

Ville Ylitalo

ASIAKASPOISTUMAN TUTKIMINEN KONEOPPIMISEN AVULLA

Diplomityö
Johtamisen ja talouden tiedekunta
Tarkastajat: Marko Seppänen ja Jari Turunen
Lokakuu 2022

TIIVISTELMÄ

Ville Ylitalo: Asiakaspoistuman tutkiminen koneoppimisen avulla
Diplomityö
Tampereen yliopisto
Johtamisen ja tietotekniikan DI-tutkinto-ohjelma
Lokakuu 2022

Valtavasti kasvaneet datamäärät, tallennuskapasiteetin kasvu sekä tekoälyn kehitys ovat mahdollistaneet data-analyysin entistä monipuolisemman hyödyntämisen asiakassuhteen laadun tutkimisessa. Yritykset ovat heränneet siihen, että uusien asiakkaiden hankkiminen on huomattavasti kalliimpaa kuin vanhoista kiinnipitäminen. Tämän vuoksi moni yritys on alkanut tutkia asiakaspoistumaa. Kohdeyritys toimii rautakauppa-alalla, ja yrityksellä oli tarve luoda automatisoitu malli asiakaspoistuman tunnistamiseksi. Tutkittavan asiakasryhmän kauppa ei perustu hankintasopimuksiin, joten asiakkailla ei ole velvollisuutta ilmoittaa, mikäli he siirtävät ostonsa muualle. Kohdeyritys haluaa päästä hyvissä ajoin selville hiipuvista asiakkaista, jotta ne voidaan yrittää voittaa takaisin, ennen kuin ostot loppuvat kokonaan. Tarkoituksena on luoda koneoppimiseen perustuva malli, joka tuottaisi helposti tulkittavan poistumisriskiprosentin suoraan kohdeyrityksen CRM-järjestelmään.

Työssä keskitytään koneoppimisen mallin muuttujien määrittelyyn sekä mallin toimivuuden validointiin. Yleisimmät muuttujat asiakaspoistuman tutkimisessa ovat niin sanotut RFM-muuttujat, jotka määrittelevät aikaa viimeisestä ostoksesta (recency), ostostiheyttä (frequency) sekä rahallista arvoa (monetary). Aluksi määritettiin tilastollinen asiakaspoistuma. Kohdeyrityksen myyntihistoriadataa tutkimalla määriteltiin varmasti poistuneet asiakkaat. Tilastollisen asiakaspoistuman määrittelyyn käytettävien muuttujien parametrejä säätämällä malli oppi tunnistamaan poistuneet asiakkaat riittävän hyvin, jotta mallia voidaan käyttää varsinaisten poistumaennusteiden tekemiseen. Tilastollisen asiakaspoistuman määrittelyyn muuttujina käytettiin ostostiheyttä sekä painotettua ostostiheyden ja viimeisestä ostoksesta kuluneen ajan avulla laskettua pistelukua (RF-score).

Varsinaisessa ennustemallissa aineisto jaettiin opetus-, validointi- ja testiaineistoon, jotka ovat kaikki eri ajanjaksoilta. Malli koulutettiin opetusaineiston avulla, tulokset vahvistettiin validointiaineiston avulla ja ennuste tehtiin testiaineistolla. Testiaineisto on aina uutta, mallille entuudestaan tuntematonta dataa. Mallia testattiin kolmella eri koneoppimisen algoritmilla, ja niiden tuottamat tarkkuusarvot olivat kaikki kahden prosenttiyksikön sisällä toisistaan. Tarkin algoritmi oli XGboost, joka saavutti testiaineistolla ROC-AUC-arvon 0,8998 sekä F1-scoren 0,8218. Kummatkin arvot osoittavat, että mallin suorituskyky on hyvä ei-sopimuspohjaisessa asiakaspoistuman tutkimisessa.

Mallin tuottama poistumariskiprosentti näkyy suoraan yrityksen CRM-järjestelmässä, ja mikäli prosentti ylittää ennalta määrätyn kynnyksarvon, myyjälle lähtee ilmoitus asiakkaan poistumariskistä. Tämän jälkeen myyjä soittaa asiakkaalle ja yrittää selvittää, onko riski todellinen. Ensisijaisena tarkoituksena on tavoittaa asiakas ennen kuin ostot loppuvat kokonaan. Ensimmäiseen koneoppimisen mallin kehitysversioon ei vielä saatu kaikkia toivottuja muuttujia mukaan. Jatkokehityksenä malliin olisi hyvä saada lisättyä muuttujiksi vastuumyyjän vaihtuminen, aika myyjän viimeisestä yhteydenotosta sekä asiakkaan tuoteryhmäkohtaiset ostokäyttäytymismuutokset. Nämä muuttujat lisäämällä voitaisiin parantaa mallin tarkkuutta ja ennustuskykyä. Laajemmin tarkastellen kohdeyrityksen kannattaa myös tarkastella asiakaskannattavuutta, jotta kannattamattomia asiakkaita ei yritetä houkutella takaisin. Lisäksi yrityksen kannattaa pyrkiä hyödyntämään keräämäänsä dataa esimerkiksi asiakkoitaisten kampanjoiden suunnitteluun sekä asiakkaan elinkaaren arvon määrittelyyn. Tekoälyä hyödyntämällä kohdeyrityksen on mahdollista tehostaa asiakassuhteiden hoitoa ja saavuttaa kannattavampi asiakaspohja.

Avainsanat: asiakaspoistuma, B2B, CRM-järjestelmä, ei-sopimuspohjainen kauppa, koneoppiminen, rautakauppa

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

ABSTRACT

Ville Ylitalo: Studying customer churn with machine learning
Master's thesis
Tampere University
Master's Degree Program in Management and Information Technology
October 2022

The huge growth in quantities of data and storage capacity and the development of artificial intelligence have enabled more versatile usage of data analysis when studying the quality of customer relationship. Companies have become more aware that getting new clients is more expensive than holding on to the old ones. Therefore, many companies have started to study customer churn. The target company is a hardware store, and they had a need to develop an automatized model to recognize customer churn. The study is limited to one customer group, whose purchases aren't based on procurement contracts and therefore the customers are not obligated to notify if they plan to stop doing business with the target company. The target company wants to know as soon as possible about fading customers, so that they can try to win them back before they stop all their purchases. The aim is to create a machine learning model, which could produce an easily interpretable churn risk percent to the company's CRM system.

This study focuses on defining the variables of the machine learning model and assessing if the model works correctly. The most common variables in the study of customer churn are RFM variables, which mean time since the last purchase (recency), frequency of purchases (frequency) and the monetary value of purchases (monetary). First, we determined the statistical churn flag. We studied the company's historical sales data to determine the already churned customers. By adjusting the parameters of the variables used to determine the statistical churn flag the model learned to recognize the churned customers well enough, so that we could use the model to make the actual churn predictions. The variables used to determine the statistical churn flag were recency and weighted RF score, which combines the recency and frequency scores.

In the actual model the data is divided into training, validation and test sets, which are all from different time periods. The model is taught with the training set, the results are confirmed with the validation set and the prediction is made with the test set. The test set is always unprecedented new data to the model. The model was tested with three machine learning algorithms and the accuracy values of these algorithms were all within two percentage point of each other. The most accurate algorithm was XGboost, which achieved the ROC-AUC score of 0,8998 and the F1 score of 0,8218 in the test set. Both values can be considered good when studying non-contractual customer churn.

The churn percent produced by the model can be found in the company's CRM system and if the churn percent exceeds a predefined threshold value a notification to a sales representative is sent. After the notification a sales representative tries to phone the customer to determine if the churn risk is real. The main objective is to reach the client before they stop all their purchases. All the desired variables could not be added to the first development version of the machine learning model. For further development the following variables should be added: the change of the responsible sales representative, time since the last contact by a sales representative and changes in product group based purchase behavior. By adding these variables, the accuracy and the prediction power of the model could be enhanced. In the bigger picture the target company should also examine the customer profitability, so that they do not try to win back unprofitable customers. Also, the company should try to utilize the collected data to plan customer specific campaigns and to determine the value of customers' life cycle. By utilizing artificial intelligence, the company could streamline their customer relationship management and achieve a more profitable customer base.

Keywords: customer churn, B2B, CRM system, hardware store industry, non-contractual sales, machine learning

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

ALKUSANAT

Tässä työssä perehdytään asiakaspoistuman tutkimiseen koneoppimisen avulla rautakaupan toimintaympäristössä. On ollut todella mielenkiintoista perehtyä koneoppimiseen syvällisemmin ja uskon sen sovellutuksilla olevan paljon käyttöä tulevaisuudessa.

Haluan kiittää ohjaajiani sekä yliopiston että kohdeyrityksen puolelta avusta ja ohjauksesta tämän työn valmistumisen eteen. Suurimmat kiitokset ansaitsevat vaimoni ja lapseni kaikesta tuesta ja kannustuksesta koko opiskelujeni ajan sekä siitä, että minun on ylipäätään ollut mahdollista opiskella töiden ohessa. Vaimolleni vielä erityiskiitos korvaamattomasta tuesta diplomityöprosessin aikana. Kiitokset myös ystäväilleni ja läheisilleni tsemppaamisesta ja tiedusteluista miten opinnot etenevät.

Naantalissa 12.10.2022

Ville Ylitalo

SISÄLLYSLUETTELO

1. JOHDANTO	1
1.1 Työn taustaa.....	1
1.2 Tutkimuskonteksti.....	2
1.3 Työn tavoitteet ja rajaukset	5
2. KIRJALLISUUSKATSAUS	7
2.1 Asiakassuhteen hallinta.....	7
2.1.1 Big datan käyttö asiakassuhteen hallinnassa.....	8
2.1.2 CRM-järjestelmä asiakassuhteen hallinnassa.....	11
2.1.3 Myynnin johtamisen uudet roolit ja työkalut.....	14
2.1.4 Asiakassuhteen säilyttäminen	16
2.2 Asiakaspoistuma	18
2.2.1 Aikaisemmat tutkimukset.....	20
2.2.2 Koneoppimisen toimintaperiaate ja sen hyödyntäminen asiakaspoistuman tutkimuksessa	22
2.2.3 Mallien muuttajat ja mallin toimivuuden arviointi	24
2.2.4 Asiakaspoistuman ehkäiseminen	27
2.3 Yhteenveto	27
3. MENETELMÄ, AINEISTON KUVAUS JA ANALYYSI.....	29
3.1 Tutkimusmenetelmä	29
3.2 Aineiston kuvaus ja analyysi	30
4. TULOKSET	40
5. PÄÄTELMÄT JA KEHITYSEHDOTUKSET.....	45
5.1 Päätelmät tuloksista	45
5.2 Kehitysehdotuksia	46
5.2.1 Uudet muuttajat	46
5.2.2 Laajemmat kehityskohteet.....	47
LÄHTEET	49

KUVA- JA TAULUKKOLUETTELO

<i>Kuva 1. Rakentamisen määrän kehitys Suomessa (Rakennusteollisuus, 2020)</i>	3
<i>Kuva 2. Rautakauppojen liikevaihdon kehitys (RaSi Ry, 2022)</i>	3
<i>Kuva 3. Rautakauppatoimialan tuotteiden hintakehitys (RaSi Ry, 2022)</i>	4
<i>Kuva 4. Kannattava asiakaspohja mukaillen Peppers & Rogers (Peppers & Rogers, 2016, s. 5–7)</i>	8
<i>Kuva 5. Big datan viisi V:tä mukaillen Ramadan (Ramadan, 2017)</i>	9
<i>Kuva 6. CRM-prosessi, mukaillen Reinartz et al (W. Reinartz et al., 2004)</i>	12
<i>Kuva 7. CRM prosessi mukaillen Ngai et al. (Ngai et al., 2008.)</i>	13
<i>Kuva 8. Liiketoiminnan analytiikkaekosysteemi (Vidgen et al 2007)</i>	15
<i>Kuva 9. Esimerkki RFM-analyysin tuloksesta (Makhija, 2021)</i>	16
<i>Kuva 10. ACSI-malli mukaillen Buttle & Maklan (Buttle & Maklan, 2019, s.49)</i>	17
<i>Kuva 11. Yrityksen kannattavuuden kehitys asiakkaittain (Pfeifer et al., 2005)</i>	20
<i>Kuva 12. Koneoppiminen mukaillen Tuominen et al. (Tuominen et al. 2019)</i>	22
<i>Kuva 13. Aineiston jakoperusteet mukaillen Alpaydin (Alpaydin, 2021, s.56 - 58.)</i>	23
<i>Kuva 14. Esimerkki ROC-käyrästä (Knuuti, 2021)</i>	26
<i>Kuva 15. Aikajana asiakaspoistuman määrittelystä</i>	29
<i>Kuva 16. Esimerkki RFM-analyysistä (Makhija, 2021)</i>	31
<i>Kuva 17. Esimerkki F-scoresta (Makhija, 2021)</i>	32
<i>Kuva 18. Esimerkki RFM-score (Makhija, 2021)</i>	32
<i>Kuva 19. XGboost-algoritmin sekaannusmatriisi</i>	40
<i>Kuva 20. ROC-kuvaaja</i>	41
<i>Kuva 21. Asiakaspoistumatodennäköisyyksien jakauma</i>	42
<i>Kuva 22. XGBoost algoritmin SHAP-arvot</i>	43

<i>Taulukko 1. Asiakaspoistumaesimerkki mukaillen Buttle & Maklan (Buttle & Maklan, 2019, s. 32)</i>	19
<i>Taulukko 2. Aikaisemmat asiakaspoistumatutkimukset</i>	21
<i>Taulukko 3. Sekaannusmatriisi mukaillen Vafeiadis (Vafeiadis et al., 2015)</i>	25
<i>Taulukko 4. Mallin kouluttamiseen käytetyt muuttujat</i>	30
<i>Taulukko 5. Esimerkki luokamuutosten painoarvoista</i>	33
<i>Taulukko 6. Yrityksen tietojärjestelmästä ajatut luvut</i>	33
<i>Taulukko 7. Analyysiesimerkki</i>	34
<i>Taulukko 8. Poistumariskissä olevat asiakkaat vertailu 1</i>	35
<i>Taulukko 9. Ei-poistumariskissä olevat asiakkaat vertailu 1</i>	35
<i>Taulukko 10. Poistumariskissä olevat asiakkaat vertailumuutosten jälkeen</i>	36
<i>Taulukko 11. Ei-poistumariskissä olevat asiakkaat muutosten jälkeen</i>	37
<i>Taulukko 12. Cohenin kappa-analyysi</i>	38
<i>Taulukko 13. Ennusteessa käytetyt muuttujat</i>	39
<i>Taulukko 14. Eri algoritmien ROC AUC pisteet</i>	39
<i>Taulukko 15. XGBoost-algoritmin metriikka</i>	41

LYHENTEET JA MERKINNÄT

ACSI	American Customer Satisfaction Index, malli joka kuvaa asiakassuhteen laatuun vaikuttavia asioita
AUC	Area Under the Curve
B2B	Business to Business, yritysten tekemää myyntiä yrityksille
B2C	Business to Consumer, yritysten tekemää myyntiä kuluttajille
CRM-järjestelmä	Customer Relationship Management, asiakassuhteen hallintajärjestelmä
NLP	Natural Language Processing, luonnollisen kielen prosessointi, eräs koneoppimisen sovellutus
NPS	Net Promoter Score, nettosuosittelemiseksi asiakastyytyvyyden mittaamiseen
OTIF	On Time, In Full, kuvaa toimitusten oikea-aikaisuutta ja täysimääräisyyttä
RFM-muuttujat	Recency, Frequency, Monetary, kuvaavat tässä aikaa viimeisestä ostoksesta, ostostiheyttä sekä rahallista arvoa
SHAP	SHapley Additive exPlanations, käytetään selittämään koneoppimisen mallien muuttujien vaikutuksia
ROC	Receiver Operator Characteristics

1. JOHDANTO

1.1 Työn taustaa

Yritykset keräävät nykypäivänä valtavat määrät dataa. Ainoastaan pieni osa datasta on valmiiksi strukturoitua ja käyttövalmista. Suurin osa datasta on strukturoimattomassa muodossa, ja sen hyödyntäminen on alkanut kiinnostaa yrityksiä yhä kasvavissa määrin. (Ahvonen et al., 2022; Harbert, 2021; Marr, 2019.) Dataa kerätään useista eri lähteistä ja monessa yrityksessä alkaa olla jo datan analysointiin erikoistuneita työntekijöitä. Tallennuskapasiteetin kasvun myötä dataa pystytään varastoimaan yhä suurempia määriä ja tietokoneiden laskentatehon kasvu sekä tekoälyn ja koneoppimisen hyödyntäminen data-analysissa mahdollistavat entistä suurempien datamäärien käsittelyn. (M. Jones, 2019.)

