūra.....	51

Ivadas

Vartotojų sąryšio informacijos valdymas – VSIV (*Customer Relationship Management - CRM*) yra viena iš svarbesnių sričių kiekvienai organizacijai. Kiekviena organizacija turi nemažai tobulintinų vartotojų sąryšio valdymo sričių. Nepaisant daugelio konsultavimo kompanijų teigimo, jog daug VSIV sistemų neduoda organizacijoms apčiuopiamos naudos, tokių sistemų paklausa išlieka gana didelė ir daugelis įstaigų daug investuoja į tokių sistemų kūrimą bei plėtojimą.

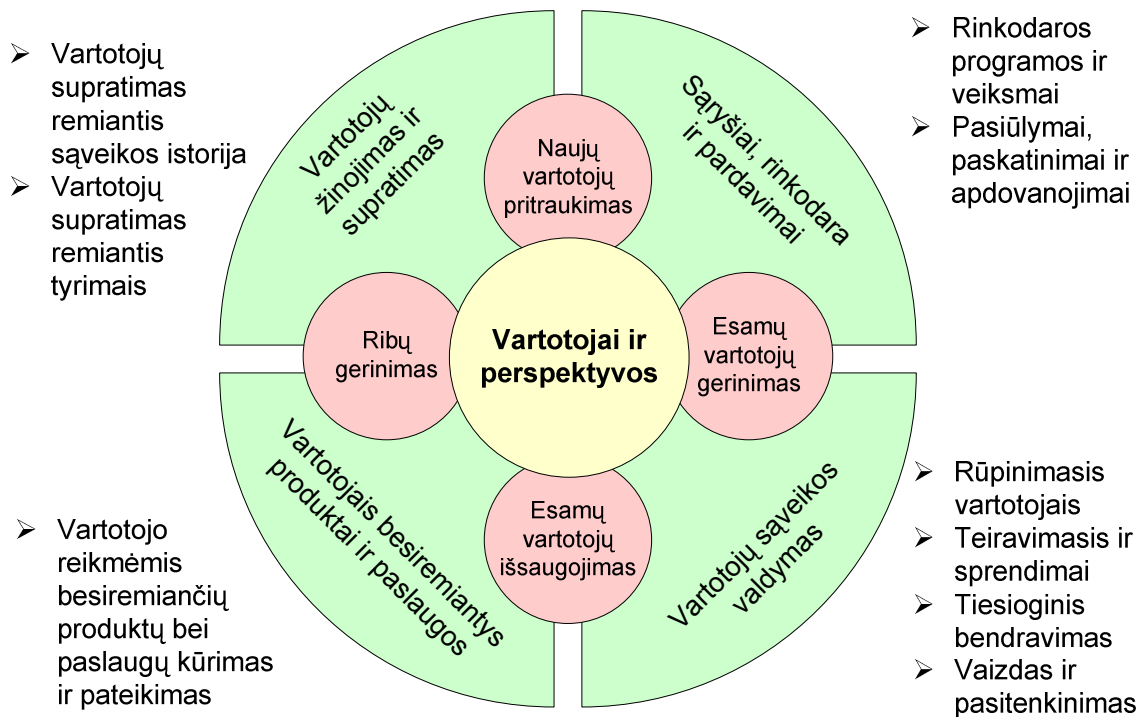
Vienas iš svarbiausių išbandymų kuriant VSIV sistemas yra didelių vartotojų informacijos kiekių valdymas. Suprantama, kad norint sukurti ir išlaikyti ryšius su vartotojais ar klientais, organizacijos veikla turi prasidėti nuo kruopštaus bei detalaus požiūrio į kiekvieną vartotoją. Tai daroma tam, kad sukurtas tiesioginis ryšys su kiekvienu vartotoju būtų efektyvus ir abipusiai naudingas. Todėl didelio vartotojų sąryšio informacijos kiekio efektyvus valdymas ir jos pritaikymas naudojimui yra ypač svarbus kuriant VSIV sistemas bei užtikrinant sėkmingą jų veiklą.

Be to, remiantis tam tikrų nesėkmes patyrusių VSIV projektų tyrimais buvo nustatyta, kad pagrindinė priežastis yra vartotojų informacijos ignoravimas. Siekdamas suprasti VSIV ir investicijų gražos (*return on investment*) įtaką, kompanijos per pastaruosius 5 metus sėkmingai kuria vartotojų duomenų bazių sistemas. Kitas svarbus aspektas, darantis įtaką organizacijai, yra gebėjimas tiksliai atpažinti jos klientus tam, kad būtų sukurtas tinkamas požiūris analitiniais ir veiksmingais pagrindais. Gebėjimas atpažinti geriausius vartotojus, grupuoti jų poreikių požymius, vartoti juos klientų išlaikymui ir naujų pritraukimui, reikalauja iš organizacijos, kad ji pradėtų atidų kiekvieno iš jos vartotojų analizavimą ir atpažinimą.

Gebėjimas kruopščiai įvertinti vartotojų požymius teikia galimybes išvengti duomenų susiskaldymo problemos, suprasti kiekvieno vartotojo elgesį, ypatybes ir rizikas. Toks atidus požiūris, remiantis pilna vartotojų informacija, leidžia sukurti pačius tinkamiausius ir savalaikius pasiūlymus ir juos pateikti tokiu laiku, kad ryšys su vartotoju duotų patį didžiausią įmanomą efektyvumą.

1. Problemos analizė

Problema, su kuria susiduria daugelis organizacijų šiandien – nesugebėjimas tiksliai identifikuoti vartotojo [1]. Vartotojo atpažinimas gali būti apibrėžtas kaip gebėjimas tiksliai identifikuoti vartotojų grupes pagal jų veiklos pobūdį. Remiantis informacijos atpažinimu, sujungti visus vidinius vartotojo požymius ir užuominas kaip vientisą informaciją. Vartotojų identifikavimas apima duomenų modelius, realaus laiko atpažinimo komponentus, aukšto lygio duomenų apdorojimo komponentus ir sąsajas. Dėl šito atpažinimo proceso organizacija gauna: 1) pradinę infrastruktūrą testinam vartotojų aspektų, požymių, užuominų sinchronizavimui ir 2) greitą vartotojų atpažinimo ir nesulyginamos klientų informacijos sujungimo galimybę.



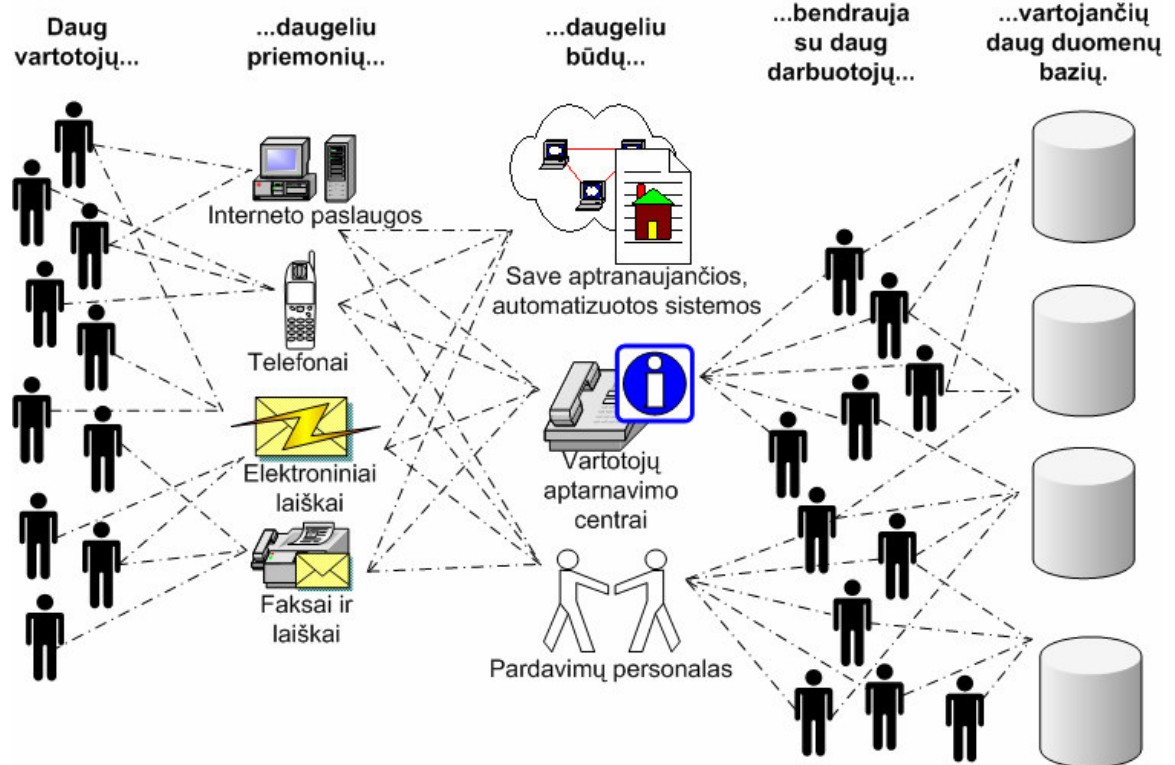
1 pav. Vartotojų informacijos svarba organizacijos veiklai

Neteisingas vartotojų identifikavimas dėl prastos duomenų kokybės, nepakankamų ir netikslių duomenų, riboto jų panaudojimo veda prie neigiamų pasekmių organizacijoje, kas mažina jos veiklos efektyvumą. Pasinaudojant tokia informacija kontaktavimui su klientais yra kenkiama pačiam organizacijos – vartotojo ryšiui, kadangi tam tikri organizacijos pasiūlymai gali neatitikti realių jo poreikių.

Tokias ir panašias problemas sukeliančių priežasčių yra keletas. Tai:

- Apribota priežiūra dėl daugelio sistemų ir nesugebėjimas visko sujungti į visumą.

- Nesugebėjimas apibrėžti investavimo gražos iš kokybės įvertinimo ir tobulinimo projektų.
- VSIV sistemos tobulinimų neįgyvendinimas remiantis praeitų metų specifikacija.
- Susitarimų tarp rinkodaros, pardavimų ir vartotojų paslaugų komandų dėl bendro artėjimo link VSIV stoka. Taktinių vietoj strateginių patobulinimų forsavimas.
- Vidinių žinių, reikalingų VSIV sistemoms tobulinti, stoka.



2 pav. Vartotojų sąryšio informacijos daugialypumas ir susiskaldymas

Vartotojų lojalumas ir pasitikėjimas organizacija dažnai yra griaunami dėl to, kad nuolat didėja duomenų rinkimo mastai, tačiau daugelis organizacijų neturi realaus plano, kaip toji informacija bus panaudota organizacijos ir pačių jos klientų gerovei, arba nesugeba šių planų realizuoti, dėl iškylančių su tuo susijusių problemų.

1.1. Tyrimų metodologija

Šiandien VSIV sistemų reikalingumas praktiškai yra nenuginčijamas. Atlikta daug tyrimų kurie patvirtina, jog tokio tipo sistemų poreikis yra tikrai realus ir tokių sistemų įdiegimas bei vartojimas yra neatskiriama sėkmingos organizacijos veiklos dalis.

2001 metų pabaigoje kompanijos Acxiom, QCi kartu su Northwestern universitetu atliko tyrimą, kodėl daugelis organizacijų patiria didelius sunkumus, susijusius su jų klientų duomenų administravimu ir panaudojimu [1]. Tyrimams buvo panaudotas vartotojų administravimo įvertinimo įrankis (Customer Management Assessment Tool – CMAT™), sukurtas kompanijos QCi, pagrįstas QCi vartotojų administravimo modeliu. Modelis susideda iš aštuonių skirtingų VSIV dalių:

- **Analizė ir planavimas:** organizacijos turi suvokti savo klientų pagrindus, jų įtaką organizacijos veiklai ir sugebėti apibrėžti kiekvieno iš klientų vertę kombinuojant pardavimų ir rinkodaros kainas, valdymo kaštus, strategijas ir paslaugas. Šitoks išvalgumas sudaro pagrindą visai klientų administravimo veiklai.
- **Pasiūlymas:** pasiūlymai turi atspindėti reikmes, formaliai apibrėžtas analizės ir planavimo dalyje. Tada pasiūlymai turi būti tiesiogiai perduoti darbuotojams, kurie bendrauja su klientais, bei patiems klientams.
- **Žmonės ir organizacija:** klientų valdymo kompetencijos turi būti aiškiai apibrėžtos visose darbų rūšyse ir sustiprintos taip, kad sugebėtų apibrėžti bet kokius trūkumus. Informacijos privatumo politika – vienas iš svarbiausių organizacijos įrankių.
- **Informacija ir technologija:** visos organizacijos privalo turėti išsamų ir prieinamą klientų informacijos planą ir technologijų infrastruktūrą, kuri apima informacijos vertę, įgijimo prioritetus, informacijos valdymą ir panaudojimą, kontrolės reikalavimus, įskaitant ir pirmumų valdymą.
- **Procesų valdymas:** visi klientų valdymo procesai turi būti reguliariai ir formaliai tikrinami iš abiejų – organizacijos ir kliento perspektyvų, tam, kad užtikrinti abipusį priimtinumą.
- **Klientų valdymo aktyvumas:** kliento vertė turėtų būti apibrėžta pardavimų, rinkodaros, paslaugų prasmėmis. Įgijimas, išvalgumas, išlaikymas ir klientų aptarnavimo efektyvumas – pagrindinės kryptys, į kurias reiktų atkreipti dėmesį.
- **Poveikio apibrėžimas:** visos rinkodaros veiklos turėtų būti aprašytos. Bendraujantys su vartotojais darbuotojai privalo būti patyrę darbe su klientais, jų pritraukimo, išlaikymo ir plėtros srityse.

- **Klientų patirtis:** kiekviena organizacija turi suvokti savo poziciją klientų valdyme pagrindinių konkurentų atžvilgiu. Įvykiais besiremiantys tyrimai turi būti vartojami visuose visiems klientų valdymo etapuose: esamų klientų valdyme ir naujų klientų pritraukime.

Šiam tyrimui buvo naudojamas CMAT tyrimų atestavimas, kurį sudaro dviejų valandų trukmės interviu su vienu ar daugiau organizacijos darbuotojų, atsakingų už klientų valdymą arba rinkodarą. Tyrimo metu didžiausias dėmesys buvo skiriamas klientų duomenų integralumui, taipogi informacijos valdymui ir panaudojimui klientų valdyme. Į tyrimą buvo įtrauktas nemažas kiekis didesniųjų JAV organizacijų, veikiančių finansų, technologijų, draudimo, telekomunikacijų, automobilių pramonės ir leidybos srityse. Surinkus, išanalizavus bei suskaičiavus visus gautus atsakymus interviu metu, buvo nustatyta, kad net 46 proc. geriausių organizacijų veiklos atvejų būtų neįmanomi be reikšmingo priėjimo prie jų vartotojų duomenų. Taipogi nemažas procentų likusių organizacijų tiesiogiai nenaudoja klientų duomenų bazių, tačiau tam tikra forma yra paveikiamos panašaus ar netgi to paties tipo informacijos.