Samaan aikaan yritykset ovat alkaneet kiinnostua entistä enemmän asiakaskokemuksesta. Miten asiakas kokee asiointinsa sujuneen meillä? Jos tunnistautuu asioidessaan vaikkapa kodinkoneliikkeessä, ei mene montakaan minuuttia, kun kännykkään kilahtaa tekstiviestipyyntö: kuinka todennäköisesti suosittelisit yritystämme ystävillesi ja sukulaisillesi? Tähän on syynsä. Yritykset ovat heränneet siihen, että uuden asiakkaan hankkiminen on huomattavasti kalliimpaa kuin vanhasta kiinni pitäminen (Ascarza et al., 2018; Gallo, 2014). Siksi asiakastyytyväisyys onkin nykyisin usean yrityksen yksi tärkeimmistä mittareista. Asiakastyytyväisyyttä mitataan esimerkiksi nettosuositeluindeksillä. (NPS) (Tullis & Albert, 2013, s.146.)

Kohdeyritys toimii rautakauppa-alalla jälleenmyyjänä. Kasvanut yrityksen keräämä datamäärä ja kiinnostus pitää kiinni olemassa olevista asiakassuhteista loivat pohjan tällekin työlle. Tarve työlle syntyi jo alun perin vuoden 2019 lopulla, kun kohdeyritys havaitsi, että noin kolmasosalla erään asiakasryhmän asiakkaista ostot olivat laskussa. Toimeen käytiin ensin perinteisin menetelmin. Hiipuneille asiakkaille alettiin soittamaan ja selvittämään syytä hiipumiselle. Koska työ oli manuaalista, osa asiakkaista oli ehtinyt siirtää jo kaikki ostot kilpailijalle ennen kuin asiaan ehdittiin reagoida. Tehokas puhelinsoitto kierros tuotti tulosta ja tilanne stabiloitui, kunnes seuraavana vuonna törmättiin samaan ongelmaan uudestaan.

Heräsi kysymys, miten voitaisiin reagoida hiipuviin asiakkaisiin tehokkaammin ja ennen kaikkea nopeammin. Pian kävi selväksi, että taustalle tarvitaan automatisoitu

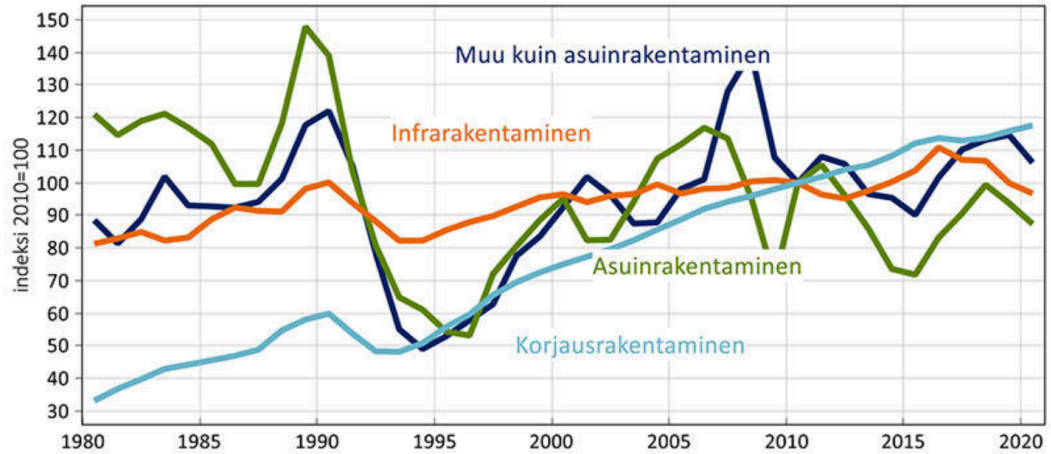
järjestelmä, joka varoittaisi hiipumisuhan alla olevista asiakkaista. Asiakaspoistuma on siis todellinen ongelma, ja sen ehkäisemiseksi haluttiin luoda koneoppimiseen perustuva järjestelmä, joka osaisi ennakoida asiakkaat, joilla on riski kääntää ostot kilpailijalle. Tarkoituksena on laskea helposti ymmärrettävä pisteluku, joka indikoisi riskiä asiakaspoistumalle. Mallin tuottama tulos on tarkoitus viedä yrityksen CRM-järjestelmään, jolloin myyjät näkevät suoraan, mitkä asiakkaat ovat poistumariskissä.

1.2 Tutkimuskonteksti

Rautakaupan asiakaspoistumasta ei tiettävästi ole aiemmin julkaistu tutkimuksia. Alalla on muutamia ominaispiirteitä, jotka on syytä ottaa huomioon tutkimusta tehdessä. Eri-tyispiirteet on kuvattu mahdollisimman tarkasti.

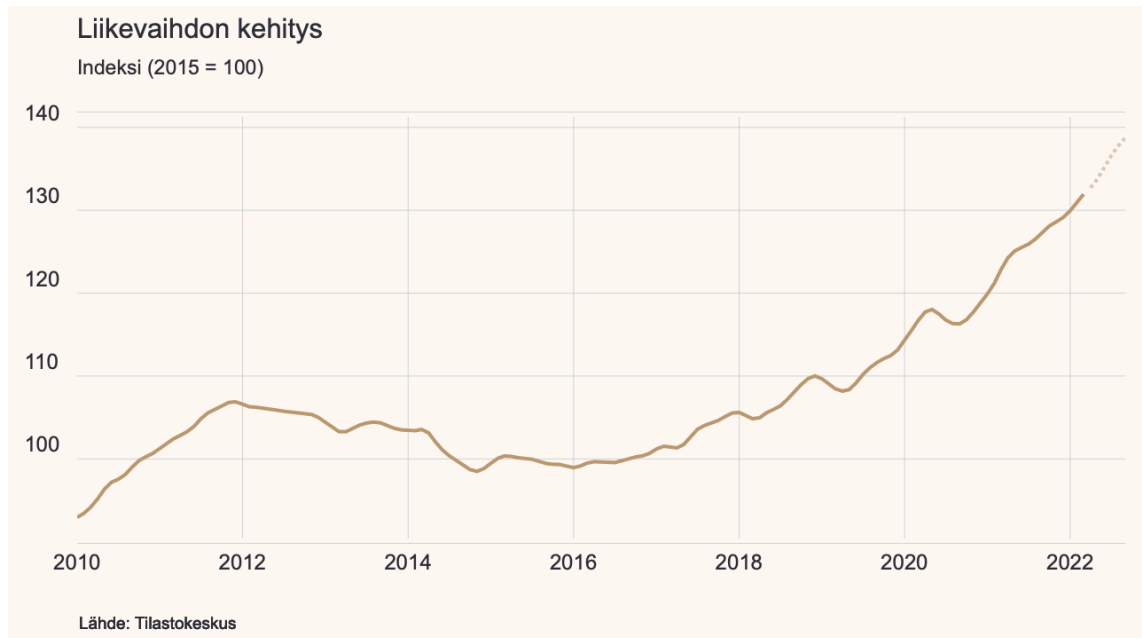
Suomen neljä vuodenaikaa aiheuttavat rakennusalalle kausiluonteisuutta. Erityisesti kulluttajakaupassa tämä näkyy selvästi, mutta se vaikuttaa myös yrityspuolelle. Talvella rakennetaan huomattavasti vähemmän kuin muina vuodenaikoina. Tämä näkyy luonnollisesti myös rautakaupan asiakasmäärissä ja myynnissä. Lisäksi vaihtelua ostokäyttäytymisessä aiheuttaa rakentamisen projektiluonteisuus. Uusi kohde ei välttämättä lähde käyntiin heti vanhan loputtua, tai kohteen alkumetreillä rautakaupan tuotteita ei välttämättä heti tarvita työmaalla. Kausivaihtelu tulee siis ottaa huomioon koneoppimisen mallia suunniteltaessa, jotta asiakkaalle ei soiteta turhaan.

Rakentamisen suhdanteilla on myös suuri merkitys rautakauppojen myyntiin. Kuvassa 1 on esitetty rakentamisen määrän vaihtelu viimeisen 40 vuoden aikana. Suurimmat erot voidaan havaita asuinrakentamisen ja muun rakentamisen kohdalla. Korjausrakentamisen määrä on kasvanut melko tasaisesti 90-luvun pientä laskua lukuun ottamatta.



Kuva 1. Rakentamisen määrän kehitys Suomessa (Rakennusteollisuus, 2020)

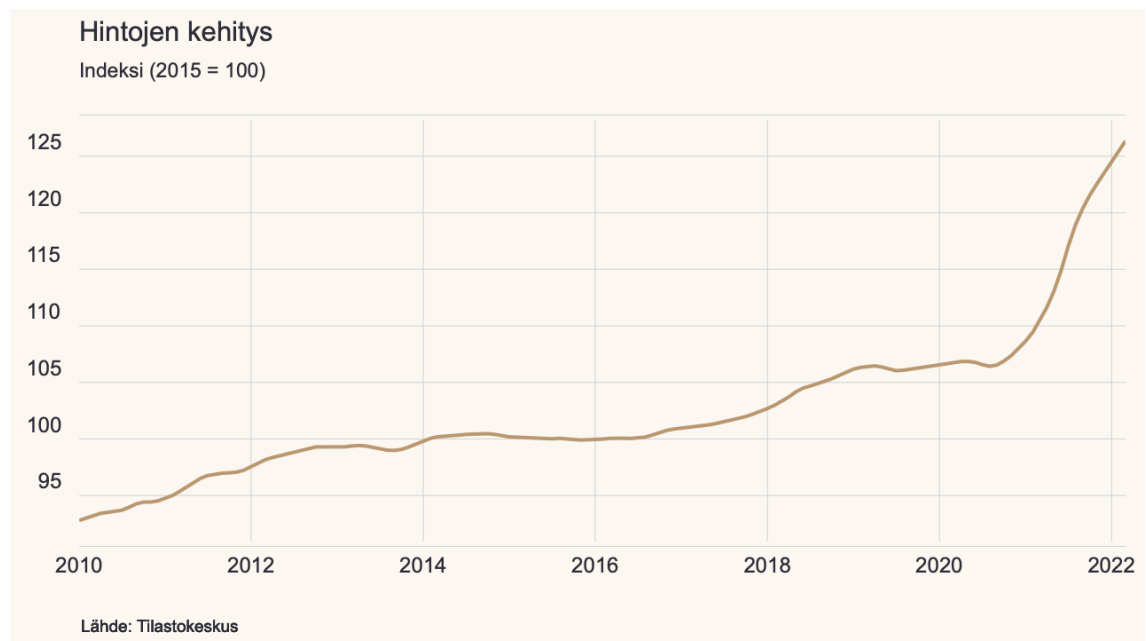
Rautakauppojen myynnin kehitys on ollut vuosikaudet melkoisen tasaista. COVID-pandemia teki tähän muutoksen, kun ihmiset eivät voineet enää matkustaa ja liikkua vapaasti vaan viettivät enemmän aikaa kotona ja mökeillä (Kuva 2). Tämä näkyy hyvin rautakauppojen myynnin kehityksestä. Kasvu on jatkunut vakaana, ja vasta Ukrainan kriisi on tuonut epävakausta markkinaan. Voimakkaat suhdannemuutokset voivat myös omalta osaltaan vaikeuttaa asiakaspoistuman tutkimista erityisesti, jos markkina taantuu.



Kuva 2. Rautakauppojen liikevaihdon kehitys (RaSi Ry, 2022)

COVID-pandemia on aiheuttanut alalla myös ennennäkemättömät hinnan muutokset. Maailmanlaajuiset raaka-aineiden toimitusvaikeudet sekä tiettyjen markkinoiden kova

imu ovat aiheuttaneet hinnan nousupaineita myös Suomen markkinoille (Kuva 3) (Kortelainen, 2022). Hintojen nousun lisäksi pandemian seuraukset ovat heijastuneet rautakauppoihin pidentyneinä toimitusaikoina sekä ajoittaisina saatavuushaasteina. Kaikki alan toimijat eivät välttämättä ole vieneet ennätysuuria hinnankorotuksia täysimääräisenä markkinaa, mikä on osaltaan vaikuttanut asiakkaiden ostokäyttäytymiseen. Osa toimijoista on puolestaan onnistunut tavaran saatavuuden varmistamisessa paremmin kuin toiset. Asiakaspoistuman tutkimisen kannalta nämä seikat ovat vaikeasti huomioon otettavia, mutta tarjoavat luonnollisen syyn, mikäli asiakas on välillä ostanut jotain kilpailijaltakin.



Kuva 3. Rautakauppatoimialan tuotteiden hintakehitys (RaSi Ry, 2022)

Rautakaupan asiakkaat segmentoidaan yleensä kuluttajiin (B2C) ja yrityksiin (B2B). B2B-puoli on jaettu usein vielä tarkempiin asiakasryhmiin, kuten jälleenmyyjiin, kaupunkeihin ja kuntiin, paikallisiin- ja valtakunnallisiin rakennusliikkeisiin ja teollisuuteen. Näistä edellä mainituista asiakasryhmistä jälleenmyyjien, kaupunkien ja kuntien, teollisuuden sekä valtakunnallisten rakennusliikkeiden ostot perustuvat usein hankintasopimukseen tai vahvoihin suosituksiin hankintapaikasta (sopimus pohjainen kauppa). Kuluttajien ja paikallisten rakennusliikkeiden kanssa toimittaessa tällaisia sopimuksia ei yleensä ole vaan hankintapaikkaan vaikuttavat oleellisesti aikaisemmat kokemukset, hinta, laatu, palvelut ja sijainti (ei-sopimus pohjainen kauppa).

Rakennustarvikkeiden myynnissä tyypillistä on se, että asiakas voi ostaa samoja tuotteita useamman toimijan liikkeistä. Tällöin tuotteiden hintavertailu on helppoa ja hinta onkin vahva ostopaikkaa määrittävä kriteeri. Rautakaupat ovatkin alkaneet tuoda

enenevissä määrin omien brändiensä tuotteita markkinoille, jotta pystyisivät paremmin erottautumaan kilpailijoistaan. B2B-kaupassa hinnan lisäksi merkittäviä tekijöitä ovat tuotteen laatu, saatavuus sekä erilaiset logistiset palvelut. Varsinaisilla brändituotteilla on myös suuri merkitys, ja asiakkaat ovat tietyissä tuotteissa vahvasti brändiuskollisia.

Rakennustarvikeliikkeiden valikoimissa on paljon eroja. Osa liikkeistä on keskittynyt pelkästään puupohjaisiin tuotteisiin, osa myy pelkästään kiinnikkeitä, osa pelkästään maa- leja sekä sisustustuotteita, osa on erikoistunut sähkö- ja LVI-tuotteisiin ja niin edelleen. Lisäksi on niin sanottuja täyden valikoiman myymälöitä, joista asiakas pystyy hankki- maan kaikki esimerkiksi omakotitalon rakentamiseen tarvittavat tuotteet. Kaikki liikkeet eivät välttämättä myy kuluttajille tavaraa ollenkaan, vaan keskittyvät pelkästään B2B- kauppaan. Suuresta valikoimien erilaisuudesta johtuen B2B-asiakkailla voi olla tilejä auki useamman ketjun liikkeisiin, jolloin ostopaikan vaihtaminen on helppoa.

Yrityksillä on yleensä eri liikkeissä oma vastuumyyjä, joka hoitaa asiakkaan tilauksia. Varsinkin pidemmät asiakassuhteet henkilöityvät usein, ja mikäli vastuumyyjä eläköityy tai vaihtaa työpaikkaa kilpailijalle, riski sille, että asiakas vaihtaa ostopaikkaa, kasvaa merkittävästi. Vastuumyyjän vaihtumisesta on tehty myös tutkimus, jossa todetaan, että vastuumyyjän vaihtumisella voi olla jopa 13,2 – 17,6 prosentin negatiivinen vaikutus asi- akkaan vuosiosastoihin (Shi et al., 2016).

1.3 Työn tavoitteet ja rajaukset

Tutkimuksen aiheena on tutkia ei-sopimus pohjaisen B2B-kaupan asiakaspoistumaa. Tarkoituksena on luoda koneoppimiseen perustuva malli, joka tunnistaisi poistumaris- kissä olevat asiakkaat ja varoittaisi etukäteen poistumariskistä. Tavoitteena on pienentää kohdeyrityksen asiakaspoistumaa, parantaa asiakassuhteiden hoitoa ja löytää ennalta- ehkäiseviä toimia asiakaspoistuman hillitsemiseen. Varsinainen koneoppimisen malli toi- mii Azuren Databricks-ympäristössä ja se on kirjoitettu Python-ohjelmointikielellä. Työn tulosten analyysissä arvioidaan myös, miten paljon erityisosaamista mallin kehittäminen ja ylläpito vaatii.

Asiakaspoistumaa on aikaisemmin seurattu manuaalisesti asiakkaan ostojen kehitystä seuraamalla. Kohdeyrityksen asiakkaat on jaettu useaan eri ryhmään niiden liiketoimin- nan perusteella, ja aluksi koneoppimisen malli yritetään saada toimimaan yhdellä noin 3000 asiakkaan ryhmällä. Kyseisen asiakasryhmän ostot eivät perustu varsinaisiin han- kintasopimuksiin. Tämän vuoksi on vaikea määrittää, milloin todella on kyse asiakas- poistumasta, koska taustalla ei ole sopimusta eikä asiakas ole velvollinen ilmoittamaan

ostojen lopettamisesta. Lisäksi rakennusalan voimakas kausivaihtelu sekä rakentamisen projektiluonteisuus aiheuttavat lisää haasteita asiakaspoistuman määrittelemiselle.

On tärkeää ymmärtää tutkittavan asiakasryhmän eli paikallisten rakennusliikkeiden ostokäyttäytymistä ohjaavia tekijöitä. Rakennusliikkeillä on enää harvoin omia varastoja ja rautakaupat varastoivatkin usein tarvikkeet niiden puolesta. Näin ollen saatavuus ja toimitusvarmuus nousevat tärkeään rooliin. Pahimmassa tapauksessa työmaa pysähtyy, jos rautakaupasta on jokin tuote lopussa tai työmaatoimitus myöhässä. Rautakaupan sijainnilla työmaahan tai asiakkaan toimipaikkaan verrattuna on myös merkitystä. Tavaraa ei lähdetä hakemaan toiselta puolelta kaupunkia, elleivät rautakaupan palvelu ja valikoima ole poikkeuksellisen hyviä. Ammattilaiset arvostavat hyvän saatavuuden lisäksi erityisesti myös nopeaa asiointia. Varsinaista asiointinopeutta on vaikea tutkia, mutta NPS-luku antaa hyvän kuvan, kuinka tyytyväisiä asiakkaat ovat kaupan palveluun (Tullis & Albert, 2013, s.146). Rautakaupan asiakaspoistumaa tutkittaessa tulee ottaa huomioon myös näiden tekijöiden mahdolliset vaikutukset.

Mikäli koneoppimisen malli saadaan toimimaan luotettavasti, sen käyttöä voidaan laajentaa myös yrityksen muihin asiakasryhmiin. Tehokkaalla asiakaspoistuman hallintamallilla yritys voi saavuttaa kilpailuetua.

Kirjallisuuskatsauksessa käydään ensin läpi asiakassuhteen hallinnan perusteita. Perehdytään, miten big dataa ja CRM-järjestelmää pystytään hyödyntämään asiakassuhteiden hallinnassa. Sen jälkeen tarkastellaan, miten myynnin johdon täytyy muuttaa toimintaansa, jotta valtavasta datamäärästä saadaan maksimaalinen hyöty irti. Lisäksi tarkastellaan asioita, jotka vaikuttavat asiakassuhteen säilymiseen.

Toisessa osiossa perehdytään itse asiakaspoistumaan ja tarkastellaan aikaisemmin tehtyjä tutkimuksia. Seuraavaksi perehdytään koneoppimisen toimintaperiaatteisiin ja datan käsittelyyn. Sitten käydään läpi asiakaspoistuman todennäköisyyden laskemisessa käytettävät muuttujat sekä esitellään mittareita, joiden avulla mallin toimivuutta voidaan arvioida. Lopuksi käydään vielä läpi keinoja asiakaspoistuman ehkäisemiseksi.

2. KIRJALLISUUSKATSAUS

2.1 Asiakassuhteen hallinta

Riippuen alasta ja tutkimuksesta uusien asiakkaiden hankkimisen on arvioitu maksavan noin 5–25 kertaa enemmän kuin olemassa olevien asiakassuhteiden ylläpito. Tästä syystä yrityksissä on alettu entistä enemmän kiinnittää huomiota asiakassuhteen hallintaan. (Gallo, 2014.) Asiakassuhteen hallinnalla pyritään nostamaan yrityksen koko asiakaspohjan arvoa. Asiakassuhteen säilyttäminen luo haasteita monelle yritykselle nykypäivän liiketoimintaympäristössä. Asiakassuhteen säilyttäminen on yksi tärkeimmistä osa-alueista asiakassuhteen hallinnassa ja tärkein komponentti asiakassuhteen elinkaarin arvoa määriteltäessä. (Ascarza et al., 2018.) Asiakassuhteen hallinta on jatkuva prosessi, jossa markkinatietoa hyödyntämällä pyritään rakentamaan ja kehittämään tuottoa maksimoiva asiakasportfolio (Zablah et al., 2004).

Peppers ja Rogers määrittelevät asiakaspohjan arvon nostamisen kuvan 4 mukaisella tavalla. Yrityksen täytyy pyrkiä hankkimaan kannattavia asiakkaita ja pitämään niistä kiinni. Tärkeää on kiinnittää huomiota hiipuviin asiakkaisiin ja pyrkiä voittamaan takaisin niiden joukosta kannattavat asiakkaat. Kannattaville asiakkaille tulisi lisätä mahdollisimman paljon ristiin- ja ylösmyyntiä sekä vähentää palvelu- ja operaatiokustannuksia. Tärkeää on myös pyrkiä hyödyntämään näiden asiakkaiden suositteluja uusien asiakkaiden hankinnassa. Huolehtimalla kaikista kuvassa 4 esitetyistä vaiheista yrityksen asiakaspohja pysyy elinvoimaisena sekä mahdollisimman kannattavana. (Peppers & Rogers, 2016, s. 5–7.)

HANKI

- Hanki kannattavia asiakkaita

PIDÄ

- Pidä kiinni kannattavista asiakkaista
- Voita takaisin kannattavia asiakkaita
- Luovu tuottamattomista asiakkaista

KASVATA

- Myy asiakkaille parempia tuotteita (ylösmyynti)
- Muista ristiinmyynti
- Pyri hyödyntämään suosituksia
- Vähennä palvelu- ja operaatiokustannuksia

Kuva 4. Kannattava asiakaspohja mukaillen Peppers & Rogers (Peppers & Rogers, 2016, s. 5–7)

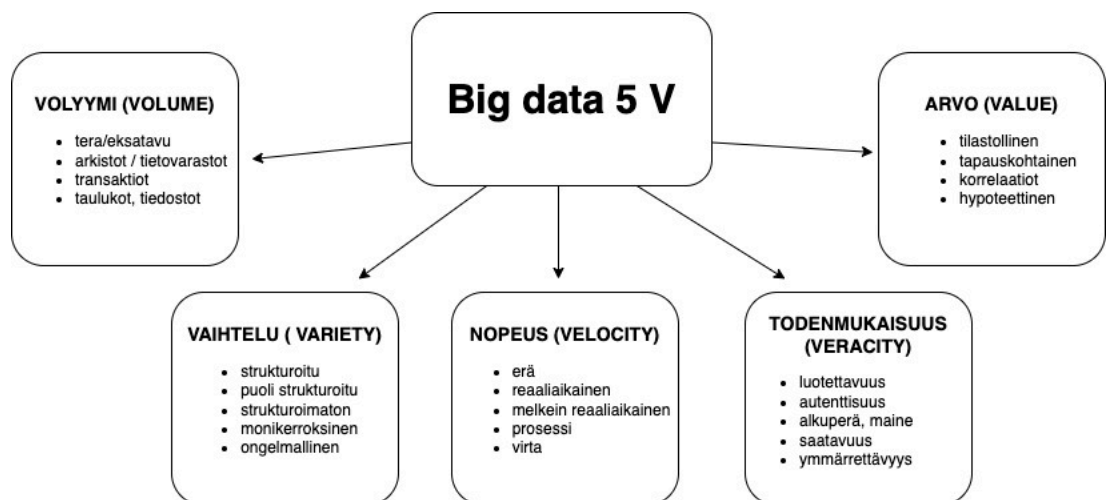
Wiersema toteaa tutkimuksessaan, että asiakas- ja markkinatietoa tarvitaan tulevaisuudessa yhä enemmän asiakkaan muuttuvien tarpeiden ymmärtämiseksi sekä asiakaspoistuman vähentämiseksi (Wiersema, 2012). Myös tietoa asiakkaista on laajemmin saatavilla. Aikaisemmin tietoa saatiin myyjiltä, mutta nykyisin tietoa on lisäksi saatavilla CRM-järjestelmistä, erilaisista sosiaalisen median palveluista sekä data-analyyysien tuloksista. Myyjien tulee tuntea asiakkaan liiketoiminta, jotta he pystyvät perinteisen myynnin lisäksi kehittämään asiakassuhdetta ja tuottamaan asiakkaille lisää arvoa. (Marcos Cuevas, 2018.) Jotta valtavia datamääriä pystytään käsittelemään tehokkaasti ja riittävän nopeasti, tarvitaan käsittelyyn tekoälyn apua. Tekoälyn käyttöönotto tulee helpottamaan ja muuttamaan asiakassuhteiden hallintaa merkittävästi. (Cao et al., 2021; Chatterjee et al., 2021; Mustak et al., 2021.) Kehittyvät teknologiat luovat uusia mahdollisuuksia vuorovaikutukseen asiakkaiden kanssa ja mahdollistavat samalla entistä tarkemman ja personoidumman datan keräyksen asiakkaista (Gupta et al., 2020).

2.1.1 Big datan käyttö asiakassuhteen hallinnassa

Erilaiset business intelligence -järjestelmät ja niihin liittyvät data-analytiikan sovellukset ovat nousseet tärkeään rooliin viimeisen kahden vuosikymmenen aikana. Tarjolla olevan datan määrä on valtavassa kasvussa. (M. Jones, 2019.) Dataa kertyy niin verkkopohjaisista palveluista kuin mobiililaitteiden ja antureiden tuottamana. Enää ei puhuta

gigatavuista vaan tera- tai jopa eksatavun mittakaavoista. Käyttöön onkin vakiintunut termi big data. (H. Chen et al., 2012.) Vaikka termi big data onkin trendikäs muotisana niin tutkijoiden kesken kuin yrityksissäkin, sillä on edelleen useita eri määritelmiä. Lukuisat määritelmät ovat vaikeuttaneet aiheen tutkimusta ja tehokasta kehittämistä. (De Mauro et al., 2015; Schoenherr et al., 2005.) Termiä big data käytetään usein mainostamaan fiksumpaa ja merkityksellisempää data-analyysiä. Harva yritys kuitenkaan lopulta kykenee hyödyntämään koko big datan tarjoaman potentiaalin. Aidosti big dataa hyödyntävät yritykset käyttävät reaaliaikaista tietoa ymmärtääkseen ympärillään tapahtuvia asioita ja hyödyntävät tätä tietoa tuottaakseen uusia tuotteita ja palveluita sekä reagoivat muutoksiin heti niiden tapahduttua. (Davenport et al., 2012.)

Big dataa voidaan luonnehtia sen viiden ulottuvuuden avulla (Kuva 5). Volyymi (volume) kuvaa datan määrää. Valtava datan määrä voi aiheuttaa yrityksille haasteita datan varastoinnin ja käsittelyn kanssa. (Ishwarappa & Anuradha, 2015.) Vaihtelulla (variety) tarkoitetaan, että data voi olla strukturoitua, puoliksi strukturoitua tai strukturoimatonta. Strukturoitu data on helppoa analysoida ja laittaa tietokantaan, kun taas strukturoimaton data on satunnaista ja sen hyödyntäminen vaatii paljon työtä. (Sagiroglu & Sinanc, 2013.) Yli 90 prosenttia datasta on strukturoimatonta. Data-analyysin sujuvuuden kannalta on erittäin oleellista ottaa huomioon, millaista dataa on käytettävissä. (Ishwarappa & Anuradha, 2015.)



Kuva 5. Big datan viisi V:tä mukailen Ramadan (Ramadan, 2017)

Nopeudella (velocity) tarkoitetaan jatkuvasti kehittyvää nopeutta, jolla dataa syntyy. Tämä aiheuttaa myös vaatimuksia yrityksen datan prosessoinnille. Data pitäisi pystyä prosessoimaan mahdollisimman reaaliaikaisesti suurimman hyödyn saavuttamiseksi. (Demchenko et al., 2013; Jabbar et al., 2020.) Valtavan datan määrän ja lähteiden

kasvun myötä on tullut tarpeelliseksi myös tarkastella datan todenmukaisuutta (veracity) (Ishwarappa & Anuradha, 2015). Datan todenmukaisuutta määriteltäessä täytyy ottaa huomioon kaksi näkökulmaa; datan johdonmukaisuus, joka voidaan määritellä tilastollisella luotettavuudella, sekä datan luotettavuus, johon vaikuttavat datan alkuperä, keruu ja käsittelymenetelmät (Demchenko et al., 2013). Arvolla (value) tarkoitetaan sitä lisäarvoa, jonka yritys saa hyödyntämällä dataa prosesseissaan, analyyseissaan ja ennusteissaan. Arvon maksimoiseksi yritykselle täytyy olla selvää, mitä data-analyysin avulla halutaan saavuttaa, jotta osataan kerätä oikeanlaista dataa. (Günther et al., 2017.)

Siinä, miten yritykset hyödyntävät big dataa, on edelleen suuria eroja. Datan kerääminen ei ole enää suuri rasite, vaan enemmän vaivaa joudutaan näkemään datan prosessoinnissa ja analysoinnissa. (Zhang et al., 2020.) Organisaatiot ovat alkaneet tekemään merkittäviä investointeja kyetäkseen luomaan lisää arvoa datan avulla. Strukturoimattomassa datassa piilee paljon mahdollisuuksia, joilla yritys voi muuttaa liiketoimintaansa ja saavuttaa kilpailuetua. Näiden mahdollisuuksien hyödyntäminen korostuu dynaamisessa ja nopeatempoissa liiketoiminnassa, jossa täytyy kyetä tekemään nopeita dataan perustuvia päätöksiä. (Mikalef et al., 2019.)