1.2. Informacijos valdymas ir vartojamumo strategija

Tam, kad kuo tiksliau apibūdinti informacijos valdymo sritis, kurių naudojimas duotų organizacijai tam tikrą naudą, galima apibrėžti tokias analizės sritis:

- **Vizija ir strategija** – tai bendras vaizdas apie informaciją, kuri turi naudą organizacijos veikloje.
- **Investicijos** – nustatyti sritis, į kurias investuojama ir apibrėžti, ar investuojama į tas sritis naudingai, kad būtų pasiekiamas maksimalus efektyvumas. Pilnai suvokti investicijų gražos prasmę ir įtaką.
- **Ištekliai** – įvertinti reikalingų darbuotojų skaičių ir kvalifikacinį tinkamumą darbui organizacijos veiklos srityse. Ištirti kaip gerai dirba skirtingos darbuotojų grupės kartu.
- **Informacijos turinys** – nustatyti kokia informacija yra pasiekama galutiniams vartotojams ir koks yra jos naudingumas.
- **Informacijos panaudojimas** – įvertinti informacijos pasiekiamumą tinkamiems asmenims ir apibrėžti kiek efektyviai ji panaudojama srityse, kur toji informacija turi potencialą.
- **Informacijos valdymas** – nustatyti ar informacija yra valdoma kaip pagrindinis verslo turtas. Kaip informacija yra renkama, tikrinama, atnaujinama ir platinama tinkamu laiku reikalingoms sritims.
- **Technologinis palaikymas** – Įvertinti koks yra informacijos valdymo aprūpinimas technologinėmis priemonėmis. Nustatyti tokio aprūpinimo reikmes ir galimybes ateityje.
- **Integralumas.** Informacijos valdymas ir jos panaudojimas tinkamai papildo vienas kitą kaip viena bendra veiklų sritis. Rasti ar egzistuoja tam tikros skirtingų veiklų sritys, kurios konfliktuoja tarpusavyje, tuo sukeldamos neefektyvumą ir išlaidas bendrai organizacijos veiklai.

Šios sritys atskirai gali būti naudingos izoliacijos atveju, tačiau pats efektyvumas pasiekiamas tinkamai bendrai jas panaudojant. Organizacija gali pasiekti itin efektyvią veiklą vienoje ar keletoje sričių, tačiau nesistengiant apimti naudą, kurią teikia visos sritys, galiausiai toks bendras veiklos efektyvumas gali būti labai žemas. Organizacija turi stengtis ne tik kuo geriau įgyvendinti, apibrėžti ir atsakyti į kiekvienos iš šių analizės sričių reikalavimus, tačiau

nepamiršti ir bendro šių sričių integralumo – geriausias rezultatas pasiekiamas visas šias sritis panaudojant kartu.

1.3. Veiklos planavimas

Atsižvelgiant į anksčiau paminėtus tyrimo rezultatus bei analizės etapus galima apibrėžti eilę įgyvendinimo sričių – etapų, kurie gali padėti organizacijai sukurti bendrą veiklos planą, skirtą jos klientų informacijai valdyti ir panaudoti. Šios sritys apima trumpalaikius, vidutinius ir ilgalaikius veiklos kriterijus, kuriuos organizacija turėtų įgyvendinti norėdama pasiruošti savo turimą bei žadamą įgyti klientų informaciją tolimesniam ir efektyvesniam jos vartotojų valdymui ir vartojimui. Iš viso galima apibrėžti 10 kriterijų, kurių įgyvendinimas turi įtakos sėkmingam organizacijos veiklos procesui:

1. Sukurti vartotojų informacijos valdymo verslo planą.
2. Įsigyti reikiamus išteklius apibrėžtam planui įgyvendinti.
3. Sukurti duomenų kokybės įverčius.
4. Apibrėžti bendrą vartotojo vaizdą.
5. Kaupti praeities informaciją
6. Remiantis klientų informacija stengtis suprasti kliento vertę, egzistavimo laikotarpį, pirmenybės ir išsaugojimo kriterijus.
7. Sukurti klientų infrastruktūrą, pasižyminčią vartotojų atpažinimu ir priėmimu.
8. Trečių asmenų įtaka duomenims (tarpininkavimas).
9. Privatumo supratimas.
10. Apsvarstyti darbų perdavimo išorinėms kompanijoms galimybes.

1.3.1. Vartotojų informacijos valdymo verslo plano kūrimo procesas

Pagrindinis tikslas – užtikrinti, kad egzistuoja organizacijos verslo planas klientų informacijai tvarkyti ir vartoti. Visi pagrindiniai kriterijai, darantys įtaką organizacijos veiklai – rinkodara, pardavimai, klientų aptarnavimas, finansai, eksploatavimas bei IT – turi atsispindėti bendrame veiklos plane. Organizacijos išleidžia milžiniškas pinigų sumas savo vartotojų informacijos kaupimui, valdymui ir efektyviam panaudojimui, ir tokios duomenų bazės dažnai yra laikomos vienomis iš didžiausių organizacijos turtų. Išsamus pirminis veiklos planas turėtų padėti apibrėžti kaip toji kaupiama informacija gali ir turi būti pritaikyta praktikoje. Tyrimai byloja, kad didžioji dauguma organizacijų susiduria su problemomis tada, kai bando įdiegti tokį veiklos planą. Dažnas atvejis, kai atskiros plano dalys egzistuoja skirtinguose departamentuose, tačiau jos yra skirtos būtent tiems skyriams, tačiau visumoje jie nesudaro esminės organizacijos veiklos vizijos. Ekspertai teigia, kad kaip ir daugelyje sričių, pirma reikia apibrėžti savo veiklos tikslus ir metodus, kaip jie bus pasiekiami. Niekas nestato namo prieš tai neparuošę detalios specifikacijos popieriuje, tačiau kol kas dažnas atvejis, kai kompanijos imasi įdiegti VSIV sistemas, neturėdamos didesnio supratimo kaip tai galės pritaikyti tolimesnėje ateityje.

1.3.2. Išteklių planavimo procesas

Įsitikinkite, kad jūsų turimi resursai yra pakankami numatytam klientų informacijos valdymui ir panaudojimui įgyvendinti. Reikalavimuose apibrėžkite ir tai, kad tai speciali sritis, kur galimas specialistų trūkumas. Apsvarstykite strateginę dalinių darbų perdavimo galimybę. Vartotojų informacijos valdymas ir vartojimas reikalauja aukštos kvalifikacijos specialistų bei daug žinių. Dažnas atvejis, kai atvira rinka negali pasiūlyti reikiamo tokių specialistų kiekio. Tokių resursų trūkumas VSIV diegimo eigoje gali skaudžiai atsiliiepti organizacijos veiklai. Viena VSIV sistemų žlugimo priežasčių yra nesugebėjimas ar nenoras atkreipti dėmesį į žinias. Jokie pasaulio pinigai negali išgelbėti VSIV projekto, jeigu jo įdiegimo eigoje su sistema pradės dirbti neprofesionalūs bei neapmokyti darbuotojai. Organizacijos neturėtų pamiršti apmokyti darbuotojų, kaip reikia dirbti su VSIV sistemomis bei įrankiais, kad jie galėtų kuo efektyviau bendrauti su klientais.

1.3.3. Duomenų kokybės įvertinimo procesas

Sukurkite stiprią kokybės kriterijų sistemą. Jeigu reikia imkitės koreguojamųjų veiksmų, nevenkite paskatinimų ar sankcijų jūsų darbuotojams, kurie bendrauja su klientais. Privalu

suprasti kokybės informacijos reikšmę. Organizacija turi sugebėti apibrėžti patiriamą naudą ar nuostolius, susijusius su turimų duomenų kokybe. Duomenų kokybės įverčiai turi apimti tokius aspektus kaip išlaidos, susijusios su blogu duomenų rinkimu, nuostoliai, patirti atliekant pakartotinį darbą. Darbuotojai, turintys tiesioginių ryšių su klientais, turi būti ypač aukštos kvalifikacijos, kadangi tai tiesioginis jūsų informacijos kaupimo šaltinis – jus pasiekiantys duomenys pirmiausia yra apdorojami būtent klientų aptarnavimo personalo, todėl jų nepastebėtos klaidos ar netikslumai turi ypač didelę reikšmę. Todėl reikia stengtis kiek įmanoma daugiau mažinti tokių klaidų tikimybę.

Duomenų kokybę galima įvertinti pagal:

- **Duomenų išbaigtumą** – tai santykis visų įmanomų duomenų šaltinių ir duomenų šaltinių kiekio, kuri organizacija įtraukė į savo VSIV sistemos veiklą.
- **Duomenų tikslumą** – tai bendras duomenų tikslumas, apibrėžiantis kiekvieno kliento kontaktinę ir kitokio pobūdžio informaciją. Pagrindinis klausimas – kiek tiksli yra informacija, turima apie kiekvieną iš organizacijos klientų.
- **Grupavimo tikslumą** – tai kriterijus, kuriuo remiantis organizacija gali sujungti į visumą informaciją, gautą iš skirtingų šaltinių. Kai organizacija paima reikiamą informacijos kiekį iš visų šaltinių ir atlieka pagrindines informacijos tikslumo užtikrinimo operacijas, sekantis svarbus procesas yra – kaip visa tai sujungti į visumą, kad tiksliau apibrėžti kiekvieno iš klientų bendrą vaizdą.
- **Duomenų prieinamumą** – tai duomenų pateikimo greitis, per kurį organizacija gali apdoroti turimą informaciją ir pateikti tinkama vartojimui forma sprendimų priėmimo personalui, programinei įrangai ir kitiems su tuo susijusiems darbuotojams.

Organizacija savo duomenų kokybę valdo naudojantis eile procesų, dažnai bendrai vadinamų – klientų duomenų integracija. Klientų duomenų integracijos eigoje yra sukuriamas išsamaus kliento portretas, kuris prireikus gali būti prieinamas bet koku laiku ir bet kokioje organizacijos vietoje reikiamų verslo sprendimų priėmimui. Integracijos procesas apima daugelį sudėtingų duomenų valdymo uždavinių ir sukuria vieną iš pagrindinių sėkmingos VSIV sistemų veiklos komponentų – kliento suvokimą.

1.3.4. Vartotojo poreikių nustatymo procesas

Tam, kad organizacija galėtų sukurti kuo efektyvesnius pasiūlymus jai reikia turėti kuo detalesnį jos klientų suvokimą – turėti bendrą vartotojo vaizdą. Kaip buvo minėta anksčiau, tam tikras vartotojo suvokimas jau yra formuojamas duomenų kokybės užtikrinimo etape, tačiau vienas būdas negali būti pats efektyviausias, todėl geriausi rezultatai yra pasiekiami bendrai vartojant keletą metodų kartu. Galima paminėti keletą duomenų analizės tipų, kuriuos pasitelkus apibrėžiamas dar detalesnis organizacijos kliento portretas. Tikslus klientų identifikavimas pasiekiamas analizuojant šių tipų duomenis:

- Atpažinimo duomenis – vardai, pavardės, kontaktinė informacija
- Demografinius duomenis – populiacija, vietovė, gamtinės sąlygos
- Klientų sąveikos duomenis – kas, kur, kaip, kada, kodėl...
- Rinkodaros bandymus bei programas – pateikti pasiūlymai, akcijos
- Rizikos faktorius – patirtos nesėkmės, problemos
- Pelningumą, rentabilumą – bendravimo laikotarpio įvertinimas, kliento „gyvavimo“ trukmės įvertinimas
- Išorinė informaciją – rinkodaros informacija, charakteristikos, finansai, kreditingumas.

Gautas bendras kliento vaizdas yra gana kintantis bei skirtingas kiekvienai organizacijai, todėl būtinas tam tikras periodišką analizės vykdymas, tam kad informacija nuolatos būtų atnaujinama bei kliento portretas kiek galima tiksliau atitiktų realybę.

1.3.5. Praeities įvykių kaupimo procesas

Organizacijos veiklai visuomet yra naudingas praeities informacijos analizavimas. Siūloma saugoti mažiausiai trijų metų senumo informaciją. Trys metai atspindi maždaug du ar daugiau pardavimo ciklų. Tokios informacijos analizavimas gali duoti nemažai naudos, gauti duomenys ir tendencijos gali būti vartojami tolimesnėje veikloje.

1.3.6. Kriterijų įvertinimo procesas

Sėkmingas informacijos apie klientus kaupimas ir vartojimas naujiems klientams pritraukti nėra vienintelis kriterijus, užtikrinantis sėkmingą veiklą. Ne mažiau svarbus yra tos informacijos pasitelkimas ryšiams su esamais klientais gerinti. Organizacijai visuomet naudingas

objektyvus vieno kliento vertės suvokimas. Tai leidžia paprasčiau nustatyti kliento gyvavimo ciklo vertę, pirmumo ir tinkamumo informacijos svarbą ir išlaikymo aktyvumą.

Kliento vertė. Sėkmingai savo veiklą plėtojanti organizacija turėtų sugebėti apibrėžti vieno kliento vertę jos atžvilgiu. Žinant, kokią įtaką daro vienas klientas, kur kas lengviau nustatyti pardavimų apimtį, valdymo kaštus, paslaugų teikimo galimybes. Turėdama tokio tipo informaciją, organizacija gali priimti rimtus sprendimus, susijusius su rinkos aktyvumu.

Gyvavimo trukmės vertė. Tam kad pasiekti maksimalų abipusį naudos koeficientą – organizacija turi aiškiai suvokti potencialią jos esamų ir būsimų klientų gyvavimo trukmę ir panaudoti ją rinkos biudžetų ir prioritetų skirstymui. Realus šio kriterijaus įvertis turėtų atsispindėti „kainoje už paslaugą ar prekę“. Tokio įverčio nustatymas yra gana sudėtingas dalykas, todėl geras tokios problemos sprendimo būdas gali tapti analizės darbų perdavimas specialistams.

Pirmenybės teikimas. Klientų pirmenybės teikimo informacijos rinkimas yra gana naujas ir greitai besiplečiantis reiškinys. Remdamasi tokia informacija organizacija gali sukurti ypač efektyvius pasiūlymus ir paslaugas, aplenkdamas konkurentus ir pirmoji pasiūlydama rinkai tai, ko laukia vartotojai. Bendravimas su klientais remiantis tokio tipo informacija yra ilgalaikio pelningo ir patikimo bendravimo garantas.

Išlaikymo aktyvumas. Norint kuo efektyviau valdyti esamų klientų išlaikymo procesą, turėtų būti fiksuojamas kiekvieno kliento praradimas ir registruojami faktoriai, įtakoję tokį vartotojo apsisprendimą. Nusiskundimų paslaugos kainomis, kokybe ir kitų nepasitenkinimą keliančių faktorių analizė gali padėti nustatyti blogai vykdomą rinkos politiką ir ateityje padėti išvengti tokių klaidų pasikartojimo.

1.3.7. Vartotojų atpažinimo procesas

Ypač naudingas veiklos kriterijus yra sukurta naujų vartotojų identifikavimo infrastruktūra. Organizacijai yra naudinga, kai remiantis tam tikrais kriterijų rinkiniais, nauji klientai yra atpažįstami pačio pirmo jų kontakto su organizacija metu. Atpažinus vartotoją gali būti pradedami tam tikri veiksmai, skirti naujam, potencialiam klientui pritraukti. Tokio tipo poveikimas gali sąlygoti efektyvesnį naujų klientų atsiradimą – dvejojantys klientai gali būti pastūmėti apsispręsti, o apsisprendusiems sustiprinamas teisingo pasirinkimo įspūdis. Taipogi tai gali būti puikus klientų apsisprendimo duomenų rinkimo šaltinis.