Big datan hyödyntäminen edellyttää uusien teknologioiden käyttöönottoa. Ne pystyvät käsittelemään isoja määriä nopeasti liikkuvaa dataa. (Libai et al., 2020; Mikalef et al., 2020.) Strukturoimattoman datan analysointi vaatii paljon resursseja, ja ohjelmistotekniikan kehitys on mahdollistanut tekoälyn ja koneoppimisen käyttämisen data-analyysin apuna (Duan et al., 2019). Tehokkaalla data-analyysillä yritykset voivat personoida ja kustomoida myyntitoimintoja sekä asiakaspalvelua, mikä auttaa lujittamaan asiakassuhteita. Datasta on myös mahdollista tunnistaa asiakkaiden odotuksia ja ennakoida asiakkaiden tarpeita. (Hallikainen et al., 2020.)

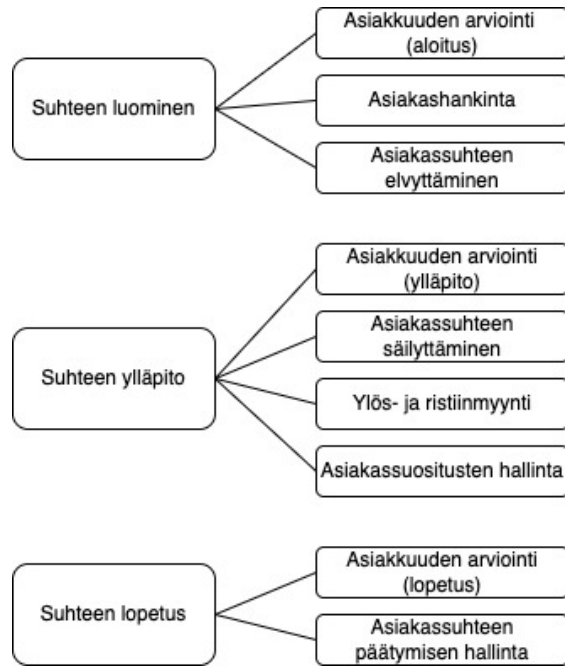
Yrityspuolella hyödynnetään vielä tällä hetkellä hyvin vähän sosiaalisen median dataa asiakastiedon rikastuttamisessa. Yritysten julkiset sosiaalisen median tilit eivät ole parhaita syvemmän asiakastiedon lähteitä, mutta yksityisten tilien tiedoissa on paljon hyödyntämätöntä potentiaalia, joka voi olla merkityksellistä tulevaisuudessa. (Lipiäinen, 2015.)

Big datan hyödyntäminen vaatii muutakin kuin investointeja teknologiaan. Yritysten pitäisi kyetä palkkaamaan sellaisia ihmisiä, joilla on hyvä tekninen- ja johtamisymmärrys big datasta. (Mikalef et al., 2020.) Datan valtavan määrän vuoksi erityisesti datan laatuun pitää kiinnittää huomiota. Näin ollen on erittäin tärkeää, että datan kanssa työskentelevillä henkilöillä on myös liiketoimintaymmärrystä ja ennen kaikkea käsitys, miten tiedon avulla voidaan saavuttaa kilpailuetua. (Chiang et al., 2018; Surbakti et al., 2020.)

2.1.2 CRM-järjestelmä asiakassuhteen hallinnassa

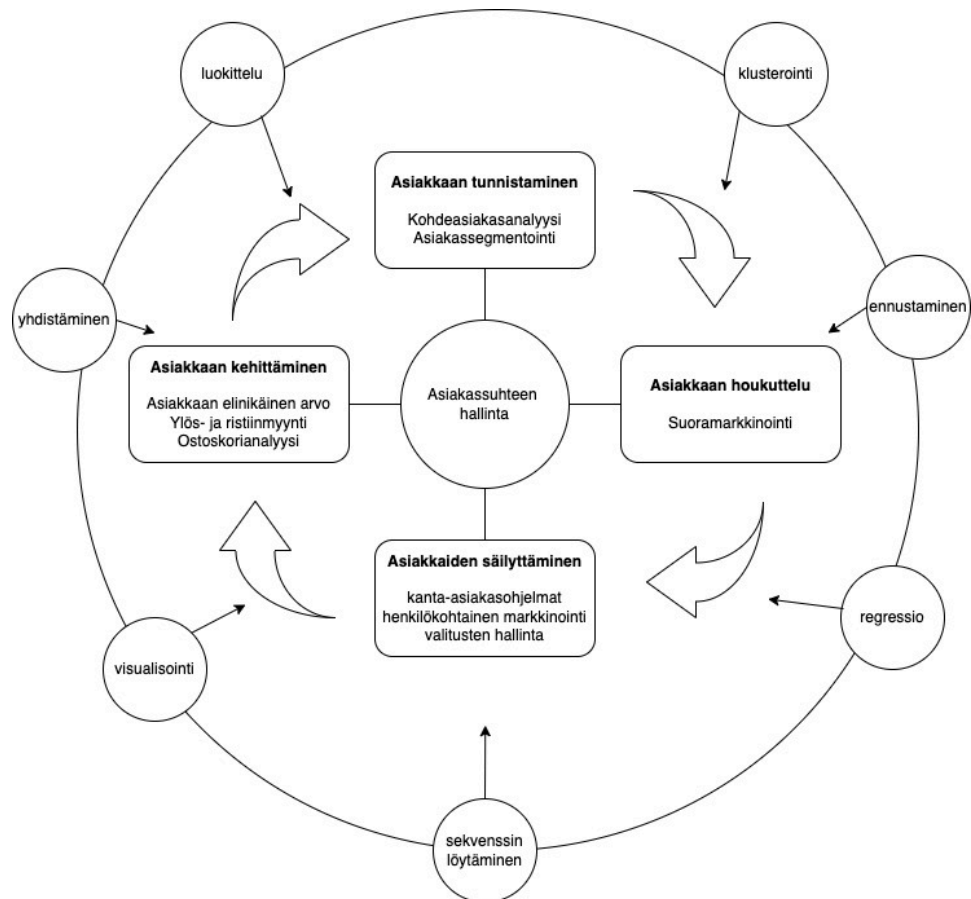
Asiakassuhteen hallinta (CRM) on ihmisten, prosessien ja teknologian yhdistelmä, jonka avulla pyritään ymmärtämään yrityksen asiakkaita. CRM-järjestelmällä tarkoitetaan asiakkuudenhallintajärjestelmää. Oleellisessa roolissa on asiakassuhteiden säilyttäminen ja niiden kehittäminen. (I. J. Chen & Popovich, 2003.) CRM:n avulla saadaan kasvatettua asiakkaan elinkaaren arvoa (King & Burgess, 2008). Osa yrityksistä kokee CRM-investoinnit vain satsauksena teknologiaan tai ohjelmistoon, osa puolestaan kaukonäköisemmin vakaiden asiakassuhteiden kehittämiseen. CRM voidaan määritellä ainakin kolmella tapaa, suppeasti ja taktisesti erityisenä teknisenä ratkaisuna, laajana valikoimana asiakaslähtöisiä it- ja internetratkaisuja tai strategisena ja kokonaisvaltaisena ratkaisuna, joka korostaa asiakassuhteiden hallintaa. (Payne et al., 2005.) On tärkeää löytää CRM-prosessit, jotka vaikuttavat positiivisesti yrityksen suorituskykyyn ja tuottavuuteen (W. Reinartz et al., 2004).

Asiakassuhteet kehittyvät vaiheittain ajan kuluessa, ja näin ollen Reinartz et al. jakaa CRM-prosessin kolmeen eri vaiheeseen: suhteen luomiseen, suhteen ylläpitoon ja suhteen lopettamiseen (Kuva 6). Heidän näkökulmansa perustuu asiakaslähtöiseen tasoon. Jokaisen kohdan ensimmäinen vaihe on asiakkuuden arviointi. Ensimmäisessä vaiheessa keskitytään asiakashankintaan sekä hiipuneiden asiakkaiden aktivointiin. Toisessa vaiheessa pyritään ylläpitämään sekä kehittämään asiakassuhteita, ja kolmannessa vaiheessa pyritään hallitsemaan päättyviä asiakassuhteita. (W. Reinartz et al., 2004.) Kaikki prosessin vaiheet tähtäävät asiakasportfolion arvon maksimointiin. Oleellisinta prosessissa on tunnistaa yritykselle potentiaaliset ja kannattavimmat asiakkaat ja lähteä määrätietoisesti kehittämään niitä asiakassuhteita. (Richards & Jones, 2008.)



Kuva 6. CRM-prosessi, mukailten Reinartz et al (W. Reinartz et al., 2004)

Ngai puolestaan jakaa CRM:n neljään vaiheeseen, joiden tarkoitus on luoda syvempi ymmärrys asiakkaista ja maksimoida asiakasarvo pitkällä aikavälillä. Prosessista muodostuu suljettu sykli, joka kuvaa systemaattista työtä, jota yritysten tulee tehdä kartuttaakseen asiakastuntemusta (Ngai et al., 2008). Vaiheet on kuvattu kuvassa 7. Suurimpana erona Reinhartzin CRM-prosessiin on se, että Ngai katsoo prosessin olevan suljettu sykli. Ngai on luonut asiakkaiden houkuttelusta täysin oman vaiheen, jossa potentiaalisia asiakkaita yritetään houkuttaa suoramainonnalla yrityksen asiakkaiksi. Lisäksi Ngai et al. alleviivaa erityisesti asiakassuhteen säilyttämisen olevan CRM:n keskeinen prosessi. (Ngai et al., 2008.)



Kuva 7. CRM prosessi mukailten Ngai et al. (Ngai et al., 2008.)

CRM on ennen kaikkea myynnin päivittäinen työkalu eikä pelkästään asiakasrekisteri tai aktiivisuusmittari. Asiakkaan historiatietojen avulla voidaan selvittää paremmin asiakkaiden todelliset tarpeet. CRM-järjestelmää käyttämällä saavutetaan parempi kyky kohdistaa toimenpiteitä kannattaviin asiakkaisiin. (Ascarza et al., 2018.) Asiakaspoistuman tutkiminen linkittyy vahvasti CRM:n vaiheisiin: asiakassuhteen elvyttämiseen sekä asiakassuhteen säilyttämiseen (Zerbino et al., 2017).

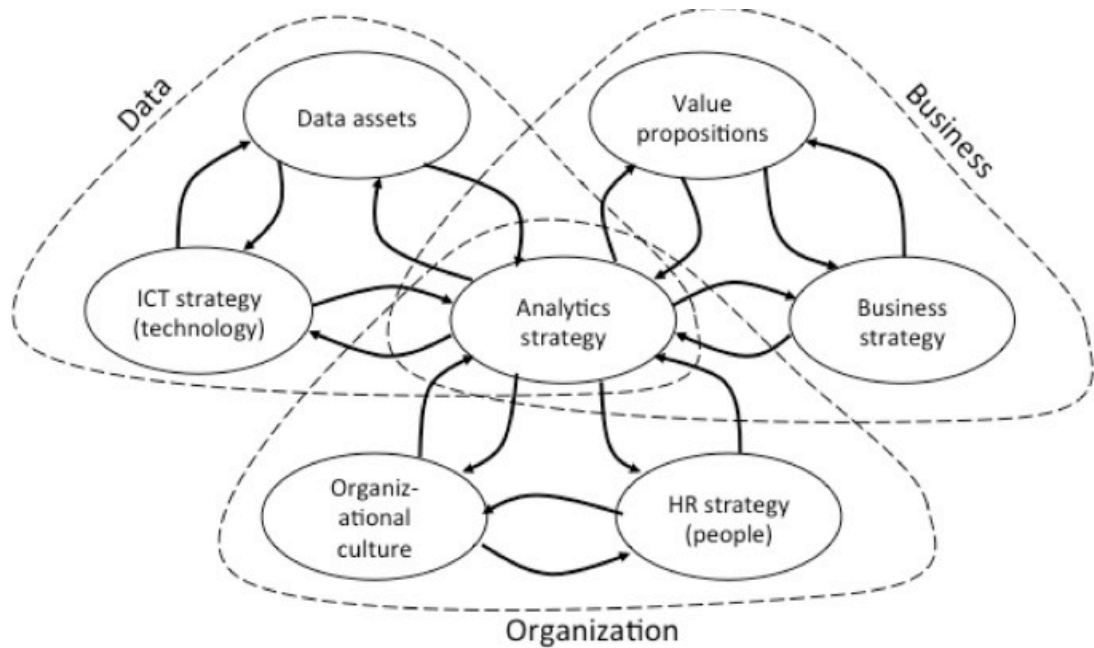
Seuraava taso perinteisestä CRM-järjestelmästä on tekoälyyn pohjautuva CRM (AI CRM), jossa toimintoja pyritään automatisoimaan entistä enemmän (Chatterjee et al., 2021). Jotta tekoälyä pystytään täysipainoisesti hyödyntämään, CRM-järjestelmään tarvitaan mahdollisimman paljon dataa. Big datan aikakaudella tämä on tullut mahdolliseksi, ja CRM-järjestelmään saadaan kerättyä tietoa, joka mahdollistaa asiakaskokemuksen paremman ymmärtämisen sekä asiakkaan palveluiden paremman kustomoinnin. (Anshari et al., 2019.) Tekoälyä voidaan hyödyntää asiakashankinnassa sekä asiakassuhteen kehittämisessä ja ylläpidossa. Sen avulla voidaan ennustaa asiakassuhteen elinkaaren arvo, joka auttaa asiakkaiden priorisoinnissa. (Libai et al., 2020.)

2.1.3 Myynnin johtamisen uudet roolit ja työkalut

Digitalisaatio haastaa myyntiorganisaatioita maailmanlaajuisesti muuttamaan toimintatapojaan. Muutospaineet kohdistuvat erityisesti myynnin johtoon, jotta he saavat tiiminsä toimimaan uudella tavalla ja pystyvät hyödyntämään maksimaalisesti datan luomat mahdollisuudet. (Wengler et al., 2020.) Tärkeintä on oppia pois vanhasta ja päästä eroon ajattelutavasta ”näinhän tämä on aina tehty”. Liikaksi vanhoista toimintatavoista kiinni pitäminen voi aiheuttaa vaikeuksia uusien toimintatapojen omaksumiselle (Mattila et al., 2021).

Myynnin johdon tehtävä on mahdollistaa onnistumiset sekä antaa myyjille sopivassa suhteessa vastuuta ja vapautta. Data-analytikot saavat johdettua datasta uusia potentiaalisia asiakasryhmiä, mahdollisia hiljaisia signaaleja tai kehityskohteita sekä nousevia trendejä. Myynnin johdon tehtävänä on välittää tarvittava tieto myyntitiimeilleen. (Jokiniemi, 2020.) Myyjien tulee pystyä saamaan käsitys asiakkaan tarpeista pystyäkseen ehdottamaan uusia ideoita, rakentamaan sitoutumista sekä tuottamaan asiakkaalle pitkäaikaista arvoa (Mattila et al., 2020).

Yritykset tarvitsevat selkeän data- ja analytiikkastrategian sekä oikeat ihmiset viemään dataohjattua kulttuuria eteenpäin saavuttaakseen kilpailuetua. Esimerkki analytiikkastrategian integroimisesta yrityksen ekosysteemiin on esitetty kuvassa 8. Tärkeimpiä vaatimuksia dataan perustuvassa johtamisessa ovat riittävän laadukas data, data-analyysin käyttö päätöksenteon tukena, olemassa oleva big data - ja analytiikkastrategia, datan saatavuus ja organisaation riittävät datan hyödyntämistaidot. (Vidgen et al., 2017.)