1.3.8. Tarpininkavimo įvertinimo procesas

Būtina suvokti tarpininkavimo įtaką informacijos valdymo procesui. Dažnai pasitaiko toks atvejis, kad tarp organizacijos ir jos galutinių vartotojų stovi taip vadinamieji „treči asmenys“. Tai gali sukelti tam tikras problemas klientų duomenų rinkime, valdyme ir panaudojime. Būtina efektyviai įvertinti tarpininkų vaidmenį organizacijos veikloje ir imtis ryžtingų veiksmų iškilusių keblumų šalinimui.

1.3.9. Privatumo įvertinimo procesas

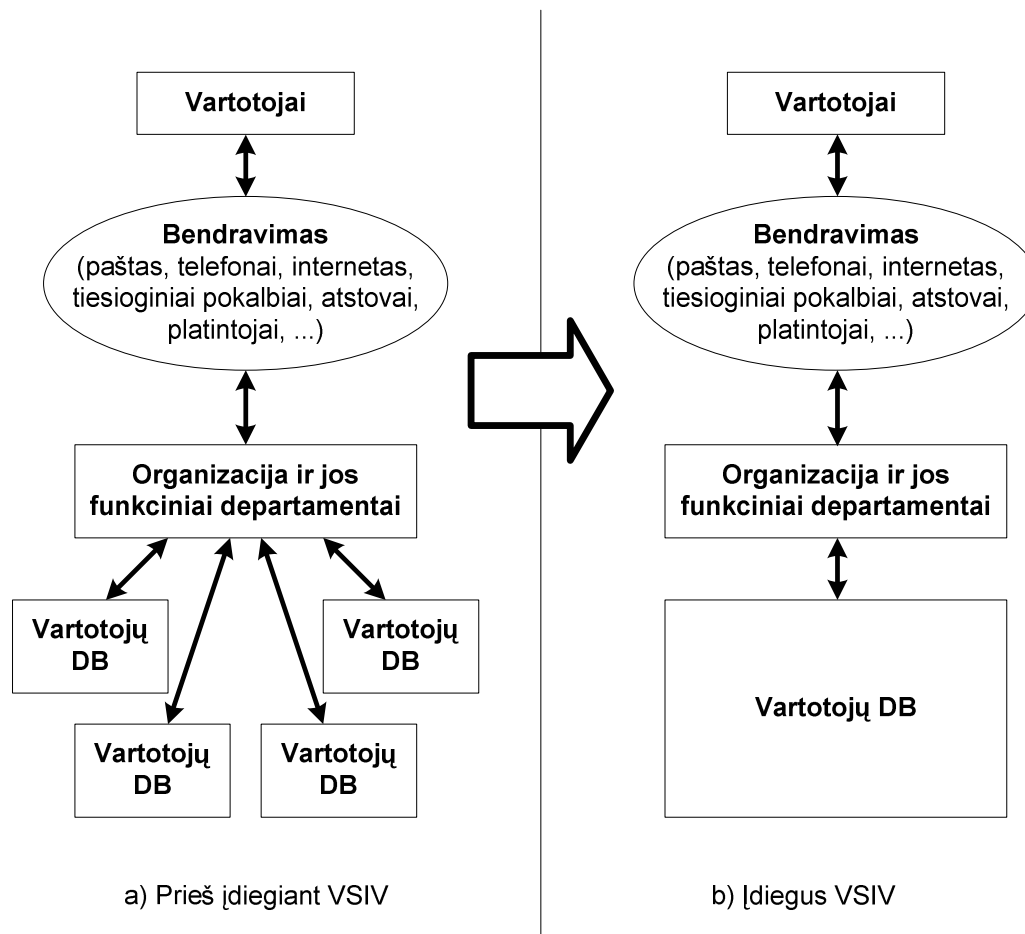
Ypač svarbus kriterijus organizacijos veikloje yra duomenų privatumas. Ši privatumo sąvoka apima tiek pačios organizacijos turimą vidinę informaciją, tiek jos klientų informaciją. Duomenų privatumas yra ypač svarbus. Bet koku privačių duomenų atskleidimu gali pasinaudoti organizacijos konkurentai ir gali būti prarasti milžiniški pinigai. Taipogi jei organizacija vykdo VSIV politiką, ji privalo turėti reikiamas priemonės jos klientų duomenų apsaugai, kadangi tokios informacijos atskleidimas gali paveikti ne tik pačią organizaciją, tačiau ir galutinius jos vartotojus. Taipogi privalu atsižvelgti į šalies įstatymus, kurios teritorijoje vykdoma organizacijos veikla.

1.3.10. Darbų perdavimo galimybių nustatymo procesas

Klientų informacijos valdymas yra sudėtingas procesas, reikalaujantis nemažų investicijų, griežto planavimo, aukštos kokybės specialistų. Taigi, užsiimti tuo gali ne kiekviena organizacija. Kai kuriais atvejais atlikti tam tikras funkcijas pačiai organizacijai gali būti nenaudinga, tačiau tai nereiškia, kad tokios veiklos išvis neturi būti vykdomos. Efektyviausias sprendimas tokiose situacijose galėtų būti dalies veiksmų perdavimas išorinėms kompanijoms. Organizacija turėtų apsvarstyti įmanomas tokios veiksmų sekos galimybes.

1.4. Įgyvendinimo priemonių analizė

Pagrindė, vartotojų sąryšio informacijos valdymas yra rinkodaros įrankis, įgalinantis valdyti bet kokią informaciją gautą bendradarbiaujant su vartotojais. Tokios sistemos turi didelę reikšmę duomenų rinkimo procesuose. Jos sujungia visą organizacijos informaciją, apimančią žurnalus, naujienų leidinius, konferencijas ir įvykius, tam kad eliminuotų duomenų pasikartojimą ir netinkamų duomenų kaupimą. Tokios sistemos visą turimą gausią informaciją transformuoja į ryšiais pagrįstas duomenų bazes – jungia vardus su adresais, prekybos aktyvumu, bendradarbiavimo istorija ir daugeliu kitų kriterijų. Pavyzdžiui vartotojui prisijungus prie kažkokios prezentacijų sistemos ir nurodžius savo asmeninį e-paštą, sistema automatiškai susieja vartotoją su turimais kitais jo aktyvumo įrašais – naujienų, demonstracinių produktų užsakymais ir pan.



3 pav. Skirtumai įdiegus VSIV sistemą

3 pav. iliustruoja du esminius atvejus – a) atveju organizacija nevalo VSIV ir b) organizacijoje yra įdiegta tokio tipo sistema [2].

Naudojantis vartotojų atpažinimo sistemomis organizaciją per trumpesnį laiką gali sukurti daug tikslesnius pasiūlymus ir veiklos strategijas. Organizaciją reikiamu laiku gali pateikti reikiamą pasiūlymą, kuris yra pagrįstas kiekvieno iš vartotojų asmeniniais bendradarbiavimo su ta organizacija duomenimis. Tokios galimybės sukuria tikslesnį veiklos modelį ir investavimo galimybes. Taip pat kartu yra didinamas bendradarbiavimo efektyvumas ir investicijų gražos nauda. Suteikiamos galimybės investicijų gražos analizavimui

Ar VSIV yra tinkamas jūsų organizacijai? Suprantama kad tokios sistemos vertė pirmiausia priklauso nuo organizacijos veiklos, produktų bei paslaugų įvairovės – ir kuo didesnis šios įvairovės tarpusavio sąryšis, veiklų persidengimas – tuo daugiau šansų, kad tokia sistema per tam tikrą laiką pasieks investicijų gražos naudą. Organizacija turi kritiškai pažvelgti į kiekvieną savo veiklą ir atlikti išsamią analizę – kad įvertinti ar duos naudos procesų automatizavimas, ar būtų galima tikėtis naudos iš tokių sistemų įdiegimo ir ar apylamai verta investuoti pinigus į VSIV sistemų kūrimą. Bet kokių atveju, VSIV sistemų kūrimo žingsnių ir galimų įgyvendinimo kliūčių suvokimas įgalina realiau pažvelgti į esamą veiklą ir nuspręsti, ar toks automatizuotas vartotojų informacijos valdymas gali būti naudingas ir neatneš pražūtingų pasekmių bei investicijų praradimą.

1.4.1. Pagrindiniai veiklos automatizavimo įrankiai

Pateikiame tris pagrindinius kriterijus, kurie padeda įvertinti kaip VSIV sistemos gali įgyvendinti efektyvesnį, pigesnį ir pelningesnį veiklos organizavimą [3]:

Kampanijų perspektyvų identifikavimas. Vartotojų atpažinimas ir grupavimas gali būti vartojamas labiausiai reaguojančių klientų segmentų nustatymui, kampanijų sąnaudų sumažinimui ir galiausiai pelningumo per užsakymą padidinimui.

Rinkodaros specialistai dažnai skirsto vartotojus į grupes, kuriose vartotojai pasižymi panašiomis charakteristikomis. Pavyzdžiui klientų grupių identifikavimas gali būti vartojamas nustatyti klientams, kurie susidomėjo pirmo pasiūlymo metu ir nuspręsti kuriems iš jų, atėjus laikui, būtų galima pateikti sekantį pasiūlymą.

Grupės taipogi gali būti skirstomos pagal polinkius: tikimybes, kad vartotojas atsilies į tam tikras kampanijas. VSIV sistemos gali padėti aiškiau nustatyti kurios vartotojo ankstesnio bendradarbiavimo charakteristikos daugiau ar mažiau gali nusakyti jo norą susidomėti pasiūlymų kampanijomis. Remiantis procentiniais tokių charakteristikų įvertinimais galima nuspręsti kuriems vartotojams gali būti siunčiami tiesioginiai laišakai su pasiūlymais.

Automatinis kampanijų generavimas. Programuojant VSIV sistemas su specifinėmis taisyklėmis, pasiūlymų kampanijos gali būti vykdomos automatiškai, kas didina veiklos efektyvumą. Jeigu vartotojas ateina į organizaciją pirmąkart, tuomet VSIV sistema gali būti nustatyta taip, kad automatiškai atpažintų pirmąjį vartotojo apsilankymą ir jam nusiųstų personalizuotą sveikimo laišką.

Taipogi galimas daugelio kitų automatizavimo taisyklių nustatymas – kaip kampanijų vykdymas atėjus naujiems metams ar kitoms šventėms, periodišką naujienų siuntimas, naujo pasiūlymo siuntimas baigiantis tam tikrai sutarčiai ir daugelis kitų taisyklių. Kuriamos taisyklės gali būti nustatomos taip, kad ne tik automatizuotų reikiamus darbus, tačiau veiktų efektyviai vartotojų atžvilgiu – vertintų kiekvieno iš vartotojų privatumą bei jo apsisprendimą – pavyzdžiui, jei vartotojas po keleto pasiūlymų praėjus tam tikram laiko tarpui neatsiliepė nei vienąkart – jis gali būti pašalinamas iš tokių automatinų kampanijų sąrašų. Taipogi turi būti nustatomi ir kitokie kriterijai, kaip stengtis išlaikyti organizacijos gerą vardą ir reputaciją – pavyzdžiui, trijų pasiūlymų siuntimas per vieną dieną gali sukelti vartotojų nepasitenkinimą ir priešišumą organizacijai. Todėl organizacija turi nustatyti tokias politikas, kaip susisiekimų su vartotojais kiekis per dieną, per savaitę, per mėnesį ar per metus.

Realaus laiko personalizacijos įgyvendinimas. „Vienas su vienu“ bendravimas dažnai vartojamas užmegzti kontaktui su vartotojais. Toks bendravimas yra pagrindinis VSIV sistemų etalonas. Tai ne tik pagerina patį bendravimą, tačiau padidina ir patį vartotojo bendravimo su organizacija efektyvumą ir jo pasitenkinimą esama sistema.

Amazon.com yra vienas iš geriausių ir efektyviausių personalizacijos įgyvendinimo atvejų [7]. Egzistuojančiam vartotojui apsilankius puslapyje, VSIV sistema automatiškai atpažįstą jį ir pateikia jam personalizuotą sveikinimą vartojant registracijos metu vartotojo pateiktą informaciją, tokią kaip vardas, pavardė, gimimo data[8]. Toliau analizuojama vartotojo užsakymų istorija, pateikiami panašių prekių siūlymų sąrašai, kuriais vartotojas gali susidomėti. Kuriant tokio tipo VSIV sistemas, galima vadovautis panašiais kriterijais – į puslapį atėjus registruotam vartotojui – gali būti vartojama jo asmeninė informacija glaudesnio ryšio su vartotoju sukūrimui, kas teikia naudą tiek pačiam vartotojui – jis jaučia rodomą jam dėmesį, tiek pačiai organizacijai – ji užmezga artimesnius ryšius su kiekvienu iš vartotojų.

2. Teorinis problemos sprendimas

Analizės dalyje išsiaiškinome, kad efektyviai organizacijos veiklai ypač svarbus yra jos vartotojų identifikavimas, jų poreikių analizė ir klasifikavimas. Šiaip problemai spręsti pateikiame teorinį vartotojų atpažinimo sistemos modelį. Šio modelio pagrindinis tikslas – naudojantis jau žinomais duomenų gavybos būdais ir turimais vartotojų ankstesnės sąveikos su organizacija duomenimis kurti taisykles, padedančias sujungti vartotojus į grupes, pasižyminčias vienodais elgesio ypatumais. Tokios grupės leidžia lengviau valdyti panašiais interesais pasižyminčias vartotojų grupes ir priimti tikslesnius sprendimus, realiau atitinkančius vartotojų poreikius.

Pradėsime nuo duomenų gavybos būdų, kurie labiausiai atitinka tokio modelio reikmes ir efektyviausiai gali analizuoti vartotojų sąveikos su organizacija duomenis.

2.1. Duomenų gavybos metodai

Duomenų gavyba yra procesas, kurio metu iš didelių informacijos kiekių, saugomų duomenų bazėse ar kitose saugyklose, yra randama nauja informacija, tokia kaip – šablonai, asociacijos, pasikeitimai, anomalijos ir prasmingos struktūros. Dėl didelių informacijos kiekių, saugomų elektroninėje formoje ir neišvengiamo poreikio paversti tokius duomenis į naudingą informaciją ir žinias, galimas panaudoti kitose srityse kaip rinkodaros analizė ar sprendimų priėmimas, duomenų gavyba pastaraisiais metais tapo svarbia sritimi informacijos industrijoje. „Jei įprasti duomenų analizės metodai parodo norimų kintamųjų priklausomybę, tai duomenų gavyba unikali tuo, kad analizės rezultatas yra naujų priklausomybių, apie kurių egzistavimą buvo ar net nebuvo įtariama, radimas.“ (R.Šileikienė, 2000) [4].

Pagrindė, žinių radimo procesas susideda iš šių pasikartojančių žingsnių sekų:

- *duomenų valymo*, kuris dirba su netvarkingais, klaidingais, trūkstamais ar nereikšmingais duomenimis,
- *duomenų integravimo*, kurio metu pasikartojantys, heterogeniniai duomenys yra suliejami į visumą,
- *duomenų pasirinkimo*, kurio metu duomenys, reikalingi analizei, yra imami iš duomenų bazės,

- *duomenų transformavimo*, kurio metu duomenys yra pervedami ar apjungiami į kitas saugojimo formas, patogesnes duomenų gavybai,
- *duomenų gavybos*, kurios metu naudojant specialius metodus iš turimų duomenų yra gaunami šablonai, asociacijos ir kitokios naujos žinios,
- *duomenų įvertinimo*, kurio metu duomenys yra grupuojami pagal prieš tai sudarytus šablonus bei taisykles, ir
- *žinių pateikimo*, kur naudojantis pateikimo ir vizualizacijos technologijomis, sukaupta nauja informacija yra pateikiama vartotojui.

Yra daug visokių duomenų gavybos algoritmų, tačiau kai kurie iš jų yra labiau specializuoti ir sunkiai pritaikomi biznio sprendimuose [5,9]. Todėl labiausiai tinkamus vartotojų sąryšio informacijos valdymui galima suskirstyti į šias tris grupes:

1. *Prognozė*. Vartotojų praeities elgesio informacijos naudojimas ateities elgesiui nustatyti. Prognozavimo algoritmai pateikia modelius ar struktūras, kurie gali nusakyti būsimus vartotojų veiksmus. Pavyzdžiui, prognozėmis besiremiantis modelis gali nusakyti sekantį labiausiai tikėtiną produktą, kurį vartotojas gali įsigyti, remiantis jo paties praeities pirkimų informacija arba praeities informacija į jį panašių vartotojų, kurie pirko tuos pačius produktus.
2. *Seka*. Ši analizė apibrėžia veiklos sekas, kurios pasireiškia tam tikra tvarka. Sekos gali būti vartojamos nustatyti klientams, kurie tam tikrus veiksmus atlieka konkrečia tvarka. Tokių sekų nustatymas gali padėti lengviau identifikuoti vartotojų elgesį duomenyse, gautuose iš įvairių šaltinių, ir sukurti tikslesnius modelius.
3. *Asociacijos*. Asociacijų analizė nustato panašių elementų ar segmentų grupes. Tai yra naudinga, kai norima nustatyti elementus ar įvykius kurie atsiranda kartu. Asociacijų algoritmai dažnai yra naudojami rinkodaroje, kad lengviau ir tiksliau nustatyti produktus, kurie dažniausiai yra perkami kartu. Lengviau suvokiant bendrus vartotojų ir produktų bruožus, organizacija gali tiksliau nuspręsti, kuriuos produktus patartina labiau reklamuoti ir kuriems iš jų teikti nuolaidas.

Toliau apžvelgsime keletą duomenų gavybos metodų ir pabandysime juos įvertinti.

2.1.1. Artimiausių kaimynų metodas

Artimiausių kaimynų metodas (dar žinomas kaip k -artimiausių kaimynų arba k -NN) yra vienas iš prognozavimo metodų. Skirtingai nuo kitų prognozės algoritmų, čia bandomieji duomenys nėra peržvelgiami ar apdorojami tam kad sukurti modelį. Priešingai, šiame metode bandomieji duomenys ir yra pats modelis. Tuomet, kai modeliui yra pateikiamas naujas duomenų egzempliorius, algoritmas analizuoja visus duomenis kad nustatytų duomenų aibes, kurios yra artimiausios šiam naujam egzemplioriui. Artimiausių kaimynų algoritme yra naudojami du parametrai: a) artimiausių vartojamų egzempliorių skaičius k ir b) artimumo apibrėžimo matas. Kiekvienas algoritmo panaudojimas reikalauja, kad būtų nustatytas sveikasis skaičius k , kuris apibrėžia kiek duomenų egzempliorių turi būti peržiūrėta, kad nustatyti kuriai grupei gali priklausyti naujasis egzempliorius.

Artimiausių kaimynų metodas remiasi nuotolio sąvoka. Nuotolis yra vartojamas apskaičiuoti atstumams tarp duomenų egzempliorių. Nuotolio nustatymo metrika yra labai svarbus dalykas, kadangi nuo sąvokos „geras“ matas priklauso sprendimai. Atstumas tarp egzempliorių turi būti įvertintas skaitinėmis reikšmėmis, tam kad būtų galima atlikti palyginimus ir nustatyti tinkamiausius variantus. Vartojant skirtingas gerumo metrikas tiems patiems tiriamiesiems duomenims yra gaunami skirtingi rezultatai, todėl yra labai svarbus teisingų metrių pasirinkimas. Automatinis tokių įverčių nustatymas yra neįmanomas, todėl čia yra reikalingas žmogaus įsikišimas.

Šio metodo eigoje yra sprendžiama kuriai iš esamų grupių gali priklausyti naujasis egzempliorius. Tam yra skaičiuojami nuotoliai tarp naujojo egzemplioriaus ir esamų kitų egzempliorių. Gautos reikšmės lyginamos su iš anksto apibrėžtomis panašumo įvertinimo reikšmėmis ir ieškoma grupė, kuri yra artimiausia naujajam egzemplioriui. Radus tokią grupę daroma prognozė, kad naujasis egzempliorius priklauso tai grupei.

Pavyzdys. Tarkime turime duomenis apie banko vartotojus. Tikslas yra nustatyti ar tai bus gera ar bloga rizika suteikti kreditą vartotojui. Bandomieji duomenys pateikiami lentelėje:

1 lentelė

Vardas	Skola	Pajamos	Šeima?	Rizika
Petras	Didelė	Didelės	Taip	Gera
Sigita	Maža	Didelės	Taip	Gera

Vardas	Skola	Pajamos	Šeima?	Rizika
Jonas	Maža	Didelės	Ne	Bloga
Marija	Maža	Mažos	Taip	Bloga
Francas	Maža	Mažos	Taip	Bloga

Pavyzdyje paprastumo dėlei, visų stulpelių, išskyrus vardą, reikšmes gali būti tik dvi. Vartosime artimiausių kaimynų algoritmą, pasirenkant $k = 3$ (sutrumpintai 3-AK). Nuotoliui išmatuoti naudosime paprastą metriką, sumuodami taškus kiekvienam iš šių trijų stulpelių, kur stulpelio taškų reikšmė bus lygi 0, jeigu dviejų egzempliorių reikšmės yra vienodos, ir 1, jeigu jos skiriasi. Taigi paskaičiuojame nuotolius tarp kiekvieno iš vartotojų:

2 lentelė

	Petras	Sigita	Jonas	Marija	Francas
Petras	0	1	2	1	2
Sigita	1	0	1	2	1
Jonas	2	1	0	3	2
Marija	1	2	3	0	1
Francas	2	1	2	1	0

Šioje lentelėje aprašomi visi atstumai kiekvienam iš vartotojų su likusiais vartotojais. Toliau vartosime 3-AK techniką, kad įvertinti kaip jis suklasifikuos mūsų duomenis. Kadangi pasirinkome, kad $k=3$, tai mus domina tik trys artimiausi vartotojai kiekvienam iš duomenų aibės vartotojų. Pavyzdžiui, vartotojui *Petras* artimiausi kaimynai yra *Petras*, *Sigita* ir *Marija*. Kaip matome, *Petrui* vienu iš artimiausių kaimynų yra parinktas jis pats. Tai gali atrodyti nelogiška, tačiau šiuo atveju tai neturi jokios įtakos, kaip ir bet kuriam kitam modeliui, testuojamam su bandymo duomenimis. Štai kodėl mums visuomet reikia atskirą duomenų rinkinį modelio

patikrinimui. Taigi, trijų artimiausių *Petro* kaimynų rizikos yra *Gera*, *Gera* ir *Bloga*. Taigi daroma prognozė, kad *Rizikos* reikšmė Petruui yra dažniausia tarp k kaimynų – šiuo atveju – *Gera*. Išrenkant artimiausius kaimynus vartotojui *Sigita* kyla problemos – be jos pačios dar yra trys kaimynai su vienodais nuotoliais. Taigi mes galėtume įtraukti visus tris, neištraukti šių trijų arba įtraukti visus tris su proporcingomis reikšmėmis ($2/3$ kiekvienam). Sprendimas tiesiogiai priklauso nuo algoritmo įgyvendinimo. Mūsų pavyzdyje vartosime proporcingus įverčius. Taigi rizikų reikšmės bus – *Gera*, $2/3$ *Gera*, $2/3$ *Bloga* ir $2/3$ *Bloga*.

Sekančioje lentelėje pateikiamos prognozės, gautos naudojantis 3-AK algoritmu.

3 lentelė

Vardas	Skola	Pajamos	Šeima?	Rizika	3-AK prognozė
Petras	Didelė	Didelės	Taip	Gera	Gera
Sigita	Maža	Didelės	Taip	Gera	Gera
Jonas	Maža	Didelės	Ne	Bloga	-
Marija	Didelė	Mažos	Taip	Bloga	Bloga
Francas	Maža	Mažos	Taip	Bloga	Bloga

Kadangi tiek k , tiek *atstumo* reikšmės turi didelę įtaką modelio kokybei, jų reikšmes reikia parinkinėti labai atsargiai.

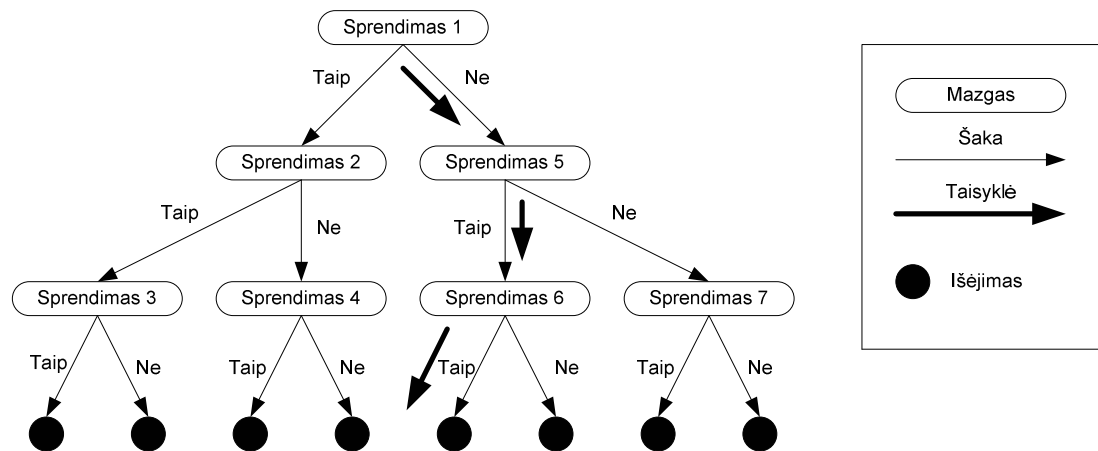
Išvados. Artimiausių kaimynų metodas neatlieka jokių veiksmų, susijusių su mokymusi. Deja tai daro patį metodą ne itin efektyviu, ypač prie didelių duomenų kiekių. Kiekvienas naujas egzempliorius turi būti palygintas su kiekvienu iš jau esančių ir tai užima nemažai laiko ir resursų. Dėl to sprendimai negali būti atliekami realiu laiku ir turi būti laukiama, kol bus atliktos prognozės. Nepaisant to, kad pati algoritmo idėja yra gan paprasta, tačiau jos parametrų k ir *nuotolio* parinkimas taipogi yra šiek tiek komplikuoatas. Norint tiksliai nustatyti šias reikšmes yra būtinas tam tikro lygio testavimas. Tikrų mokymosi procesų stoka išskiria šį metodą iš kitų prognozavimo metodų.

2.1.2. Sprendimų medžiai

Sprendimų medžiai yra viena iš populiariausių duomenų gavybos technologijų. Juos lengva naudoti, jų rezultatai yra lengvai suprantami vartotojams. Jų privalumas yra tas, kad jie

gali išspręsti daugelį klasifikacijos problemų, jie lengvai naudojami su įvairiais duomenų formatais ir yra efektyvūs analizuojant didelius duomenų kiekius [9].

Norint sukurti sprendimų medį iš bandomųjų duomenų – duomenis reikia suskaldyti į mažesnius jų rinkinius. Kiekviena iteracija analizuoja duomenis tik viename mazge. Pirmoji iteracija analizuoja pirmąjį mazgą, kuris yra viso medžio pradžia ir apima visus duomenis. Toliau einama giliau per išvestinius mazgus ir tikrinami tie mazgai, kurie vis dar apima duomenis. Kiekvienos iteracijos metu reikia pasirinkti nepriklausomą atributą, kuris efektyviausiai gali suskaldyti duomenis. Tai reiškia, kad duomenys perskelti atsižvelgiant į nepriklausomą atributą turi būti kiek įmanoma daugiau homogeniški. Iteracijos yra kartojamos tol, kol pasiekiamas išėjimo mazgas. Išėjimo mazgas standartiškai pasiekiamas tada, kai suskaidyti duomenis tampa neįmanoma arba netikslinga. Bendra sprendimų medžio schema pateikiama 4 pav.



4 pav. Sprendimų medžio schema

Pereinant medį nuo pagrindinio mazgo iki išėjimo mazgo yra gaunama taisyklė, kurią pritaikius duomenims, jie bus priskirti kažkokiai nors kategorijai. Taisyklės apibrėžiamos JEIGU-TADA formomis.

Pavyzdys. Vartosime tuos pačius bandomuosius duomenis kaip ir artimiausių kaimynų metodo pavyzdyje. Šiame pavyzdyje įvesime entropijos sąvoką, kuri skirta nustatyti kiek nežinomybės yra duomenų rinkinyje. Entropija yra matuojama skalėje nuo 0 iki 1. Jeigu aibė buvo suskaidyta lygiu santykiu tarp gerų ir blogų rizikų, tada mes negalime būti tikri ar pasirinktas atsitiktinis vartotojas turi geras ar blogas rizikas. Tokiu atveju aibės entropija turėtų

būti 1. Kitu atveju, jei visoje aibėje buvo vien tik geros ar blogos rizikos, tada netikrumo nelieka ir entropija tada lygi 0.

Pradėsime nuo entropijos matavimo visiems bandomiesiems duomenims. Rasime gerų rizikų ir blogų rizikų dalis. Entropija yra paskaičiuojama pagal formulę:

$$entr = -\sum_i p_i \log_2 p_i$$

Aibėje turime dvi geras ir tris blogas rizikas, taigi:

$$entr = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5}\right) = \mathbf{0.97095}$$

Toliau apsvairstome visus įmanomus dalijimo variantus – pagal skolą, pagal pajamas ir pagal šeimyninę padėtį. Suskaičiuojame bendrą entropiją rezultatų aibėms kiekvienam iš šių trijų variantų.