Kuva 8. Liiketoiminnan analytiikkaekosysteemi (Vidgen et al 2007)

CRM:stä saatava data on tärkeintä ja kattavinta dataa, mitä yrityksellä on tarjota myynnin johdon käyttöön. Tämä korostuu eritoten B2B-puolella, jossa yritykset pyrkivät pitkäkestoihin asiakassuhteisiin. (Stein et al., 2013.) CRM-systeemi on usein yrityksen tiedonkeruun keskus, johon kerätään dataa sekä yrityksen ulko-, että sisäpuolelta. Tämä data luo myynninpäälliköille edellytykset johtaa omia tiimejään. (Stein & Smith, 2009.) Data-analytiikkaa myynnin johtamisessa hyödyntämällä on saavutettavissa selkeää myynnin kasvua sekä asiakassuhteen lujittumista (Hallikainen et al., 2020).

B2B-puolella yleisiä myynnin johdon seurantamittareita ovat esimerkiksi asiakassuhteen kesto, aika viimeisestä ostoksesta, asiakkaan ostostiheys ja asiakkuuden rahallinen arvo (Kandeil et al., 2014). Koneoppimisen avulla voidaan tuottaa malli, joka seuraa automaattisesti näiden mittarien kehitystä ja ilmoittaa negatiivisista muutoksista myynnin johdolle tai suoraan myyjälle. RFM-analyysi mittaa aikaa asiakkaan viimeisestä ostoksesta, ostosten välistä tiheyttä sekä rahallista arvoa. Esimerkki RFM-analyysistä on esitetty kuvassa 9. Tällaisen mallin avulla saadaan luokiteltua asiakkaat, helpotettua myynnin johdon työtä ja parannettua reagoitinopeutta. (Paschen et al., 2019.) Tekoälyn ja koneoppimisen avulla saadaan luotua nopeammin tietoa johdon ja konsulttien päätöksenteon tueksi. Tästä on etua erityisesti epävarmoissa tilanteissa, joissa usein joudutaan tekemään nopeita päätöksiä. (Wamba et al., 2017.)



Kuva 9. Esimerkki RFM-analyysin tuloksesta (Makhija, 2021)

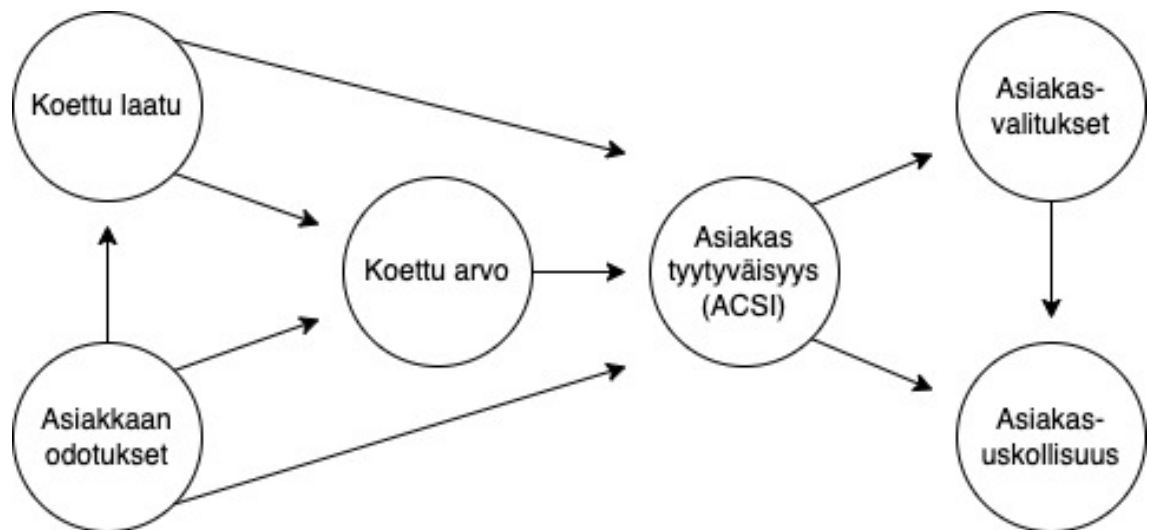
B2C-puolella data-analyysin käyttö on viety pidemmälle kuin B2B-puolella. Tämä johtuu osittain B2C-puolella tarjolla olevasta suuremmasta datamäärästä. (Valluri, 2019.) Moni yritys kerää nykypäivänä valtavat määrät asiakas- ja tuotedataa. Erilaiset ostoskorianalyysit ovat yleisiä erityisesti ruokakaupan puolella. Ostoskorianalyysillä pyritään löytämään usein yhdessä ostettuja tuotteita. Tätä tietoa pyritään hyödyntämään esimerkiksi kampanjoiden suunnittelussa. Analyysin avulla voidaan saada lisäksi tietoa esimerkiksi tuotesijoittelusta, hinnoittelusta, kampanjoiden toimivuudesta sekä tuottavuudesta. Riskinmyyntiä voidaan tehostaa asettelemalla usein yhdessä myytyjä tuotteita samalle alueelle. (Charlet Annie & Kumar, 2012.)

2.1.4 Asiakassuhteen säilyttäminen

Yrityksen kyky pitää asiakkaistaan kiinni sekä asiakkaiden uskollisuus vaikuttavat oleellisesti yrityksen tuottavuuteen. Uskollisuudella tarkoitetaan asiakkaan aikomusta käyttää yrityksen palveluita uudestaan tulevaisuudessa. (Dick & Basu, 1994.) Korkea asiakasuskollisuus on edellytys yrityksen menestykselle. Uskollisuus ja luottamus kehittyvät ajan

saatossa ja näin ollen mitä pidempään asiakassuhde kestää sitä suurempaa tuottoa yritys voi saada. (Larsson & Broström, 2020.)

Asiakastyytyväisyydellä ja asiakassuhteen säilymisellä on vahva korrelaatio. (W. J. Reinartz & Kumar, 2000). Asiakastyytyväisyyteen vaikuttavat oleellisesti tyytyväisyys yrityksen palveluihin, tuotteiden laatuun, kokemus siitä onko, rahoille saatu vastinetta, sekä miten yritys hoitaa asiakkaiden valitukset. Tyytyväiset asiakkaat sitoutuvat paremmin käyttämään yrityksen tuotteita tai palveluita ja heidän asiakasuskollisuutensa on korkeampi. (Gustafsson et al., 2005.) Kuvassa 10 on esitetty ACSI-malli (American Customer Satisfaction Index), joka kuvaa asioita, jotka vaikuttavat asiakasuskollisuuden kehittymiseen (Buttle & Maklan, 2019, s. 49).



Kuva 10. ACSI-malli mukailien Buttle & Maklan (Buttle & Maklan, 2019, s.49)

Asiakassuhteen alkuvaiheessa palvelun laadulla on suurempi merkitys asiakastyytyväisyyteen. Suhteen kehittyessä tuotteen laatu alkaa vaikuttaa puolestaan enemmän. (Johnson et al., 2006.) Mitä suurempaa arvoa asiakas kokee yrityksen palveluista saavansa, sitä tyytyväisempi ja uskollisempi hän on yritystä kohtaan (Tsai et al., 2010).

Asiakassuhteen säilymisen kannalta myös markkinoiden kilpailutilanteella on vaikutusta. Markkinoilla, joilla on paljon tarjontaa ja kilpailua, yritykset joutuvat kiinnittämään asiakastyytyväisyyteen erityistä huomiota, koska vaihtamiskustannukset ovat pienet. Vähemmän kilpailuilla markkinoilla vaihtokustannukset ovat isot ja asiakastyytyväisyys ei ole niin merkittävä tekijä. Vaihtokustannusten suuruudella on iso merkitys asiakas-pyyvyyteen. (M. A. Jones et al., 2002.) Erityisesti B2B-puolella vaihtokustannukset näyttävät merkittävää roolia, koska ostaminen on ammattimaisempaa. Mikäli asiakas vaihtaa toimittajaa, hän joutuu opettelemaan uudet toimintatavat. Tähän kuuluu aikaa,

jotta toiminta kehittyy samalla tasolle, kuin se on aikaisemmin ollut. (P. Y. Chen & Hitt, 2002.)

Nykyisin puhutaan paljon myös asiakaskokemuksesta, jolla tarkoitetaan sitä, miten asiakas on kokenut asiain yrityksen kanssa. Asiakaskokemukseen vaikuttavat sekä suorat että epäsuorat kontaktit yritykseen. Suorilla kontakteilla tarkoitetaan tuotteiden tai palveluiden ostotapahtumaa ja epäsuorilla taas esimerkiksi törmäämistä yrityksen mainontaan tai vaikkapa kuulopuheeseen yrityksen palveluista. (Meyer & Schwager, 2007.) Rautakaupan kontekstiin sijoitettuna asiakaskokemukseen vaikuttavat esimerkiksi tuotteiden saavutettavuus, valikoiman laajuus, palvelun sujuvuus myymälässä ja noutopihalla, myyjien aktiivisuus, toimitusten oikea-aikaisuus ja sisällön oikeellisuus sekä asiointinopeus.

Säännöllisellä yhteydenpidolla asiakkaisiin sekä erilaisilla kanta-asiakasohjelmilla on todettu olevan positiivinen vaikutus asiakassuhteen kehitykseen ja säilymiseen. Myös suoramainonnan on katsottu vaikuttavan positiivisesti. (Verhoef, 2003.) Kim et al. muistuttavat kuitenkin, että se, miten yritys ymmärtää säännöllisen yhteydenpitovälin, ei välttämättä ole sama, miten asiakas kokee asian. Liian tiheällä yhteydenpidolla voi olla negatiivisia vaikutuksia asiakassuhteen kehitykselle, ja se on myös yrityksen resurssien hukakäyttöä (Kim et al., 2012).

2.2 Asiakaspoistuma

Asiakaspoistuma voidaan määritellä tilanteena, jossa asiakas on lopettanut tuotteiden tai palveluiden oston yritykseltä (Gordini & Veglio, 2017; Jahromi et al., 2014). Asiakas voidaan katsoa menetetyksi, kun viimeisestä ostosta on kulunut riittävän pitkä aika. Riittävän pitkälle ajalle ei ole selkeää määrittelyä, vaan se riippuu alasta. Menetetystä asiakkaasta aiheutuu menetettyjen tuottojen lisäksi myös markkinointikustannuksia sekä uusien asiakkaiden hankintakustannuksia. (Patil et al., 2018.) Asiakaspoistumariskiä ilmenetään usein prosenttiluvulla: mitä korkeampi luku, sitä suuremmassa riskissä asiakas on poistua (De Caigny et al., 2018). Taulukossa 1 on esitetty asiakaspoistuman vaikutus kahden esimerkkiyrityksen asiakasmäärään mukaillen Buttlen ja Maklanin esimerkkiä (Buttle & Maklan, 2019, s.32). Oletuksena on, että yritykseen tulee 50 uutta asiakasta vuosittain. Viiden prosentin asiakaspoistumalla yrityksen asiakasmäärä viiden vuoden jälkeen 991 asiakasta, kun taas vastaava luku kymmenen prosentin asiakaspoistumalla on 811 asiakasta. Asiakaspoistumaa vähentämällä voidaan vaikuttaa yrityksen asiakasmäärään kehitykseen. On hyvä myös huomata uudisasiakashankinnan

merkitys. Oletuksella, että yritykseen tulee 50 uutta asiakasta vuosittain, asiakasmäärän kehitys on viiden vuoden jälkeen negatiivinen viiden prosentin asiakaspoistumalla.

Taulukko 1. Asiakaspoistumaesimerkki mukaillen Buttle & Maklan (Buttle & Maklan, 2019, s. 32)

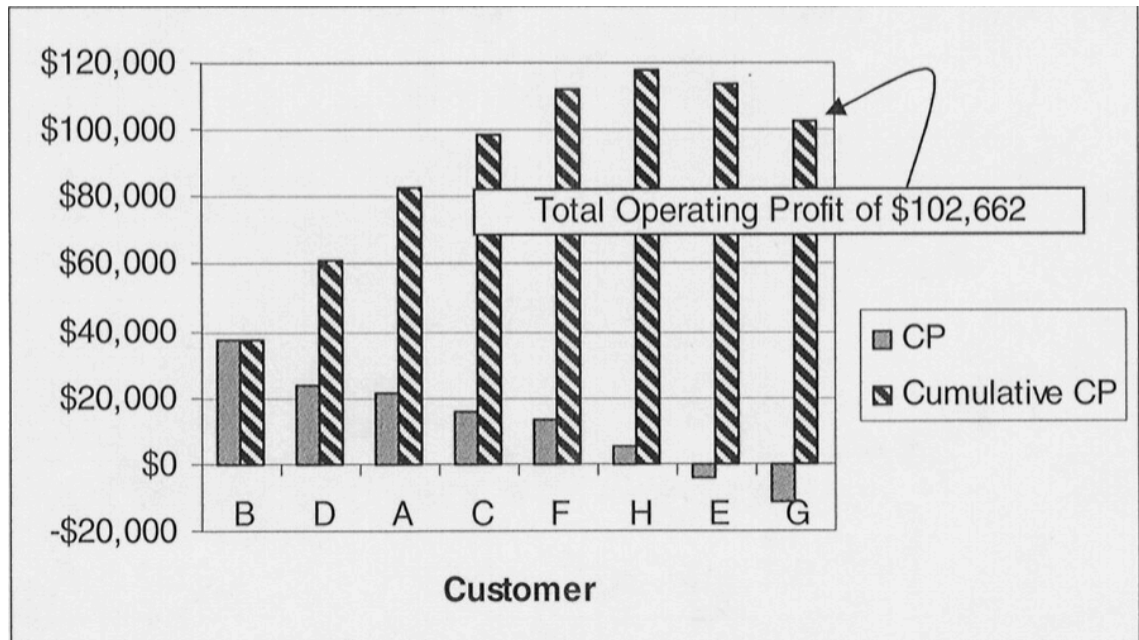
Vuosi	Yritys A (5 % asiakaspoistuma)			Yritys B (10 % asiakaspoistuma)		
	Olemassa olevat	Uudet	Kaikki	Olemassa olevat	Uudet	Kaikki
	asiakkaat	asiakkaat	asiakkaat	asiakkaat	asiakkaat	asiakkaat
1	1000	50	1050	1000	50	1050
2	998	50	1048	945	50	995
3	995	50	1045	896	50	946
4	993	50	1043	851	50	901
5	991	50	1041	811	50	861

Asiakaspoistuma on erilainen sopimuspohjaisessa ja ei-sopimuspohjaisessa kaupassa. Sopimuspohjaisessa kaupassa asiakaspoistuma on helppo todentaa, kun asiakas irtisanoo sopimuksen, kuten esimerkiksi puhelinliittymän. Ei-sopimuspohjaisessa kaupassa poistuman määrittely on hankalampaa, koska asiakkaalla ei ole velvollisuutta ilmoittaa ostojen loppumisesta vaan yrityksen pitää päätellä mahdollinen poistuma asiakkaan käyttäytymisen perusteella. (Ringbeck et al., 2019.) Ei-sopimuspohjaisessa kaupassa onkin tämän vuoksi usein syytä tarkastella osittaista asiakaspoistumaa, jolla tarkoitetaan, että asiakas lopettaa tietyn tuoteryhmän ostot ja siirtää ne muualle (Buckinx & Van Den Poel, 2005).

Asiakaspoistuman taustalla voi olla useita syitä. Ne voidaan jakaa yritykselle hallittavissa oleviin tai hallitsemattomiin syihin (Braun & Schweidel, 2011). Hallittavia syitä voivat olla esimerkiksi tuotteen tai palvelun hinta ja laatu, houkuttelevat kilpailijan tarjoukset, matalat vaihtokustannukset, yleinen asiakastyytymättömyys, kommunikaatio-ongelmat tai jakelukanava ongelmat (Jahromi et al., 2014; Raab et al., 2016, s. 81-82). Hallitsemattomia syitä voivat puolestaan olla asiakkaan konkurssi tai eläköityminen, asiakkaan toimipaikan sijaintimuutos, asiakkaan elämäntapamuutokset tai taloudelliset muutokset (Braun & Schweidel, 2011; Peppers & Rogers, 2016).