Pabandykime suskaldyti aibę pagal skolą. Lentelėje pateikiamas skaldymas pagal skolą rizikų atžvilgiu:

4 lentelė

	Gera rizika	Bloga rizika	Bendras
Didelė skola	1	1	2
Maža skola	1	2	3
Bendras	2	3	5

Aibė su *didele skola* turi vieną gerą riziką ir vieną blogą riziką, taigi:

$$entr = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = \mathbf{1.00000}$$

Aibė su *maža skola* turi vieną gerą riziką ir dvi blogas rizikas:

$$entr = -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3}\right) = \mathbf{0.91830}$$

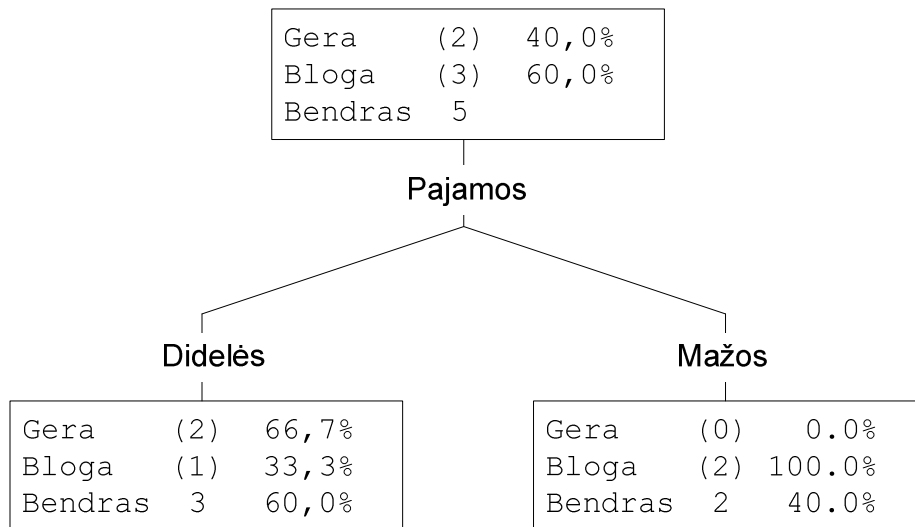
Kadangi iš viso yra dvi *didelės skolos* ir trys *mažos skolos*, bendra abiejų aibių, skaidytų pagal *skolą*, entropija bus:

$$entr = (2/5)*1 + (3/5)*0.91830 = 0.95098$$

O bendra entropija, gauta skaidant pagal *skolą*, bus:

$$entr = 0.97095 - 0.95098 = 0.01998$$

Analogiškai paskaičiavus bendras entropijas skaidant pagal pajamas ir šeimyninę padėtį gauname, kad jos yra atitinkamai 0.41998 ir 0.17095. Taigi skaidymas pagal pajamas yra pats efektyviausias duomenų skaidymas, ir gautos aibės yra labiausiai homogeniškos:



5 pav. Duomenų skaidymas pagal pajamas

Antroji suskaldytų duomenų aibė susideda iš 100 proc. *blogų rizikų*. Kadangi tai yra visiškai homogeniška aibė (ir jos entropija yra lygi 0), todėl šitoje aibėje skaidymai daugiau nebus atliekami. Pirmoji aibė ir toliau susideda iš *blogų* ir *gerų rizikų*. Jos entropija yra 0.91830 ir jis turi būti suskaldyta pagal antrąją nepriklausomą reikšmę. Dabar reikia pasirinkti, pagal kurias reikšmes bus skaldomi duomenys – pagal *skolą* ar pagal *šeimyninę padėtį*. Pirmą skaidome duomenis pagal *skolą*. Skaidymo duomenys pateikiami 5 lentelėje.

	Gera rizika	Bloga rizika	Bendras
Didelė skola	1	0	1
Maža skola	1	1	2
Bendras	2	1	3

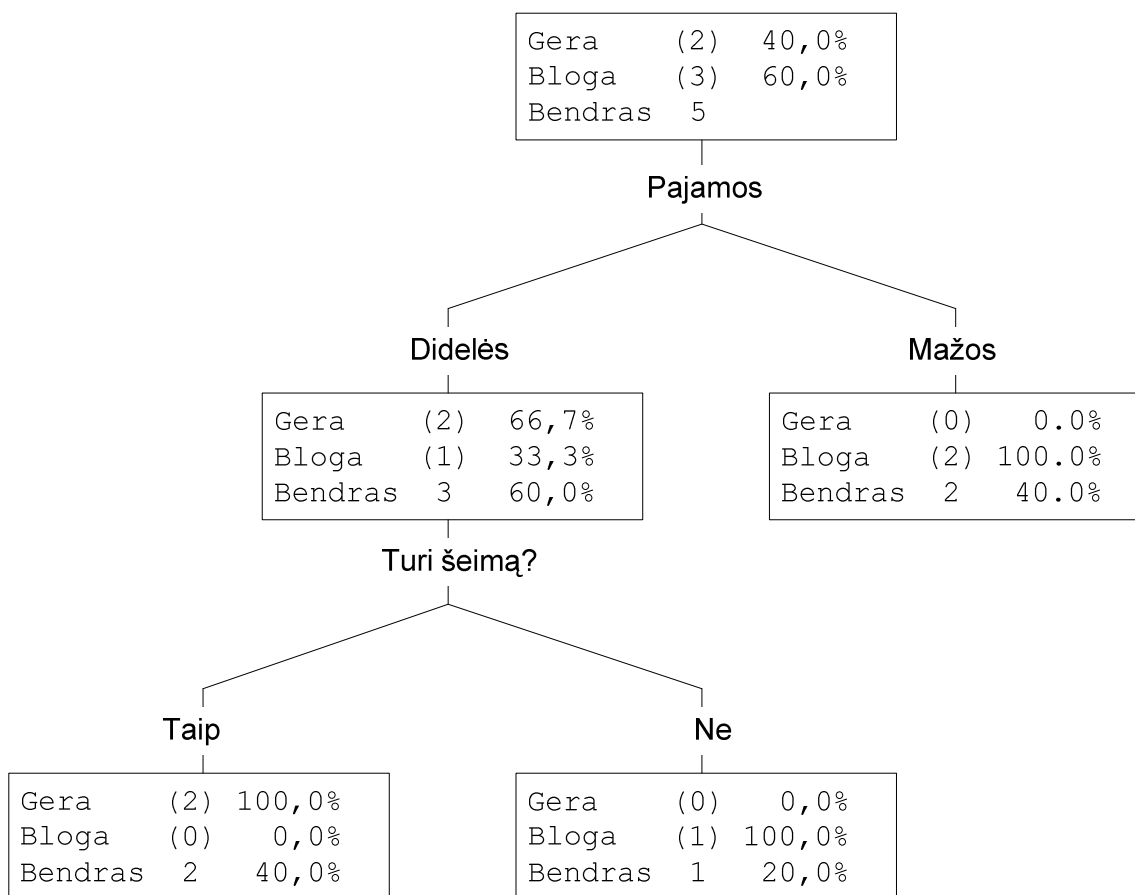
Duomenų aibė su *didele skola* turi vieną *gera riziką* ir jokių *blogų rizikų*. Ji visiškai homogeniška ir jos entropija yra lygi 0. Aibė su *maža skola* turi vieną *gerą riziką* ir vieną *blogą riziką*. Jos entropija yra lygi 1. Kadangi bendrai yra viena *didelė skola* ir dvi *mažos skolos*, tai bendra aibės, skaidomos pagal *pajamas*, entropija yra:

$$entr = (1/3)*0 + (2/3)*1 = 0.66667$$

Taigi, skaidant *pajamas* pagal *skolą* gauname, kad entropija yra lygi:

$$entr = 0.91830 - 0.66667 = 0.25163$$

Įvertinus galimybes skaidyti duomenis pagal *šeimyninę padėtį* gauname, kad gautos duomenų aibės būtų visiškai homogeniškos, ir entropija tada yra lygi 0. Taigi pasirenkame skaidymą pagal *šeimyninę padėtį*.



6 pav. Tolimesnis duomenų skaidymas pagal šeimyninę padėtį

Tolimesni skaidymai nereikalingi, kadangi jau turime sprendimų medį, kuris gali būti vartojamas naujiems atvejams. Skaičiavimų medis gali būti apibrežtas šiuo taisyklių rinkiniu:

- a) **JEIGU** Pajamos = *Didelės* **IR** Turi šeimą = *Taip* **TADA** Rizika = *Gera*
- b) **JEIGU** Pajamos = *Didelės* **IR** Turi šeimą = *Ne* **TADA** Rizika = *Bloga*
- c) **JEIGU** Pajamos = *Mažos* **TADA** Rizika = *Bloga*

Išvados. Naudojant šį metodą sudaromas lengvai suprantamas modelis. Jį galima nesunkiai analizuoti nesinaudojant jokiais papildomais įrankiais. Be to, kiekvienam naujam egzemplioriui nereikalingas pilnas palyginimas su kiekvienu iš modelio elementų. Tai didina metodo efektyvumą ir greitaveiką. Šis metodas gali dirbti netgi su nepilnais duomenimis. Prie trūkumų galima priskirti tai, kad tokio tipo metodai reikalauja duomenų paruošimo – jie turi būti paruošti specialia forma. Analizės metu gali būti lyginama tik vienalyte priklausomybė. Jei norima atlikti analizę keleto kriterijų atžvilgiu – reikia sudaryti kelis sprendimų medžių modelius.

2.1.3. Asociatyvių taisyklių metodas

Asociatyvinių taisyklių metodo esmė yra asociatyvinių taisyklių išskyrimas duomenų analizės metu. Remiantis rastomis taisyklėmis toliau galima kurti šablonus, apibrėžiančius vartotojų elgesį. Tokio tipo analizė labai tinkama tirti transakcijų duomenis [6,9] – pavyzdžiui pirkinių krepšelių ir pan. Kiekviena transakcija susideda iš kintamo duomenų elementų kiekio. Tegul $I = \{x_1, x_2, \dots, x_n\}$ būna skirtingų duomenų elementų rinkinių aibė. Duomenų rinkinys X , priklausantis aibei I yra vadinamas I aibės poaibiu. Tegul I elementų aibė sudaro duomenų bazę D . Tuomet kiekvienas $T \in D$ yra vadinamas transakcija. Laikome, kad transakcija $T \in D$ turi aibę $X \subseteq I$, jeigu $X \subseteq T$. Taisyklės yra nustatomos tarp dviejų ar daugiau transakcijos aibės elementų ir žymimos $X \rightarrow Y$, kur X ir Y yra transakcijos T elementai. Kiekviena analizės metu sukurta taisyklė yra apibrėžiama dviem parametrais – *patvirtinamumu* ir *pasikliovimu*.

Patvirtinamumas yra skirtas nusakyti šablono pasikartojimo dažnį visų transakcijų atžvilgiu. Skaitinė šio parametro reikšmė yra išreiškiama procentais ir formaliai apibrėžiama kaip:

$$patv(X \rightarrow Y) = \frac{|\{T \in D \mid X \cup Y \subseteq T\}|}{|D|}$$

Atsižvelgiant į šio parametro reikšmes galima eliminuoti nepasitvirtinusias taisykles – joms patvirtinimo elemento reikšmė yra nedidele.

Pasikliovimas apibūdina elementų susietumą ir parodo kiek vienas elementas yra priklausomas nuo kito. Jo skaitinė reikšmė taip pat yra išreiškiama procentais.

$$pas(X \rightarrow Y) = \frac{|\{T \in D \mid X \cup Y \subseteq T\}|}{|\{T \in D \mid X \subseteq T\}|}$$

Pavyzdys. Tarkime, turime pardavimų duomenų bazę, susidedančią iš transakcijų, pateiktų 6 lentelėje:

6 lentelė

Transakcija	Prekės
T001	{Kondicionierius, Oro pagalvė, Akumulatorius}
T002	{Sankaba, Radijas, Akumulatorius}
T003	{Kondicionierius, Radijas, Oro pagalvė}
T004	{Oro pagalvė, Akumulatorius}
T005	{Kondicionierius, Sankaba, Akumulatorius}
T006	{Radijas, Akumulatorius, Oro pagalvė}
T007	{Akumulatorius, Radijas}
T008	{Oro pagalvė, Akumulatorius, Radijas}

Pabandykime rasti taisyklę, kurios patvirtinamumas yra 50 proc., o pasiklojimas 60 proc. Ieškome pasitaikančių duomenų rinkinių, kurių patvirtinamumas yra ne mažiau negu 50 procentų. Imame elementą *Kondicionierius*:

$$patv(Kondicionierius) = \frac{|\{T001, T004, T005\}|}{|D|} = \frac{3}{8} = 0,375 = 37,5\%$$

Matome, kad šis elementas netenkina užsibrėžtų sąlygų. Analogiškai paskaičiuojame patvirtinamumus visiems elementams ir gauname, kad šį kriterijų tenkina šie elementai:

7 lentelė

Elementas	Patvirtinamumas
Oro pagalvė	62,5%
Akumulatorius	87,5%
Radijas	62,5%

Šie elementai sudaro tris įmanomas poras. Kiekvienai iš jų mes apskaičiuojame porų patvirtinamumą:

8 lentelė

Pora	Patvirtinamumas
{Oro pagalvė, Akumuliatorius}	50%
{Radijas, Akumuliatorius}	50%
{Oro pagalvė, Radijas}	37,5%

Kaip matome pora {*Oro pagalvė, Radijas*} netenkina reikalavimų, todėl ji yra atmetama. Likusioms poroms paskaičiuojame pasiklovimus:

9 lentelė

Pora	Pasiklovimas
{Oro pagalvė, Akumuliatorius}	80%
{Akumuliatorius, Oro pagalvė}	57,1%
{Akumuliatorius, Radijas}	57,1%
{Radijas, Akumuliatorius}	80%

Kaip matome, pagal pasiklovimų įverčius mūsų reikalavimus tenkina poros {*Oro pagalvė, Akumuliatorius*} ir {*Radijas, Akumuliatorius*}. Taigi galime daryti išvadą, kad perkant *oro pagalvę* vartotojai taip pat yra linkę įsigyti *akumuliatorių* ir perkant *radijo imtuvą* taip pat susidomi *akumuliatoriumi*. Abiem atvejais tokio apsisprendimo tikimybė yra 80 procentų.

Išvados. Metodas lengvai suprantamas ir plačiai pritaikomas, kada norima analizuoti ryšius tarp produktų ar paslaugų. Šis metodas yra labai populiarus internetiniuose prekybos sprendimuose ir jam sugalvota nemažai algoritmų ir jų modifikacijų. Ypač patogus jo naudojimas, kada duomenis galima nusakyti taisyklėmis JEI-TAI. Nereikalauja didelių resursų skaičiavimams atlikti. Lengvai gali būti vartojamas su duomenimis, kuriuos nesunkiai galima grupuoti. Metodas skirtas duomenims, kuriuose galima išvelgti asociacijas, todėl duomenims, kuriuose jų nėra, šis metodas yra netinkamas.