Asiakaspoistuman tunnistaminen on tärkeää. Silti on tunnistettava ero niiden asiakkaiden välillä, jotka ovat riskissä lähteä, ja niiden, joihin kannattaa kohdistaa asiakassuhteen jatkamistoimenpiteitä. Nämä eivät ole välttämättä samoja asiakkaita. (Ascarza et al., 2018.) Tästä esimerkkinä kuva 11, jossa on esitetty Pfeifer et al:n laatima laskelma

yrittäjien eri asiakkaiden kannattavuudesta. Huomattavaa kuvassa on, että E- ja G-asiakkaiden kannattavuus on kääntynyt negatiiviseksi, ja kääntänyt samalla koko yrityksen kannattavuuden laskuun. (Pfeifer et al., 2005.) Luonnollisestikaan E- ja G-tyyppisiä asiakkaita ei kannata houkutella takaisin, mikäli ne olisivat lähdössä.



Kuva 11. Yrityksen kannattavuuden kehitys asiakkaittain (Pfeifer et al., 2005)

2.2.1 Aikaisemmat tutkimukset

Asiakaspöytä on tutkittu selvästi enemmän B2C-puolella. Erilaiset koneoppimisen mallit asiakaspöytän tutkimuksessa ovat tulleet mukaan vahvemmin 2010-luvulla. Eniten tutkimusta on tehty telekommunikaation ja erilaisten kuukausimaksupohjaisten palveluiden parissa. B2B-puolen tutkimusta on tehty verraten vähän. (Figalíst et al., 2019.) Valluri (2019) on vertaillut väitöskirjassaan B2C- ja B2B-asiakaspöytän tutkimisen eroja ja toteaa suurimmaksi syyksi vähäiselle B2B-puolen tutkimukselle asiakkaiden ja datan pienemmän määrän verrattuna B2C-puoleen. (Valluri, 2019.) B2B-puolen tutkimuksen haasteina ovat lisäksi useamman henkilön liittyminen ostoprosessiin ja tutkijoiden puutteet alan tuntemuksessa (Lilien, 2016).

Aikaisemmissa tutkimuksissa (Buckinx & Van Den Poel, 2005; Clemente-Císcar et al., 2014; Gatterman-Itschert & Thonemann, 2021; Miguéís et al., 2013; Ringbeck et al.,

2019) on tunnistettu niin sanottu osittainen asiakaspoistuma, jolla tarkoitetaan juuri tiettyjen tuoteryhmien ostojen kääntämistä muualle. Tutkimukset osoittavat, että on erittäin merkityksellistä päästä tilanteeseen kiinni jo tässä vaiheessa, koska vielä asiakassuhteessa oleva asiakas on helpompaa voittaa takaisin verrattuna asiakkaaseen, jonka ostot ovat loppuneet kokonaan (Gatterman-Itschert & Thonemann, 2021).

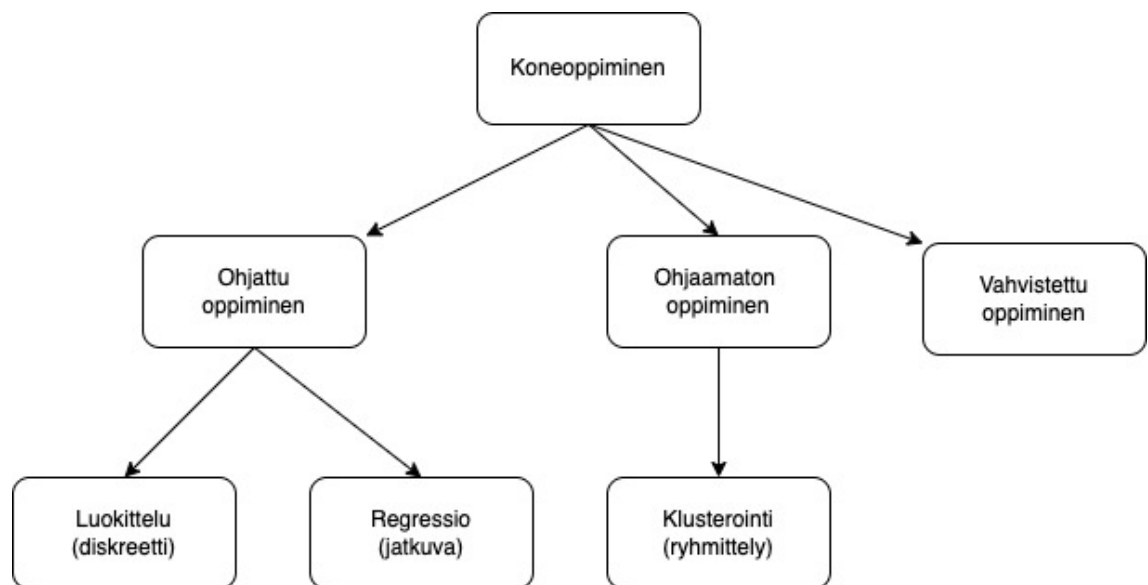
Rautakauppojen asiakaspoistumaa käsitteleviä tutkimuksia ei tiettävästi ole tehty aikaisemmin. Gatterman-Itschert ja Thonemann ovat tehneet tutkimuksen ei-sopimus pohjaisen B2B-tukkukaupan asiakaspoistumasta vuonna 2021. Tutkijoiden mukaan heidän tutkimuksensa on ensimmäinen laajempi kenttätutkimus ei-sopimus pohjaisessa B2B-kaupassa (Gatterman-Itschert & Thonemann, 2021). Sopimus pohjaiseen B2C-kauppaan liittyviä tutkimuksia on tehty huomattavasti enemmän. Taulukossa 2 esitellään kirjallisuudesta löydettyjä tutkimuksia. Taulukossa on mainittu myös kyseisissä tutkimuksissa toimivimmaksi havaittu koneoppimisen malli. Näihin malleihin perehdytään tarkemmin seuraavassa luvussa.

Taulukko 2. Aikaisemmat asiakaspoistumatutkimukset

<u>Tekijät</u>	<u>ala</u>	<u>sopimus- pohjainen</u>	<u>tyyppi</u>	<u>toimivin malli</u>
Buckinx & Van Den Poel, 2005	FMCG jälleenmyynti	ei	B2C	satunnaismetsä
Burez & Van den Poel, 2007	Maksu-TV	kyllä	B2C	satunnaismetsä
Coussement & Van Den Poel, 2008	Tilauspalvelut	kyllä	B2C	tukivektorikone
Mutanen et al., 2010	Pankkiala	kyllä	B2C	logistinen regressio
Lu et al., 2014	Telekommunikaatio	kyllä	B2C	boosting
Jahromi et al., 2014	Jälleenmyynti	ei	B2B	boosting
K. Chen et al., 2015	Logistiikka	kyllä	B2B	päätöspuu
Idris & Khan, 2016	Telekommunikaatio	kyllä	B2C	satunnaismetsä
Gordini & Veglio, 2017	Verkkokauppa	ei	B2B	tukivektorikone
Óskarsdóttir et al., 2018	Telekommunikaatio	kyllä	BC2	logistinen regressio
Ascarza et al., 2018	Ammattijäsenyydet	kyllä	B2C	logistinen regressio
Rautio, 2019	Ohjelmistoliiketoiminta	kyllä	B2B	tukivektorikone
Ringbeck et al., 2019	Jälleenmyynti	ei	B2C	satunnaismetsä
Martínez et al., 2020	Valmistava teollisuus	ei	B2B	gradient boosting
Gatterman-Itschert & Thonemann, 2021	Tukkukauppa	ei	B2B	satunnaismetsä

2.2.2 Koneoppimisen toimintaperiaate ja sen hyödyntäminen asiakaspoistuman tutkimuksessa

Koneoppimisella tarkoitetaan tietokonepohjaista mallia, joka oppii itse pohjatiedon tai kokemuksen kautta esimerkiksi ennustamaan asioita. Koneoppimisen menetot ovat kehittyneet valtavasti viimeisen kahden vuosikymmenen aikana. Tekoälyn kehitys on tuonut koneoppimisen hyödyntämisen osaksi arkipäivää. Yleisimpiä sovelluksia ovat konenäkö, puheentunnistus, luonnollisen kielen tunnistus, robotiikka sekä muut sovellukset. (Jordan & Mitchell, 2015.) Koneoppiminen voidaan jakaa ohjattuun, ohjaamattomaan sekä vahvistettuun oppimiseen (Kuva 12). Ohjattu oppiminen tarkoittaa sitä, että konetta opetetaan syöte-tavoiteparien avulla. Tavoitteena on, että kone osaa tehdä ennusteen opetusaineiston perusteella. Ohjaamattomassa oppimisessä selkeää lopputulosta ei tiedetä, vaan datasta yritetään tunnistaa korrelaatioita, suhteita ja samankaltaisuuksia. Vahvistetussa oppimisessä kone saa joko positiivista tai negatiivista palautetta ja oppii sen perusteella vähentämään negatiivisen palautteen määrää. (Tuominen et al., 2019.)



Kuva 12. Koneoppiminen mukailen Tuominen et al. (Tuominen et al. 2019)

Ohjaavassa oppimisessä malli koulutetaan opetusaineiston avulla. Aineisto voidaan jakaa kuvan 13 kuvaamalla tavalla. Opetusaineiston oletetaan olevan niin suuri ja monipuolinen, että sillä saadaan malli koulutettua. Opetusdatasta voidaan vielä erottaa osuus mallin validointiin. Koulutuksen jälkeen tarkistetaan, että sen tuottamat tulokset ovat halutunlaisia, jonka jälkeen mallista tunnistettuja merkittäviä piirteitä käytetään varsinaisen

testidatan analysointiin. Tarkoituksena ei siis ole toistaa analyysiä jo käytetylle aineistolle, vaan malli tekee ennustuksia uusista tapauksista. (Alpaydin, 2021, s.56 - 58.)



Kuva 13. Aineiston jakoperusteet mukailien Alpaydin (Alpaydin, 2021, s.56 - 58.)

Yleisimpiä ohjatun oppimisen luokittelumenetelmiä ovat logistinen regressio, k:n lähimmän naapurin menetelmä, tukivektorikone, Bayes-luokittelija, diskriminanttianalyysi, erilaiset päätöspuut ja satunnaismetsät sekä neuroverkot (Tuominen et al., 2019). Lisäksi on olemassa luokittelumenetelmiä, jotka yhdistelevät useita luokittelijoita. Niin sanotut ensemble learning-menetelmät käyttävät useita algoritmeja saavuttaakseen paremman tarkkuuden. Näitä menetelmiä ovat esimerkiksi bagging ja boosting. (Khodabandehlou & Zivari Rahman, 2017.)

Van Haver ja De Caigny ovat omilla tutkimuksissaan tehneet laajan vertailun eri luokittelumenetelmien käytöstä asiakaspoistuman tutkimuksessa. Heidän mukaansa eniten käytetyt menetelmät ovat logistinen regressio, erilaiset päätöspuumallit (erityisesti satunnaismetsät) sekä tukivektorikoneet. (De Caigny et al., 2018; Haver, 2016.) Paukkeri on tutkinut erilaisten luokittelijoiden toimivuutta kotimaisen teleoperaattorin asiakaspoistuman hallinnassa ja toteaa parhaiten toimiviksi malleiksi Gradient Boosting- sekä satunnaismetsä-menetelmät (Paukkeri, 2013). Myös Martinez on todennut Gradient Boostingin parhaaksi malliksi nimenomaan ei-sopimusperustaisen kaupan asiakaspoistuman tutkimisessa (Martínez et al., 2020).

2.2.3 Mallien muuttujat ja mallin toimivuuden arviointi

Jotta koneoppimisen malli tunnistaisi poistumisriskissä olevat asiakkaat, sen täytyy tietää, miltä poistuvan asiakkaan käytös näyttää. Asiakkaan käytöstä analysoidaan erilaisen muuttujien avulla (Gatterman-Itschert & Thonemann, 2021). Aikaisemmissa B2C-puolen tutkimuksissa on käytetty niin sanottuja RFM-muuttujia (vertaa Kuva 9). Lyhenne tulee sanoista recency, frequency ja monetary, ja niillä kuvataan aikaa asiakkaan viimeisestä ostoksesta, ostosten tiheyttä sekä rahallista arvoa. (Buckinx & Van Den Poel, 2005; Miguéis et al., 2013.) Samoja muuttujia on käytetty onnistuneesti myös B2B-puolen tutkimuksissa (Gatterman-Itschert & Thonemann, 2021; Jahromi et al., 2014; Ringbeck et al., 2019).

B2B-puolella voi lisäksi olla useita alakohtaisia tärkeitä muuttujia. Esimerkiksi aloilla, joilla läheiset suhteet asiakkaisiin ovat tärkeitä, on tutkittu aikaa myyntiedustajan viimeisestä yhteydenotosta CRM-datan perusteella (Coussement & Van Den Poel, 2008; Gatterman-Itschert & Thonemann, 2021). Chen et al. on puolestaan tutkinut taiwanilaisen logistiikkayrityksen toimituskyvyn vaikutusta asiakaspoistumaan vertailemalla toimitusvirheiden osuutta kaikista toimituksista. Tämä ominaisuus nousi merkitykselliseksi heidän tutkimuksessaan (K. Chen et al., 2015). Jälleenmyyntiin suuntautuneessa tutkimuksessa Gatterman-Itschert ja Thonemann tutkivat lisäksi hinnoittelun sekä katteen merkitystä asiakaspoistumaan ja totesivat ne merkityksellisiksi ominaisuuksiksi (Gatterman-Itschert & Thonemann, 2021).

Kaikki ominaisuudet eivät tietystikään ole yhtä tärkeitä mallin toimivuuden kannalta. On tärkeää tunnistaa merkittävät ominaisuudet valtavan datamassan joukosta (Blum & Langley, 1997). Merkityksellisimmät ominaisuudet voidaan löytää esimerkiksi SHAP-arvoja tutkimalla. SHAP perustuu peliteoriaan ja sen avulla voidaan tarkastella muun muassa koneoppimismallin suorituskykyä. SHAP-arvojen avulla voidaan nähdä, onko eri ominaisuuksilla positiivinen vai negatiivinen vaikutus lopputulokseen. (Mangalathu et al., 2020.)

Luokittelijan suorituskykyä voidaan mitata useilla tavoin. Eräs yleinen tapa on käyttää sekaannusmatriisia (Taulukko 3). Matriisin avulla voidaan tunnistaa kahdenlaisia virheitä, valenegatiiviset ja valepositiiviset. (Gollapudi, 2016, s. 25.) Asiakaspoistumakontekstiin sovellettuna valenegatiivinen tarkoittaa, että asiakas olisi oikeasti lähdössä ja malli ei kykene ennustamaan sitä. Valepositiivinen puolestaan tarkoittaa, että ennusteen mielestä asiakas olisi lähdössä, mutta todellisuudessa ei. (Jahromi et al., 2014.)

Taulukko 3. Sekaannusmatriisi mukailten Vafeiadis (Vafeiadis et al., 2015)

Todellisuus	<i>oikeasti negatiivinen (F)</i>	oikeat negatiiviset (TN)	vale-positiiviset (FP)
	<i>oikeasti positiivinen (T)</i>	vale-negatiiviset (FN)	oikeat positiiviset (TP)
		<i>ennustettu negatiivinen (F)</i>	<i>ennustettu positiivinen (T)</i>
		Ennuste	

Sekaannusmatriisista voidaan johtaa useita luokittelijan toimivuutta kuvaavia tunnuslukuja (Vafeiadis et al., 2015).

Täsmällisyys (accuracy) ilmoittaa mallin yleisen täsmällisyyden ja kertoo siis osuuden näytteistä, jotka luokittelija on luokitellut oikein. Täsmällisyys antaa parhaan kuvan silloin, kun tulosten jakauma on tasainen. Se lasketaan kaavalla (katso lyhenteiden selitykset taulukosta 3):

$$\text{täsmällisyys} = \frac{TP+TN}{TP+FP+TN+FN}. \quad (1)$$

Herkkyys (recall) tarkoittaa oikein tunnistettujen positiivisten tapausten määrää. Se lasketaan kaavalla:

$$\text{herkkyys} = \frac{TP}{TP+FN}. \quad (2)$$

Tarkkuus (precision) ilmoittaa osuuden, jossa ennustettu positiivinen tulos on oikein. Se lasketaan kaavalla:

$$\text{tarkkuus} = \frac{TP}{TP+FP}. \quad (3)$$

Spesifisyys (specificity) tarkoittaa, kuinka suuri osuus ennustetuista negatiivisista tapauksista on oikeasti negatiivisia. Se lasketaan kaavalla:

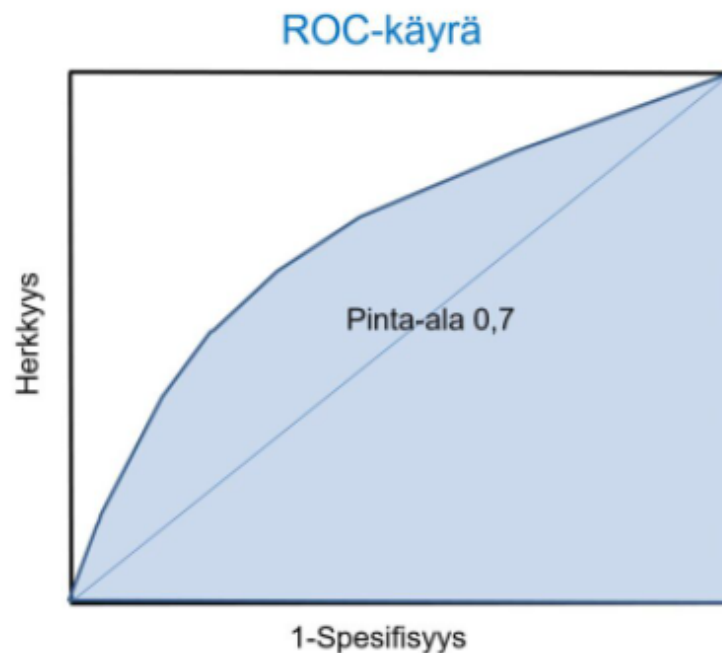
$$\text{spesifisyys} = \frac{TN}{TN+FP}. \quad (4)$$

Pelkästään täsmällisyys, herkkyys tai tarkkuus yksin ei riitä kuvaamaan luokittelijan tehokkuutta, koska hyvä taso yhdellä mittarilla ei välttämättä tarkoita samaa toisella. Tämän vuoksi usein käytetään myös F1-lukua. F1-luku on tarkkuuden ja herkkyyden

harmoninen keskiarvo. Ykköstä lähellä oleva arvo kertoo tarkkuuden ja herkkyuden hyvästä kombinaatiosta. F1-luku toimii pelkkää täsmällisyyttä paremmin silloin, kun tulosten jakauma ei ole tasainen. (Gollapudi, 2016; Vafeiadis et al., 2015.) F1-luku on käyttökelpoinen esimerkiksi asiakaspoistumaa tutkittaessa, koska oletettavasti suurempi osa yrityksen asiakkaista on jäämässä kuin lähtemässä. F1-luku lasketaan kaavalla:

$$F1 \text{ luku} = \frac{2 * \text{tarkkuus} * \text{herkkyys}}{\text{tarkkuus} + \text{herkkyys}}. \quad (5)$$

Luokittelijan tehokkuutta voidaan mitata myös ROC-analyysillä (Receiving Operator Characteristic), mikäli ennustettavan muuttujan luokat ovat epätasapainossa eli toiseen luokkaan kuuluu vaikkapa 90 prosenttia havainnoista. ROC-käyrän (Kuva 14) piirtämiseksi määritellään herkkyys ja johdonmukaisuus eri kynnyсарvoilla. Kynnyсарvojen määrittelyn jälkeen voidaan piirtää kuvaaja, jonka vaaka-akselilla on 1 – spesifisyys (kaava 4) ja pystyakselilla herkkyys (kaava 2). ROC-käyrän alapuolelle jäävä pinta-ala kuvastaa luokittelijan kykyä erotella havainnot (Kuva 14), ja siitä käytetään nimitystä AUC (Area Under the Curve). 45 asteen kulmassa oleva viiva kuvastaa satunnaista arvausta. Tämän suoran alle jäävän pinta-alan osuus on siis 0,5. AUC-arvot vaihtelevat käytännössä 0,5–1 välillä. Mitä korkeampi AUC-arvo on, sitä paremmin luokittelija pystyy jaottamaan havainnot eri luokkiin. (Kotu, 2019, kappale 8.2.)



Kuva 14. Esimerkki ROC-käyrästä (Knuuti, 2021)

2.2.4 Asiakaspoistuman ehkäiseminen

Koneoppimiseen perustuva malli tuottaa siis yrityksen datan perusteella ennusteen, onko asiakas riskissä poistua vai ei. Poistumariskiä ilmaistaan usein prosenttiluvun avulla: mitä korkeampi prosentti on sitä suurempi on riski, että asiakas poistuu (Verbeke et al., 2012). Erityisesti yrityspuolella ollaan kiinnostuneita jokaisen asiakkaan tilanteesta. Ennusteiden tulee olla mahdollisimman tarkkoja mutta lisäksi oikea-aikaisia, jotta asioihin ehditään vaikuttaa, ennen kuin on liian myöhäistä. (Hills et al., 2020.)

Poistumariskissä oleville asiakkaille halutaan usein kohdistaa toimenpiteitä, joiden avulla asiakassuhdetta pyritään jatkamaan. Yleisimpiä toimenpiteitä ovat erilaiset kampanjat, joilla tarjotaan poistumariskissä oleville asiakkaille erilaisia etuuksia tai alennuksia. Kohdistamalla kampanjat ainoastaan poistumariskissä oleville asiakkaille yritys ei turhaan jaa alennuksia asiakkaille, joiden asiakassuhde on kunnossa. Kampanjoilla on todettu olevan positiivinen vaikutus asiakaspoistuman pienentämiseen. (Jahromi et al., 2014; Ringbeck et al., 2019.) Kampanjat voidaan jakaa reaktiivisiin ja proaktiivisiin kampanjoihin. Reaktiivisella kampanjalla tarkoitetaan kampanjaa, jonka yritys kohdistaa jo lähteneisiin asiakkaisiin voittaakseen heidät takaisin. Proaktiivinen kampanja on puolestaan sellainen, jolla pyritään ennaltaehkäisemään poistumariskissä olevien asiakkaiden lähteminen. (Ascarza et al., 2018.) Myös kampanjan ajoituksella on tärkeä rooli sen onnistumisessa. Käyttämällä aikasarja-analyysia asiakaspoistuman määrittämisessä voidaan helpottaa kampanjoiden oikea-aikaisuutta. (Óskarsdóttir et al., 2018.)

B2B-puolella kampanjoiden lisäksi tärkeää on yhteydenpidon tiheys asiakkaaseen. Gatterman-Itchert ja Thonemann muodostivat omassa kenttätutkimuksessaan poistumariskissä olevista asiakkaista kaksi ryhmää. Toinen ryhmä soitettiin läpi, ja mikäli soittaja oli sitä mieltä, että asiakas on oikeasti vaarassa poistua, varattiin myyntiedustajan vierailu. Toiselle ryhmälle ei tehty mitään toimenpiteitä. Yhteydenpidolla asiakkaaseen saatiin pienenettyä ryhmän asiakaspoistumaa tarkastelujaksolla. Tutkijat laskivat, että toimenpiteillä saatiin säilytettyä noin kuusi miljoona myyntieuroa. (Gatterman-Itschert & Thonemann, 2021.)

2.3 Yhteenveto

Kirjallisuuskatsauksen perusteella voidaan summata, että kasvaneiden datamäärien ja koneoppimisen mallien kehityksen myötä asiakaspoistuman tutkiminen on helpottunut. Asiakaspoistuman hallinta on osa asiakassuhteen säilyttämisen prosessia ja integroituu

sitä kautta koko asiakassuhteen hallintaprosessiin. Asiakassuhteita hallitaan nykyisin CRM-järjestelmillä ja järjestelmistä saatava data on erittäin tärkeässä roolissa asiakaspoistumatutkimuksessa. CRM:n täysipainoinen käyttö täytyy pystyä juurruttamaan osaksi organisaatiokulttuuria ja ymmärtää sen olevan strateginen ja kokonaisvaltainen työkalu asiakassuhteen hallinnassa.

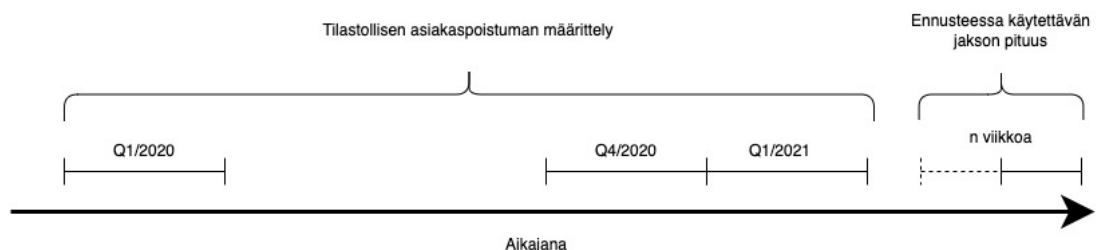
Rautakaupassa asiakaspoistuman tutkiminen koneoppimisen avulla on uutta, ja eritoten tässä tutkimuksessa ei-sopimusperustainen kauppa luo omat haasteensa tutkimukselle. Siksi onkin hyvä vertailla, mitä aikaisemmissa tutkimuksissa on saatu aikaan muualla. RFM-muuttujien käyttö mallia luodessa on selvää, mutta kirjallisuudesta on saatavissa hyviä esimerkkejä, mitä muita muuttujia voidaan hyödyntää. Mitä enemmän muuttujia, sitä enemmän tarvitaan myös erilaista dataa. Dataa joudutaan usein keräämään useista eri järjestelmistä. Tällöin on syytä varmistaa, että kaikkien järjestelmien tuottama data on riittävän laadukasta. Huonolaatuinen data voi vaarantaa koko mallin luotettavuuden.

Data-analyysin parempi hyödyntäminen tuo myös uusia työkaluja myynnin johdon käyttöön. Jotta asiakaspoistumaprosentti ei jää pelkäksi uudeksi tunnusluvuksi raportteihin, myynnin johdon tulee alkaa käyttämään sitä aktiivisesti asiakassuhteiden hallinnan johtamisessa. Eritoten on tärkeää määritellä tarkkaan, mitä sitten tehdään, jos koneoppimisen malli tunnistaa asiakkaan olevan poistumariskissä.

3. MENETELMÄ, AINEISTON KUVAUS JA ANALYYSI

3.1 Tutkimusmenetelmä

Tutkimuksen tarkoituksena on todentaa, saadaanko koneoppimiseen perustuvalla mallilla riittävän luotettavia ja käyttökelpoisia tuloksia asiakaspoistuman ehkäisemiseen. Mikäli mallista saadaan toimiva, sen käyttöä voitaisiin laajentaa rautakaupan muihin asiakasryhmiin. Mallin toimivuutta tutkitaan kvantitatiivisilla menetelmillä eli pyritään vertaamaan mallin tuottamaa ennustetta yrityksen omiin tilastoihin. Aluksi koneoppimisen mallille on tehtävä selväksi, millaiselta jo poistunut asiakas näyttää. Tähän käytetään tilastollista asiakaspoistuman määrittelyä. Asiakkaiden käytöstä tutkitaan kolmen kuukauden ajalta (Q1/2021), sitä edeltävän kolmen kuukauden ajalta (Q4/2020) sekä kolmen kuukauden ajalta edellisenä vuonna (Q1/2020), yhteensä siis kolme kolmen kuukauden periodia (Kuva 15). Käyttämällä eri aikasarjoja pyritään minimoimaan kausivaihtelun vaikutuksia. Näitä aikasarjoja käytetään koneoppimismallin opettamiseen. Mallille annetaan kynnysarvo, jonka ylittävät asiakkaat ovat mallin mukaan poistuneita tai riskissä poistua. Kynnysarvon laskemiseen käytetään ennalta määriteltäviä muuttujia, kuten aikaa viimeisestä ostoksesta tai ostoksien tiheyttä. Muuttujille annetaan tietyt painoarvot. On kuitenkin syytä huomata, että mallin opettamiseen käytettyjä muuttujia ei voida käyttää varsinaisen ennusteen tekemisessä. Tämä johtaisi mallin ylikouluttamiseen ja näin ollen epäluotettaviin tuloksiin.



Kuva 15. Aikajana asiakaspoistuman määrittelystä

Aluksi vertaillaan tilastollisen asiakaspoistuman tuottamia tuloksia yrityksen omiin analyyseihin ja yritetään todentaa, että malli onnistuu poimimaan jo poistuneet asiakkaat sekä poistumariskissä olevat asiakkaat mahdollisimman oikein. Näin pyritään varmistamaan, että mallin tuottama ennuste olisi mahdollisimman totuudenmukainen. Mikäli tilastollisessa määrittelyssä todetaan epä johdonmukaisuuksia, säädetään parametreja,

ajetaan opetusaineisto uudestaan ja tarkastellaan, korjautuivatko epäjohdonmukaisuudet. Tämä prosessi toistetaan niin monta kertaa, kunnes lopputulokseen ollaan tyytyväisiä. Tässä vaiheessa on oltava riittävän huolellinen, koska varsinaisen mallin toimivuus riippuu siitä, kuinka hyvin malli erottaa poistumariskissä olevat asiakkaat. Varsinaisen mallin tuottaman ennusteen eli testidatan aikasarjan pituus on käyttäjän määriteltävissä. Aikasarja ei voi olla sama, joita mallin opettamiseen on käytetty.

3.2 Aineiston kuvaus ja analyysi

Aineisto on ajettu kohdeyrityksen tietojärjestelmästä ja tallennettu sopivan kokoisissa osissa analyysia varten. Data koostuu eri tauluista, joissa on tietoa asiakkaan käyttäytymisestä. Tauluja ovat esimerkiksi asiakas- ja laskutustaulut. Tauluissa olevien sarakkeiden tietoja voidaan joko käyttää sellaisenaan tai laskea niiden avulla uusia tunnuslukuja. Esimerkiksi yhdistämällä asiakkaan id-numero ja tarkastelemalla laskujen päiväysten väliä saadaan selville, kuinka tiheästi asiakas on asioinut yrityksessä. Aineistoissa on yhteensä 8 taulua ja niissä yhteensä 370 saraketta. Datan tarkasteluvaiheessa jokaiselle taulun sarakkeelle määriteltiin tyyppi, onko kyseessä aika, päivämäärä, kokonaisluku, reaalityluku vai merkkijono. Kaikki sarakkeet eivät ole oleellisia asiakaspoistuman tutkimisen kannalta, mutta datan tarkasteluvaiheessa ei vielä haluttu jättää mitään tietoja pois aineistosta.

Aluksi tavoitteena on siis todentaa tilastollisen asiakaspoistuman paikkansapitävyys. Jotta varsinaisen ennusteen tekemiseen jäisi mahdollisimman paljon muuttujia käyttöön, opetusaineiston arviointiin käytettiin aluksi pelkästään viimeistä ostospäivää (recency). Tämä osoittautui kuitenkin heti aivan liian epävarmaksi, joten päätettiin ottaa käyttöön lisää muuttujia ja pelkkä recency muuttujana pudotettiin pois. Malli koulutettiin seuraavien taulukoissa 4 esitettyjen muuttujien avulla. RF-scoressa käytettiin painotusta 30 prosenttia recency ja 70 prosenttia frequency.