2.1.4. Duomenų gavybos metodų įvertinimas

Egzistuoja įvairūs metodai, skirti duomenų gavybai. Daugelis iš jų nėra universalūs ir skirti taikyti specialiose srityse. Todėl norint išsirinkti metodą kokiam nors problemai spręsti, svarbu atsižvelgti į metodo taikymo sritį, jo duodamus rezultatus ir duomenų kilmę ar tipą, su kuriuo efektyviausiai gali dirbti. Palyginimui pateikiame analizuotų duomenų gavybos metodų įvertinimus:

10 lentelė

Gavybos metodas	Privalumai	Trūkumai
Artimiausių kaimynų metodas	Lengvai suprantamas metodas. Gali dirbti su įvairaus tipo duomenimis. Nereikalauja jokių žinių apie tiriamų duomenų struktūrą	Mokymosi metodų trūkumas. Reikalauja daug resursų skaičiavimams. Galimos problemos nustatant k ir <i>nuotolio</i> parametrus.
Sprendimų medžio metodas	Sudaromas modelis lengvai suprantamas. Nereikia jokių papildomų priemonių rezultatų analizavimui. Nereikalauja daug skaičiavimo resursų. Dirba su nepilnais duomenimis. Patogūs kai sprendimo rezultatas turi būti nustatomas remiantis keletu duomenų elementų.	Reikalingas duomenų paruošimas. Lyginimas pagal vieną priklausomybę. Norint analizuoti pagal keletą kriterijų – sudaromi atskiri sprendimų medžiai. Netinkamas sprendimams priimti, kurie priklauso nuo laiko faktoriaus.
Asociatyvių taisyklių metodas	Plačios pritaikymo galimybės, lengvai suprantami rezultatai. Nereikalauja daug resursų skaičiavimams atlikti.	Tinka tik duomenims, kuriuose galima išvelgti asociacijas.

2.2. Vartotojų atpažinimo sistemos modelis

VSIV sistemos sėkmingos veiklos užtikrinimui viena iš svarbiausių dalių yra vartotojų atpažinimas ir tinkamas jų grupavimas. Šiam tikslui būtina apžvelgti vartotojų klasifikavimo ir jų profilių sudarymo sąvokas.

Vartotojų klasifikavimas – tai procesas, kurio metu vartotojai yra suskirstomi į homogenines grupes, pasižyminčias vienodais požymiais (atributais). Atributai gali būti įvairūs – tai pomėgiai, skonis, ieškomos informacijos pobūdis, įpročiai ir t.t.

Vartotojų profiliavimas – tai procesas, kurio metu vartotojai yra aprašomi pagal jau suskirstytus atributus. Profiliavimas yra atliekamas sudarant vartotojų elgesio modelį ir įvertinant jam būdingus parametrus. Vartotojų profilių sudarymas – išorinių duomenų pritaikymo būdas galimiems vartotojams nustatyti. Pagrindinis vartotojo profilio sudarymo tikslas yra pagal surinktus duomenis nuspėti vartotojo elgseną. Vartotojų profiliavimo procesas yra atliekamas po vartotojų klasifikavimo.

2.2.1. Vartotojų profilių sudarymas

Vartotojų profiliai yra skirti pagerinti organizacijos bendravimą su jos klientais. Vartotojų profilių vartojimas įgalima organizaciją sukurti geresnes paslaugas ir pasiūlymus, labiau atitinkančius realius vartotojų poreikius. Kiekvienai vartotojų grupei, sukurtai vartotojų klasifikacijos etape, yra sukuriamas atskiras profilis, apimantis tik tai grupei būdingus požymius, tokius kaip vartotojų elgesio ypatumai, jų demografinė informacija, kalba ir pan. Reikia atkreipti dėmesį į tai, kad visa ši informacija gali būti gauta iš įvairių šaltinių, todėl kartu yra sprendžiamos ir informacijos susiskaldymo ir dubliavimosi problemos. Be to, vartotojų profiliai taip pat gali būti vartojami ir naujų klientų pritraukimui. Pavyzdžiui jei sistemoje besilankančio vartotojo atributų aibė yra labai panaši į vieną ar keletą jau esamų vartotojų šablonų atributų aibę, tai tokiam klientui gali būti pateikiami tikslesni sprendimai, kurie buvo sukurti remiantis ankstesne bendravimu su kitais, į jį panašiais vartotojais, patirtimi.

Vienas iš paprasčiausių vartotojų grupės pavyzdžių yra asmens byla. Tai duomenų rinkinys, turintis savyje adresą, miestą, šalį ir pašto adresą. Priklausomai nuo organizacijos vykdomos veiklos pobūdžio, vartotojų grupių profiliai yra daug sudėtingesni ir labiau pritaikyti specifiniams sprendimams ir produktams. Profiliuose saugoma ne tik informacija apie vartotojus, tačiau ir duomenys, apibūdinantys ir pačią paslaugą ar siūlomą produktą. Profilių sudarymui papildomai galima apibrėžti tokias vartotojų grupių savybių aibes:

- *Geografinė vartotojų padėtis* – pagal kokius regioninius matavimus bus sudaromos vartotojų grupės.
- *Ekonominiai aspektai* – vartotojų grupavimas pagal pajamas, išlaidas, atsiskaitymo būdus.
- *Amžiaus grupės* – vartotojų skirstymas į amžiaus grupes.
- *Pasaulėžiūra* – atskirų grupių išskyrimas atsižvelgiant į klientų vertybes, tikėjimą, požiūrius į įvairius specifinius su paslaugomis susijusius kriterijus.
- *Kultūra* – skirstymas pagal kilmę, kalbą, nacionalinius papročius.
- *Gyvenimo būdas* – skirstymo kriterijai, susiję su vartotojų gyvenimo būdu.
- *Žinios* – grupių sudarymas remiantis vartotojo žiniomis apie jam teikiamas paslaugas ar siūlomas prekes.
- *Specifiniai kriterijai* – kriterijai tiesiogiai susiję su siūlomomis paslaugomis ir sprendimais, nepapuolantys į bendrąsias grupes.

Pats vartotojo arba vartotojų grupės profilis susideda iš dviejų dalių – *faktinės* ir *elgesio*. Faktinė profilio dalis apima vartotojų grupei būdingą šabloną, kuris apsprendžia vartotojų priklausymą vienam ar kitam profiliui. Elgesio dalis yra susijusi su veiksmingais dalykais, kuriuos reiktų atlikti tuomet, kai yra nusprendžiama, kad esamas vartotojas sutampa su šablonu.

Individualaus vartotojo elgesys gali būti modeliuojamas įvairių konjugtyvinių taisyklių pagalba, įskaitant asociacijas ir klasifikacijų taisykles. Taisyklių vartojimas vartotojų elgesiui apibrėžti turi privalumų, kadangi konjugtyvinės taisyklės yra gerai išstudijuota sąvoka, plačiai naudojama duomenų rinkime. Be to, tyrinėtojai yra pasiūlę daug taisyklių paieškos algoritmų, skirtų asociacijų ir klasifikacijų taisyklėms rasti.

Duomenų rinkimo metodai dažnai sukuria daug taisyklių ir kai kurios iš jų gali būti trivialios, klaidingos ar tiesiog nereikšmingos. Todėl svarbus dalykas po visų taisyklių sukūrimo yra jų patikrinimas ir patvirtinimas. Paprasčiausias sudarytų taisyklių patikrinimo būdas yra peržvelgti visas taisykles ir nuspręsti, kurios taisyklės bus priimtose ir kurios bus atmestos. Tuomet iš atrinktų taisyklių bus suformuotas vartotojų grupės profilis. Tačiau toks taisyklių atrinkimo būdas nėra efektyvus, ypač kai dirbama su dideliais duomenų kiekiais ir sukurtų taisyklių kiekis yra didelis. Todėl yra svarbu yra sistemos sugebėjimas apdoroti turimas taisyklės automatiškai ir leisti ekspertams patikrinti didelius taisyklių kiekius per gan trumpą laiką,

reikalaujant nedidelio ekspertų įsikišimo. Šiam tikslui vartojami šie taisyklių grupavimo ir eliminavimo algoritmai:

- *Panašumais paremtų taisyklių grupavimas.* Panašios taisyklės yra grupuojamos pagal ekspertų nustatytus panašumų kriterijus. Rezultate ekspertai gali peržvelgti taisyklių grupes, kas yra žymiai patogiau negu peržvelgti kiekvieną grupę atskirai, ir gali iškart patvirtinti arba atmesti taisyklių grupes .
- *Panašumais paremtų taisyklių grupavimas.* Taisyklės yra filtruojamos pagal ekspertų nustatytus panašumų kriterijus. Ekspertai gali nustatyti priimtinus ir atmetamus šablonus. Taisyklės, kurios tenkina priimtinių šablonų keliamus reikalavimus yra priimamos ir taisyklės kurios tenkina atmetamų šablonų keliamus reikalavimus yra atmetamos. Taisyklės kurios netenkina šablonų lieka nepatikrintos ir pereina į sekantį taisyklių patikrinimo etapą.
- *Nereikalingų taisyklių eliminavimas.* Šiame metode yra eliminuojamos taisyklės, kurios gali būti išvedamos iš kitų, jau turimų, skirtų bendresniems atvejams, taisyklių ir faktų. Kitaip tariant, yra eliminuojamos taisyklės, kurios savyje neturi jokios naujos informacijos apie vartotojų elgesį.

Naudojantis tokiomis automatizuotomis taisyklių tikrinimo galimybėmis galima per gan trumpą laiką peržiūrėti didelius sukurtų taisyklių kiekius. Ekspertui belieka įvertinti rastas taisyklių grupes ir nuspręsti kada turi baigtis tikrinimo procesas. Po šio patikrinimo proceso mes jau turime galutines taisykles, iš kurių gali būti kuriami vartotojų elgesį atspindintys vartotojų grupių profiliai.

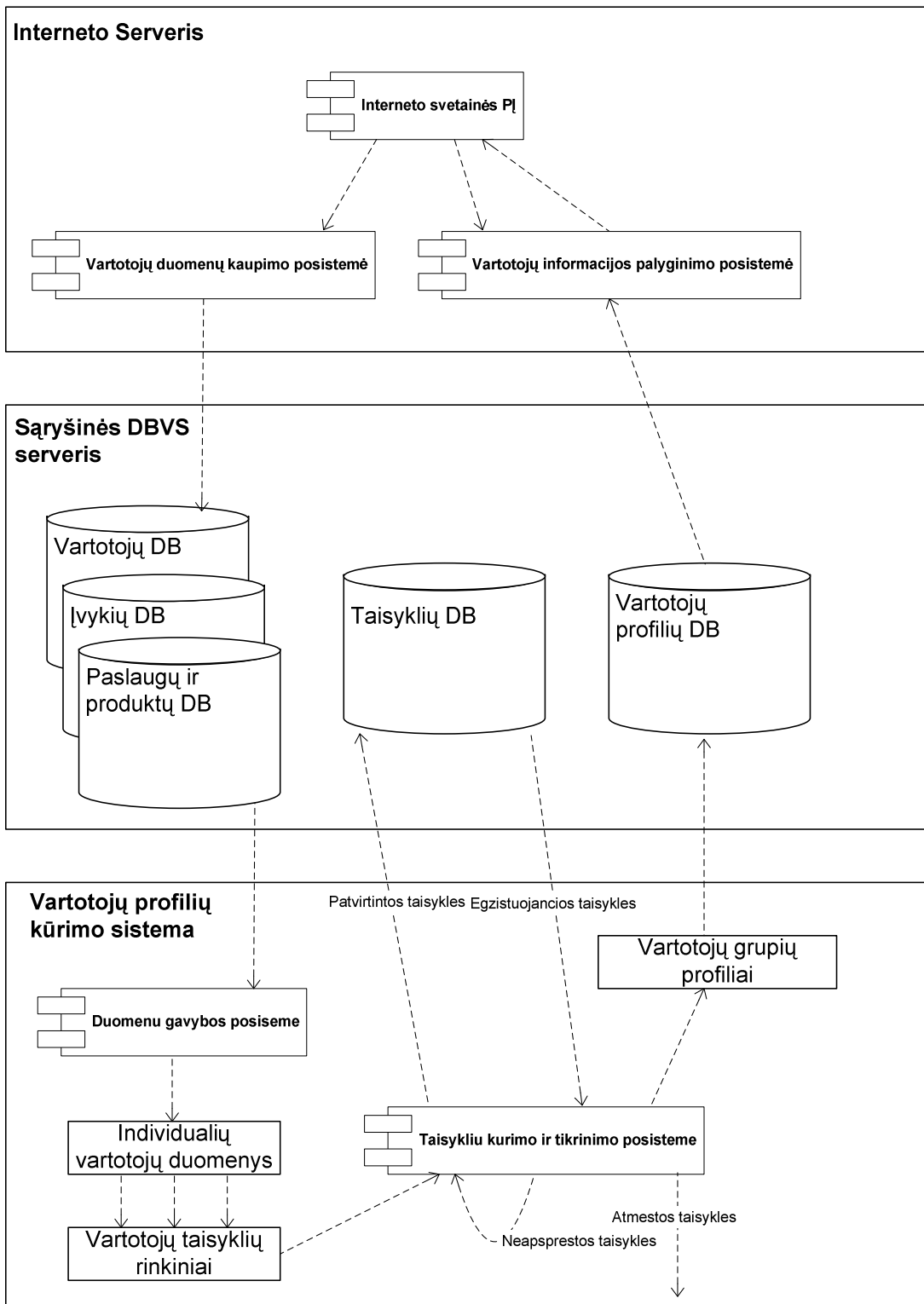
2.2.2. Konceptualusis vartotojų atpažinimo sistemos modelio aprašymas

Vartotojų atpažinimo sistema susideda iš trijų dalių (7 pav.):

- *Interneto serveryje* yra įdiegta svetainės programinė įranga, kuria naudojasi vartotojai. Kartu yra įdiegiamos ir dvi posistemės – tai vartotojų informacijos kaupimo posistemė, kuri atsakinga už vartotojų įvykių informacijos kaupimą ir saugojimą į duomenų bazes, bei vartotojų informacijos palyginimo posistemė, kuri turimai vartotojų informacijai duomenų bazėje ieško profilių atitikmenų ir radus – grąžina juos svetainės programinei įrangai, kuri atlieka tolimesnius sprendimus bendravime su vartotojais.

- *Sąryšinės DBVS serveris.* Visa informacija susijusi su vartotojais yra saugoma sąryšinėse duomenų bazėse. Iš esmės ši sistemos dalis atlieka tik duomenų saugojimą ir pateikimą posistemėms, kurioms jų reikia.
- *Vartotojų profilių kūrimo sistema* yra pagrindinė visos sistemos dalis, kur yra daromi visi sprendimai, susiję su vartotojų profilių kūrimu. Čia visa informacija apie vartotojus, gauta iš DBVS , yra analizuojama, ieškoma taisyklių, kurios atspindi vartotojų elgesį ir pagal tai kuriami vartotojų grupių profiliai, kurie vėliau yra panaudojami sistemai bendraujant su vartotojais.

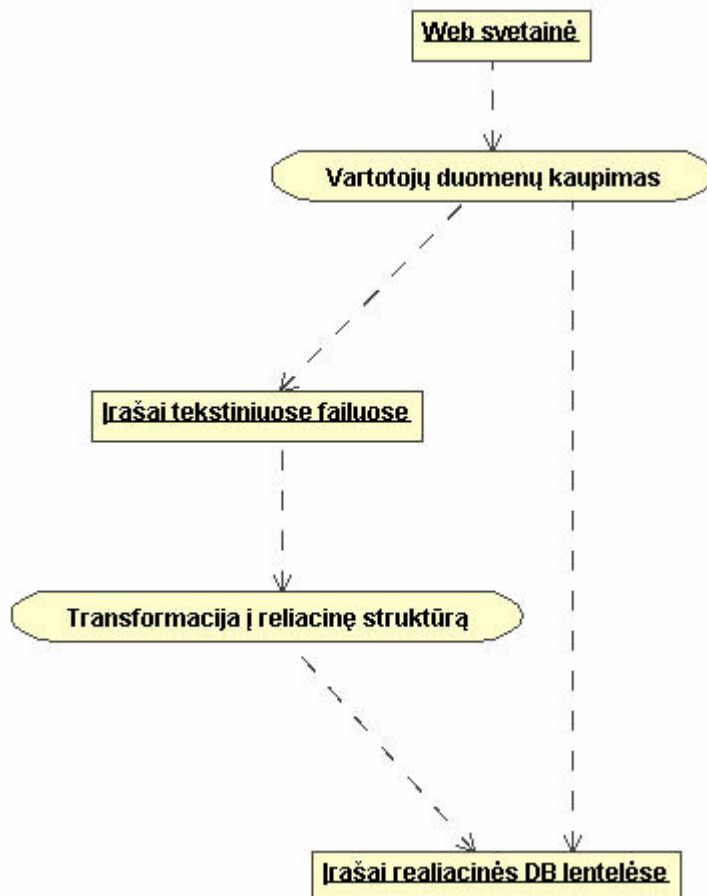
Pirmoji sistemos dalis yra pateikiama labiau supaprastintame variante, kadangi jos tiksli struktūra tiesiogiai yra priklausoma nuo organizacijos programinės įrangos, kurią ji vartoja bendravime su savo klientais. Šiame modelyje parodytos tik pagrindinės jos sudedamosios dalys, kurios yra būtinos norint daryti sprendimus remiantis vartotojų atpažinimu ir profilių jiems kūrimu.



7 pav. Vartotojų atpažinimo sistemos konceptualusis modelis

2.2.3. Duomenų perkėlimo į saugyklą procesas

Kaip buvo minėta anksčiau, visi svetainės vartotojų duomenys yra saugomi sąryšinėje duomenų bazėje. Toliau pateikiamas šių duomenų transformavimo procesas, kurio rezultate duomenys iš sistemos patenka į duomenų bazę. Procesas pavaizduotas UML veiklos diagrama 8 pav.



8 pav. Duomenų perkėlimo seka

Transformavimo žingsnių aprašymas:

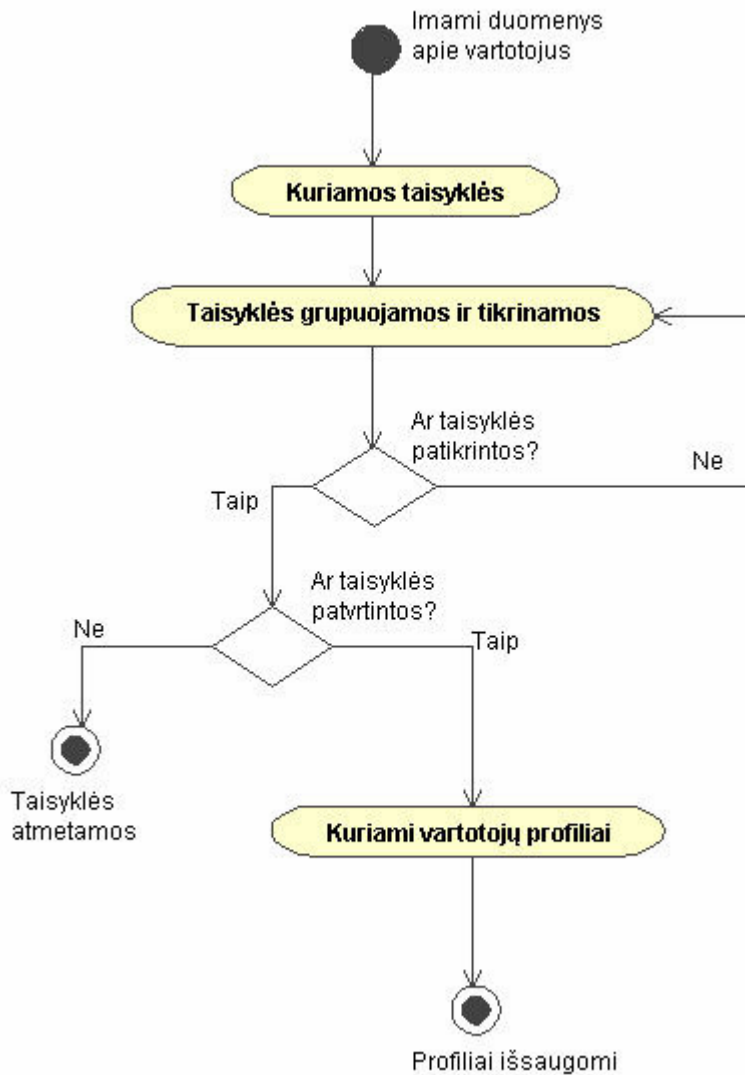
- 1. Vartotojų informacijos kaupimas.** Sistemoje yra kaupiama informacija, susijusi su vartotojų elgesiu ir atliktais veiksmais – fiksuojamas kiekvieno puslapio parodymas, koku metu jis buvo žiūrėtas, iš kur į jį ateita, specifinė vartotojo informacija – naudojama naršyklė, kalba, geografinė vartotojo buvimo vieta ir t.t. Dažniausiai visa ši informacija yra kaupiama tarpiniame taške – tekstiniuose failuose. Tai daroma tam, kad būtų išvengta didelių duomenų bazės apkrovimų. Saugojimas tekstiniuose failuose yra mažiausiai sistemos resursų

reikalaujanti operacija, todėl duomenų išsaugojimo operacijos yra atliekamos greičiausiai. Priklausomai nuo svetainės apkrautumo, duomenys gali būti saugomi tiesiogiai į DBVS sąryšines lenteles. Tokiu atveju duomenys iškart patenka į DB ir tekstinių failų transformavimo į sąryšinę struktūrą žingsnis yra praleidžiamas.

- 2. Transformavimas į sąryšinę struktūrą.** Duomenims, saugomiems tekstiniuose failuose būtinas transformavimo procesas. Tik tada jie gali būti išsaugoti sąryšinėse duomenų bazėse. Tekstiniuose failuose duomenys yra saugomi griežtai apibrėžta tvarka. Kiekvienas naujas įrašas faile yra pradedamas nauja eilute. Eilutėse duomenys skiriami specialiais simboliais (kableliais, kabliataškiais, taškais, tabuliacijos simboliais). Po atlikto transformavimo proceso, DBVS lentelėse turėsime tiek pat naujų įrašų, kiek eilučių buvo tekstiniuose failuose.

2.2.4. Vartotojų profilių kūrimo procesas

Išnagrinėsime vartotojų profilių kūrimo procesą detaliau. Šiame etape iš turimos vartotojų elgesio informacijos yra sukuriami juos apibūdinantys profiliai, kurie vėliau yra vartojami vartotojų elgsenos palyginimui ir tinkamesnių sprendimų jiems priėmimui. Procesas pavaizduotas UML veiklos diagrama 9 pav.



9 pav. Vartotojų profilių kūrimo seka

Kūrimo žingsnių aprašymas:

1. **Taisyklių kūrimas.** Iš sąryšinėse DB saugomos vartotojų elgsenos informacijos yra kuriamos kiekvieno vartotojo elgesį apibūdinančios taisyklės. Visos taisyklės tarpusavyje yra nesusietos.
2. **Taisyklių grupavimas ir tikrinimas.** Šiame etape visos taisyklės, gautos ir taisyklių kūrimo etapo, yra grupuojamos ir tikrinamos. Grupavimas yra skirtas tam, kad būtų galima nesunkiai patikrinti didesnius sukurtų taisyklių kiekius ir išvengti atvejo, kada reikia visas taisyklės tikrinti po vieną. Visos taisyklės, turinčios vienodus atributus, vienodus tikslus ar kitus iš anksto nustatytus panašumo požymius yra priskiriamos tai pačiai grupei. Atlikus grupavimą,

belieka tik patikrinti turimas taisyklių grupes ir nuspręsti, kurios iš jų yra tinkamos tolimesniam vartojimui. Šio etapo gale visos taisyklės yra suskirstomos į tris atskiras grupes:

a) Nepatikrintos taisyklės. Dėl tam tikrų priežasčių nepatikrintos taisyklės yra paliekamos pakartotinam jų grupavimui ir tikrinimui.

b) Atmestos taisyklės yra tolimesniuose etapuose nenaudojamos.

c) Patvirtintos taisyklės patenka į tolimesnį etapą, kur iš jų bus sudaromi vartotojų profiliai.

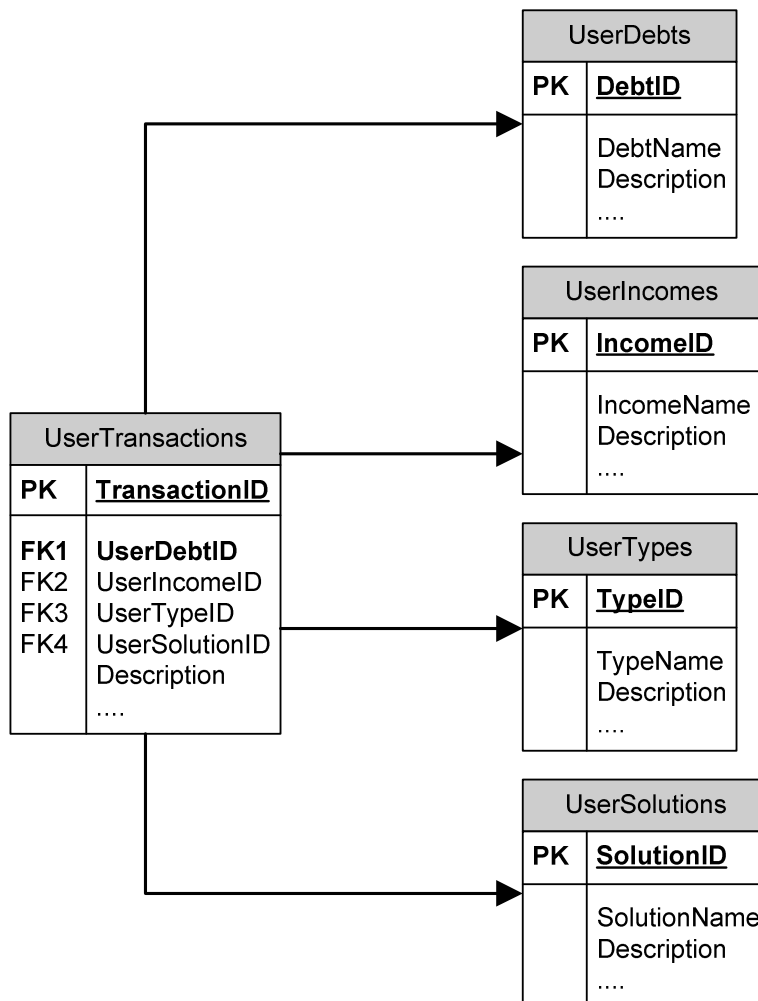
3. **Vartotojų profilių kūrimas.** Šiame etape iš patvirtintų taisyklių yra kuriami vartotojų grupių profiliai. Sukurti profiliai yra išsaugomi sąryšinėse duomenų bazėse.

3. Vartotojų atpažinimo sistemos modelio eksperimentinis tyrimas

Eksperto metu siekėme patvirtinti praktines vartotojų atpažinimo sistemos modelio pritaikymo galimybes. Šiam tikslui vartojome duomenis, nusakančius vartotojų sąveikos elgesį su tam tikra sistema. Pačios sistemos veiklos tikslas šiuo atveju mums nėra svarbus, kadangi pagrindiniai kriterijai, į kuriuos mes kreipėme dėmesį yra patys vartotojų elgesio duomenys, iš jų gaunamos vartotojų elgseną identifikuojančios taisyklės ir jų pritaikymas kitų vartotojų identifikavimui.

3.1. Tiriamųjų vartotojų elgsenos duomenų struktūra

10 pav. pateikiamas tiriamųjų duomenų sąryšinis modelis:



10 pav. Tiriamųjų duomenų modelis

Pagrindiniai tiriamieji duomenys, kuriems mes bandysime pritaikyti vartotojų elgesio atpažinimo modelį yra lentelėje *UserTransactions*. 11 lentelėje pateikiame detalų *UserTransactions* duomenų struktūros aprašymą:

11 lentelė

Stulpelis	Duomenų tipas	Papildoma inf.	Aprašymas
TransactionID	Int	PRIMARY KEY	Transakcijos įrašo unikalus identifikatorius.
UserDebtID	Int	FOREIGN KEY	Skolos tipo identifikatorius
UserIncomeID	Int	FOREIGN KEY	Pajamų tipo identifikatorius.
UserTypeID	Int	FOREIGN KEY	Vartotojo tipo identifikatorius
UserSolutionID	Int	FOREIGN KEY	Sprendimo tipo identifikatorius
Description	Text		Transakcijos aprašymas

Stulpelių *UserDebtID*, *UserIncomeID*, *UserTypeID* ir *UserSolutionID* reikšmės identifikuoja atitinkamus įrašus iš kitų lentelių, todėl *lentelė12* – *15* lentelėse pateikiami atitinkamų lentelių detalūs aprašai.

12 lentelė

UserDebts lentelės aprašas

Stulpelis	Duomenų tipas	Papildoma inf.	Aprašymas
DebtID	Int	PRIMARY KEY	Skolos tipo įrašo unikalus identifikatorius.
DebtName	Char(30)		Skolos pavadinimas
Description	Text		Skolos aprašymas

UserIncomes lentelės aprašas

Stulpelis	Duomenų tipas	Papildoma inf.	Aprašymas
IncomeID	Int	PRIMARY KEY	Pajamų tipo įrašo unikalus identifikatorius.
IncomeName	Char(30)		Pajamų pavadinimas
Description	Text		Pajamų aprašymas

UserTypes lentelės aprašas

Stulpelis	Duomenų tipas	Papildoma inf.	Aprašymas
TypeID	Int	PRIMARY KEY	Transakcijos įrašo unikalus identifikatorius.
TypeName	Char(30)		Skolos pavadinimas
Description	Text		Skolos aprašymas

UserSolutions lentelės aprašas

Stulpelis	Duomenų tipas	Papildoma inf.	Aprašymas
DebtID	Int	PRIMARY KEY	Transakcijos įrašo unikalus identifikatorius.
DebtName	Char(30)		Skolos pavadinimas
Description	Text		Skolos aprašymas

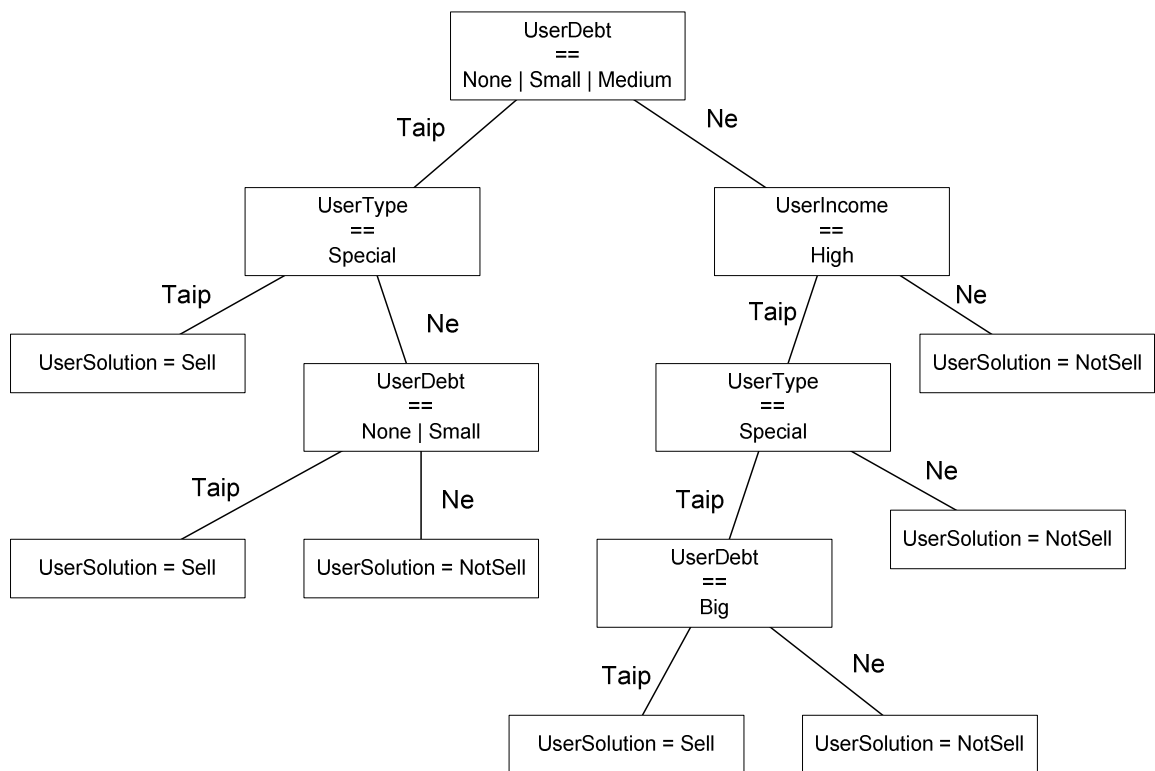
3.2. Eksperimento atlikimo eiga ir rezultatai

Tyrimui vartojome duomenų modelį, kuris susideda iš keturių pagrindinių atributų: *UserDebt*, *UserIncome*, *UserType* ir *UserSolution*. Pagrindinis tikslas rasti taisykles, nustatančias koks bus vartotojo *UserSolution* tipas, remiantis likusių trijų atributų reikšmėmis. Kiekvienas iš atributų turi aibę reikšmių kurios pateikiamos 16 lentelėje:

16 lentelė

Duomenų atributas	Galimos reikšmės
UserDebt	None, Small, Medium, Big, ExtraBig
UserIncome	Low, Medium, High
UserType	Special, Common
UserSolution	Sell, NotSell

Šių atributų kombinacijos iš anksto buvo nežinomos, todėl neturėjome jokios pradinės informacijos apie galimas sprendimų taisyklės, kurios nulemtų vieną ar kitą sprendimą. *Pirmoje* eksperimento dalyje iš 16349 įrašų, sprendimo medžio algoritmo pagalba, buvo gautas sprendimų medis, pateikiamas 11 pav.



11 pav. Sprendimų medis, gautas duomenų analizės metu

Iš šio medžio gavome taisyklių rinkinį, kuris pateikiamas 17 lentelėje.

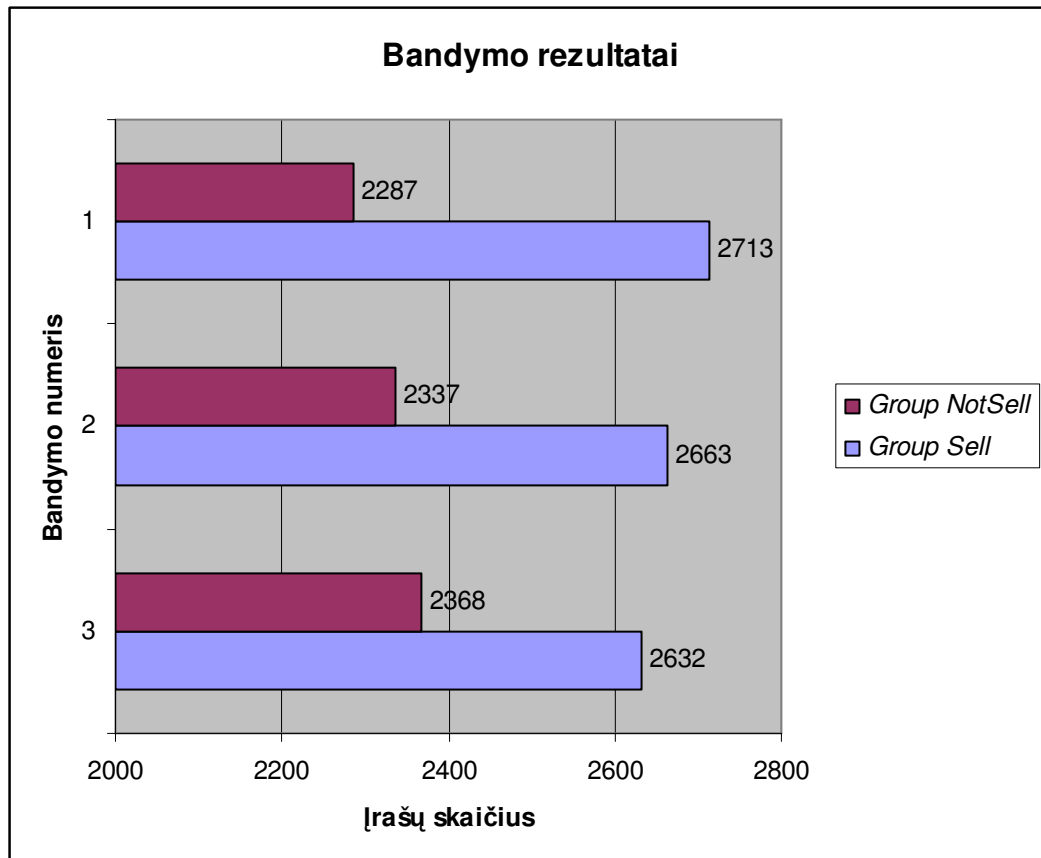
17 lentelė

Nr.	Taisyklė	Sprendimas
1	(UserType == Special) IR (UserDebt == Medium ARBA UserDebt == None ARBA UserDebt == Small)	UserSolution = Sell
2	(UserDebt == None ARBA UserDebt == Small) IR (UserType == Common)	UserSolution = Sell
3	(UserDebt == Medium) IR (UserType == Common)	UserSolution = NotSell
4	(UserDebt == Big) IR (UserType == Special) IR (UserIncome == High)	UserSolution = Sell
5	(UserDebt == ExtraBig) IR (UserType == Special) IR (UserIncome == High)	UserSolution = NotSell
6	(UserType == Common) IR (UserIncome == High) IR (UserDebt == Big ARBA UserDebt == ExtraBig)	UserSolution = NotSell
7	(UserIncome == Low ARBA UserIncome == Medium) IR (UserDebt == Big ARBA UserDebt == ExtraBig)	UserSolution = NotSell

Matome, kad pagal mums svarbų rezultatą, t.y. pagal *UserSolution* atributo reikšmę, mes galime išskirti du variantus: *UserSolution* reišmė yra *Sell* arba *UserSolution* reikšmė yra *NotSell*. Todėl *antroje* eksperimento dalyje sugrupavome panašias taisykles, ir gavome dvi taisyklių grupes. Į pirmąją grupę pateko taisyklės, kurių numeriai yra 1, 2 ir 4. O į antrąją grupę papuolė taisyklės, kurių numeriai yra 3, 5, 6 ir 7. Šioms dviejoms taisyklių grupėms buvo sukurti vartotojų elgesio šablonai. Visos taisyklės, patenkančios į vieną grupę yra vartojamos būsimiems grupės vartotojams identifikuoti. Tai bus *faktinė* vartotojų profilio dalis. *Elgesio* daliai galima priskirti tam tikrus veiksmus, kuriuos sistema turėtų įvykdyti, kai identifikuojama, jog vartotojo duomenys sutapo su profilio *faktinės* dalies turimomis viena ar daugiau taisyklių. Mes priskyrimė nesudėtingus sprendimus – rodyti specialius pranešimus priklausomai nuo atitikusio vartotojo profilio. Jeigu nusprendžiama, kad vartotojas priklauso pirmai grupei, rodomas įrašas „*Group Sell*“, o jeigu antrajai – „*Group NotSell*“.

Trečioje eksperimento dalyje buvo atlikti trys naujų vartotojų atpažinimo bandymai. Kiekvieno bandymo metu dinamiškai buvo generuojami 5000 irrašų, su atsitiktinėmis atributų *UserDebt*, *UserIncome* ir *UserType* reikšmėmis, kas leisdavo sukurti atsitiktinius šių trijų atributų rinkinius. Kiekvienas dinamiškai sukurtas atributų rinkinys buvo lyginamas su turimomis profilių taisyklėmis ir fiksuojamas sutapimas – kuriai vartotojui grupei tinka šis atributų rinkinys. Jeigu atributų rinkinys nesutampa nei su vienu vartotojų grupės šablonu – tada jis priskiriamas neatpažintų vartotojų grupei. Rezultatai pateikiami 12 pav. Atlikus tyrimą paaiškėjo, kad remiantis bandymo duomenimis sudarytos taisyklės yra labai tikslios, todėl bandant jas su atsitiktiniais duomenimis nebuvo irrašų, priskirtų neatpažintų vartotojų grupei.

Išvados. Remiantis eksperimentiniais tyrimo rezultatais galima daryti išvadą, kad toks vartotojų atpažinimo sistemos modelis yra prasmingas. Jo pagalba iš turimų ankstesnės sąveikos duomenų galima nustatyti taisykles, atspindinčias vartotojų elgesį (*pirmoji eksperimento dalis*). Taisykles apjungus į grupes, galima sudaryti panašių vartotojų grupių profilius, apibrėžiančius vartotojų elgesį (*antroji eksperimento dalis*). Sudaryti profiliai gali būti pritaikomi tolimesniam naujų vartotojų identifikavimui ir sprendimų priėmimui (*trečioji eksperimento dalis*). Taigi, toks vartotojų atpažinimo sistemos modelis gali būti sėkmingai vartojamas realiuose projektuose vartotojų grupių nustatymui, jų profilių sudarymui, bei naujų vartotojų identifikavimui.



12 pav. Eksperimento rezultatai

Išvados

1. Darbe teoriškai analizuota vartotojų elgesio informacijos svarba organizacijos veiklai. Nustatyti esminiai kriterijai, turintys įtaką organizacijai sąveikoje su jos vartotojais.
2. Apibrėžti pagrindiniai procesai, kuriuos reikia įgyvendinti norint efektyviai vystyti veiklą, besiremiančia vartotojų identifikavimu ir jų elgesio informacijos panaudojimu tolimesniems sprendimams priimti.
3. Įvertinti pagrindiniai duomenų gavybos metodai, tinkantys vartotojų identifikavimui remiantis ankstesne jų sąveikos su organizacija patirtimi.
4. Pasiūlytas teorinis vartotojų atpažinimo sistemos modelis, įgalinantis efektyviau atlikti vartotojų grupavimą ir jų poreikių identifikavimą.
5. Eksperimentiškai testuotas pasiūlytas teorinis modelis. Patvirtintas jo veikimo principo teisingumas ir tikslingumas.
6. Teorinis modelis tinka realių projektų, veikiančių vartotojų atpažinimo sistemų pagrindu, įgyvendinimui.

Summary

Customer Relationship Management (CRM) is one of the biggest problems for many companies today. By analyzing history records (profiles) of its customers, organization can effectively adapt its business activity to users needs and create better products and services. Proper analysis of customer profiles can help to predict the behaviour of the customers. After grouping customer profiles by similar attributes, company can easier handle its interactions with similar users. Such group profiling can also help to identify needs of new customers on their first interaction with the company.

The biggest problem in implementing such systems is the management of a vast array of customer data. Data mining technologies can help to solve this problem and help the e-businesses to better understand their e-customers. This work reviews data mining methods, such as *Nearest Neighbors*, *Decision Trees* and *Association Rules*, which can be effectively used for customers grouping and profiling. A new conceptual model of *Users Recognition System* is suggested. The new model uses profiles created from customer history records for identifying new customers. The suggested model has been tested experimentally and results prove the possibility of practical application of this model.

Literatūra

- [1] Myths and realities of customer information management and usage in CRM. A Research Report Published by Acxiom® Corporation in Conjunction with Northwestern University and QCi, Inc. [interaktyvus]. August 2002.
[žiūrėta 2003-05-03]. Prieiga per internetą:
http://itresearch.forbes.com/detail/RES/1050332732_388.html
- [2] The CRM Canon Project Group. Using the Appropriate Mix of Marketing Tools to Create and Manage Relationships with Customers. Volume 1 [Interaktyvus] December 2001.
[žiūrėta 2003-05-20]. Prieiga per internetą:
<http://www.cim.co.uk/cim/ser/html/knoFactFile.cfm?KCTopicID=419A2999-5325-4DBC-87DD60D57E81E6F4&objectID=87C9404B-3A3A-4A8A-BA648B9C3AD8F8F3>
- [3] Thomas S. Kraemer. KEYS TO IMPLEMENTING A CRM SYSTEM. Circulation Management [interaktyvus]. November 2002. [žiūrėta 2004-10-20]. Prieiga per internetą:
http://circman.com/ar/marketing_keys_implementing_crm/
- [4] Šileikienė R. Duomenų gavyba – kas tai? Iš *Informacinės technologijos* Nr. 11 [interaktyvus]. 2000, pavasaris [žiūrėta 2004-12-06]. Prieiga per internetą:
<http://www.it.lt>
- [5] Dyche J. The CRM Handbook. A Business Guide to Customer Relationship Management. Addison-Wesley Information Technology Series, 2002
- [6] Rauch J. Mining Association Rules in Medical Data - Two Case Studies. University of Economics, CZ [interaktyvus]. 2003. [žiūrėta 2004-04-10]. Prieiga per internetą:
<http://cyber.felk.cvut.cz/EUNITE03-BIO/pdf/Rauch.pdf>
- [7] Elektroninė parduotuvė amazon.com. [interaktyvus]. [žiūrėta 2004-04-18]. Prieiga per internetą: <http://www.amazon.com>

- [8] Greenberg P. CRM at the Speed of Light: Capturing and Keeping Customers in Internet Real Time. McGraw-Hill Osborne Media; 2nd edition. 2002
- [9] Maurice Frank. The DBMS Guide to Data Mining Solutions.[interaktyvus] [žiūrėta 2004-04-05] Prieiga per internetą:
<http://www.dbmsmag.com/9807m00.html>
- [10] Hertkorn, P. and Rudolph, S. From Data to Models: Synergies of a Joint Data Mining and Similarity Theory Approach, Proceedings SPIE Conference On Data Mining and Knowledge Discovery, Orlando, Florida, April 5-9th, 1999.