Taulukko 4. Mallin kouluttamiseen käytetyt muuttujat

frequency segment trend	frequency segmentin muutos vrt. edelliseen 3 kk
frequency segment trend ind	frequency segmentin muutos vrt. edelliseen vuoteen
RF score trend	RF-segmentin muutos vrt. edelliseen 3 kk
RF score trend ind	RF-segmentin muutos vrt. viime vuoteen

Kuvassa 16 on esitetty esimerkkiarvoja RFM-muuttujista. Arvojen perusteella asiakkaat jaetaan luokkiin 1–5 kuten kuvassa 17 (F-score sarake). Käytetyssä mallissa ei ole asetettu itse luokkarajoja vaan ne ovat dynaamiset, jotta malli olisi sellaisenaan käytettävissä mahdollisimman laajalle asiakaskunnalle. Osa asiakkaista tekee hankintoja pari kertaa vuodessa ja osa voi käydä joka päivä ostoksilla, joten dynaamiset rajat toimivat tässä tapauksessa paremmin. Mikäli asiakkaan luokka muuttuu tarkastelussa verrattuna edellisen kolmen kuukauden jaksoon (segment trend) tai viime vuoden vastaavana ajankohtana olevaan jaksoon (segment trend ind), malli antaa luokkamutoksesta arvon -1 tai +1 riippuen muutoksen suunnasta. Luokka voi pudota tai nousta myös enemmän, mikäli muutos on isompi.

CUSTOMER ID	REGENCY (DAY)	FREQUENCY (NUMBER)	MONETARY (TOTAL)
1	4	6	540
2	6	11	940
3	46	1	35
4	23	3	65
5	15	4	179
6	32	2	56
7	7	3	140
8	50	1	950
9	34	15	2630
10	10	5	191
11	3	8	845
12	1	10	1510
13	27	3	54
14	18	2	40
15	5	1	25

Kuva 16. Esimerkki RFM-analyysistä (Makhija, 2021)

CUSTOMER ID	FREQUENCY	F SCORE
9	15	5
2	11	5
12	10	5
11	8	4
1	6	4
10	5	4
5	4	3
13	3	3
7	3	3
4	3	2
14	2	2
6	2	2
15	1	1
8	1	1
3	1	1

Kuva 17. Esimerkki F-scoresta (Makhija, 2021)

RF-score lasketaan puolestaan kuvassa 18 olevan esimerkin kaltaisesti (monetary arvo poislukien). Eli malli jakaa asiakkaat dynaamisiin luokkiin recencyn ja frequencyn perusteella, ja luokkien arvoista lasketaan painotettu keskiarvo. Luokkamutoksista malli antaa edellä olevan esimerkin tapaan -1 tai +1 tai enemmän, mikäli muutos on isompi.

CUSTOMER ID	RFM CELL	RFM SCORE
1	5,4,4	4.3
2	4,5,4	4.3
3	1,1,1	1.0
4	2,2,2	2.0
5	3,3,3	3.0
6	2,2,2	2.0
7	4,3,3	3.3
8	1,1,5	2.3
9	1,5,5	3.7
10	3,4,3	3.3
11	5,4,4	4.3
12	5,5,5	5.0
13	2,3,2	2.3
14	3,2,1	2.0
15	4,1,1	2.0

Kuva 18. Esimerkki RFM-score (Makhija, 2021)

Luokkamuutoksille on annettu painoarvot, joiden avulla lasketaan varsinainen poistuman todennäköisyys. Esimerkki painoarvoista on esitetty taulukossa 5. Painoarvot ovat esimerkiksi kuvitteelliset. Jokaiselle opetusaineiston kouluttamiseen käytettävälle muuttujan luokkamuutokselle on asetettu omat painoarvot, joiden perusteella poistumatodennäköisyys lasketaan. Esimerkiksi jos asiakkaalla on yhden luokan lasku kolmessa muutujassa, malli laskee yhteen kolme kertaa 20 prosenttia, joten poistuman riskiksi saadaan 60 prosenttia. Mikäli kynnyсарvo on asetettu 60 prosenttiin, malli luokittelee asiakkaan poistumariskissä olevaksi. Mikäli luokkamuutoksista kolme on negatiivista ja yksi positiivinen, poistumatodennäköisyys on 40 prosenttia, jolloin malli ei vielä luokittele asiakasta poistumariskissä olevaksi, mutta siirtää asiakkaan tarkkailulistalle.

Taulukko 5. Esimerkki luokkamuutosten painoarvoista

Luokkamuutos	vaikutus poistumatodennäköisyyteen
- 1	+ 20 %
- 2	+ 40 %
- 3 tai enemmän	+ 60 %
+ 1	- 20 %
+ 2	- 40 %
+ 3 tai enemmän	- 60 %

Sen todentamiseksi pitävätkö mallin tuottamat tulokset poistumariskissä olevista asiakkaista paikkaansa, niitä verrattiin yrityksen tietojärjestelmästä ajettuihin lukuihin. Vertailussa käytettiin taulukossa 6 esitettyjä lukuja. Vertailun helpottamiseksi laskettiin tarkastelujaksoa edeltävän kolmen kuukauden sekä edellisen vuoden vastaavan ajanjakson luvuista indeksit. Mikäli indeksi jäi alle 70 prosenttiin tarkastelujakson luvuista, annettiin solulle arvo 1. (Taulukko 7). Taulukossa punaisella korostetut solut ilmentävät alle 70 prosentin olevia indeksejä.

Taulukko 6. Yrityksen tietojärjestelmästä ajatut luvut

asiakkaan ostot tarkasteluajanjaksolta	€
asiakkaan ostot tarkastelujaksoa edeltävältä 3 kuukaudelta	€
asiakkaan ostot viime vuodelta tarkastelujaksoa vastaavalta jaksolta	€
asiakkaan tilausmäärä tarkastelujaksolta	kpl
asiakkaan tilausmäärä tarkastelujaksoa edeltävältä 3 kuukaudelta	kpl
asiakkaan tilausmäärä viime vuodelta tarkastelujaksoa vastaavalta jaksolta	kpl

Taulukko 7. Analyysiesimerkki

Asia- kasId	sales IND 3 prev months	if IND <70 =>1	Sales IND to last Y	if IND <70 =>1	Order IND to prev 3 M	if IND <70 =>1	Order IND to last Y	if IND <70 =>1
1	7	1	83	-	31	1	89	-
2	492	-	684	-	400	-	133	-
3	33	1	47	1	59	1	50	1
4	117	-	95	-	85	-	59	1
5	129	-	72	-	100	-	71	-
6	190	-	1 326	-	273	-	178	-
7	34	1	115	-	68	1	74	-

Ostoja vertailtiin tarkastelujaksoa edeltävään kolmeen kuukauteen, jotta saataisiin käsitys mihin suuntaan asiakkaiden ostot ovat kehittymässä. Vertailu edelliseen vuoteen osoittaa puolestaan mahdollisen kausivaihtelun, eli jos asiakkaiden ostot laskevat joka vuosi samaan aikaan, niin kyse ei silloin ole hiipumisesta. Tilausmäärien vertailu puolestaan osoittaa, onko asiakkaiden aktiivisuus pysynyt samalla tasolla. Mikäli indekseistä kolme tai enemmän on punaisella, katsottiin joko että asiakkaat ovat poistuneet tai, että asiakkailla on suuri riski poistua; mitä alhaisempi indeksi sitä suurempi riski.

Opetusaineistossa oli 2803 asiakasta, joista malli tunnisti 382 asiakkaan olevan poistumariskissä tai poistuneita (Taulukko 8). Varmasti poistuneiden asiakkaiden poistumisen todennäköisyys on 100 prosenttia ja riskissä olevien asiakkaiden todennäköisyys yli 60 prosenttia. Mallin tunnistamia 382 asiakasta vertailtiin itse laskettuihin indekseihin. Mallin todettiin ennustaneen varmasti poistuneet asiakkaat 96,5 prosentin tarkkuudella. Malli löysi näiden lisäksi yhteensä 71 asiakasta, joiden poistumistodennäköisyys oli 60–80 prosentin välillä, mutta jotka indeksipohjaisessa vertailussa eivät nousseet poistuneiden listalle. Tarkemman analyysin jälkeen havaittiin, että mallin muuttujien painoarvot olivat liian suuret edellisen vuoden samaan ajanjaksoon vertailevassa kohdassa. Tämän vuoksi esimerkiksi uudet asiakkaat, joilla oli vähän tai ei ollenkaan tapahtumia edellisen vuoden tarkastelujaksolla, saivat poistumismerkinnän turhaan. Mallin koulutusosioon päätettiin siis lisätä huomioon otettavaksi tekijäksi asiakkuuden ikä. Lisäksi toinen havainto koski asiakkaita, joiden ostojen arvo oli noussut huomattavasti mutta ostojen tiheys oli laskenut yhden luokan verran. Malli ei ota vielä koulutusvaiheessa huomioon asiakkaiden ostojen rahallisen määrän muutoksia, joten ratkaisuna tähän ongelmaan päätettiin muuttaa muuttujien painoarvoja siten, että painoarvo on suurempi tarkastelujaksoa edeltävän kolmen kuukauden luokkamuutoksilla, kuin edelliseen vuoteen perustuvilla luokkamuutoksilla.

Taulukko 8. Poistumariskissä olevat asiakkaat vertailu 1

Poistumatodennäköisyys	lukumäärä	samaa mieltä	eri mieltä	% samaa mieltä
100 %	113	109	4	96,5 %
90 %	0	0	0	0,0 %
80 %	162	124	38	76,5 %
70 %	3	0	3	0,0 %
60 %	104	74	30	71,2 %
YHTEENSÄ	382	307	75	80,4 %

Malli tunnisti 1301 asiakasta, joiden poistumariski oli nolla prosenttia (Taulukko 9). Tehtyihin analyysihin verrattuna löytyi 110 asiakasta, joista oltiin sitä mieltä, että poistumatodennäköisyyden pitäisi olla yli 60 prosenttia. Mallin tarkkuus tässä luokassa oli 92 prosenttia. Näiden 110 asiakkaan indeksit olivat selvästi alle 70 prosenttia, mutta siitä huolimatta niiden RF- tai frequency-segmenttien luokat eivät olleet tippuneet. Väärä luokittelu johtui siitä, että tarkastelujaksoa edeltävällä kolmella kuukaudella oli liian pienet painoarvot, joten kolmen kuukauden painoarvoja päätettiin nostaa. Lisäksi malli löysi 255 asiakasta, joilla se katsoi poistumisriskin olevan 10–50 prosenttia, mutta itse tehtyihin analyysihin verrattuna näiden asiakkaiden poistumistodennäköisyys pitäisi olla yli 60 prosenttia. Tästä ryhmästä havaittiin selkeästi, että jos muuttujien luokkamuutos on enemmän kuin 1 luokka, niin painoarvoja täytyy nostaa. Opetusaineistoon tehtiin edellä mainitut muutokset ja sen jälkeen malli ajettiin uudestaan.

Taulukko 9. Ei-poistumariskissä olevat asiakkaat vertailu 1

Poistumatodennäköisyys	lukumäärä	samaa mieltä	eri mieltä	% samaa mieltä
0 %	1301	1191	110	91,5 %
10 %	208	192	16	92,3 %
20 %	318	261	57	82,1 %
30 %	150	135	15	90,0 %
40 %	387	232	155	59,9 %
50 %	57	45	12	78,9 %
YHTEENSÄ	2421	2056	365	84,9 %

Painoarvomuutosten lisäksi malliin luotiin prosessi uusien asiakkaiden käsittelemiseksi. Uusille asiakkaille annettiin painoarvokerroin, joka laskee poistumistodennäköisyyttä. Asiakkaan katsotaan olevan uusi, mikäli asiakassuhde on kestänyt vähemmän aikaa kuin kaksi kertaa tarkastelujakson pituus. Eli mikäli tarkastelujakso on 13 viikkoa, asiakassuhteen pitää olla vähintään 26 viikkoa vanha, jotta malli ei laske ylimääräistä poistumatodennäköisyyden vähennystä.

Tarkastelujaksojen painoarvojen merkitystä nostamalla tilastollisesta poistuman määrittelystä saatiin herkempi, jolloin malli ei nojaa niin vahvasti vuoden takaiseen asiakkaan ostokäyttäytymiseen vaan ottaa enemmän huomioon viimeaikaisia tapahtumia. Tehdyillä muutoksilla oli selkeä vaikutus mallin tunnistamiin poistumariskissä olevien asiakkaiden määriin. Aikaisemmillä asetuksilla malli tunnisti 382 poistumariskissä olevaa asiakasta (Taulukko 8) ja uusilla asetuksilla malli tunnisti 726 poistumariskissä olevaa asiakasta (Taulukko 10).

Taulukko 10. Poistumariskissä olevat asiakkaat vertailumuutosten jälkeen

Poistumatodennäköisyys	lukumäärä	samaa mieltä	eri mieltä	% samaa mieltä
100 %	464	358	106	77,2 %
90 %	65	51	14	78,5 %
80 %	2	0	2	0,0 %
70 %	131	57	74	43,5 %
60 %	64	12	52	18,8 %
YHTEENSÄ	726	478	248	65,8 %

Muutosten jälkeen itse tehtyyn analyysiin verrattuna malli muuttui epätarkemmaksi, mutta tässä on syytä huomioida, että itse tehty analyysi vaatii kolmen indeksin olevan alle 70 prosenttia, ennen kuin se luokittelee asiakkaan poistumariskin alaiseksi. Malli tunnistaa muutosten jälkeen herkemmin, mikäli esimerkiksi kolmen kuukauden ostojen tiheys ja aika viimeisestä ostoksesta on laskenut merkittävästi ja nostaa tällöin asiakkaan poistumariskiä kynnyksarvon yli. Mallin muuttumista herkemmäksi ei siis pidetä huonona asiana, vaan käytännössä se tarkoittaa, että malli tunnistaa poistumariskissä olevat asiakkaat paremmin.

Taulukko 11. Ei-poistumariskissä olevat asiakkaat muutosten jälkeen

Poistumatodennäköisyys	lukumäärä	samaa mieltä	eri mieltä	% samaa mieltä
0 %	1373	1340	33	97,6 %
10 %	130	110	20	84,6 %
20 %	181	155	26	85,6 %
30 %	146	107	39	73,3 %
40 %	110	78	32	70,9 %
50 %	137	87	50	63,5 %
YHTEENSÄ	2077	1877	200	90,4 %

Malliin tehtyjen muutosten jälkeen ei-poistumariskissä olevien asiakkaiden tunnistusprosentti parani selvästi (Taulukko 11). Nollan prosentin poistumariskin asiakkaissa tarkkuusprosentti nousi 97,6 prosenttiin, mitä voidaan pitää hyvänä (Holmes, 2022). Käytännössä tämä tarkoittaa vähemmän turhia puheluita asiakkaille. Tilastollinen asiakaspoistuman tunnistaminen näyttää antavan riittävän tarkat arvot todennäköisyyksillä 80–100 prosenttiin ja 0–20 prosenttiin. Eniten hajontaa itse tehtyihin analyyseihin nähden tulee todennäköisyyksillä 30–70 prosentin välillä. Tämä johtuu itse tehdyn analyysin karkeudesta. Analyysiin pitäisi lisätä lisää muuttujia, jotta sen tarkkuus paranisi. Käsini tämä ei mielekästä, vaan katsottiin, että tilastollinen asiakaspoistuman määrittely toimii riittävän hyvin, jotta sitä voidaan käyttää varsinaisen ennusteen apuna. Tilastollisen asiakaspoistuman määrittelyssä on muutenkin tärkeintä, että malli tunnistaa selkeästi poistuneet ja ei-poistumariskissä olevat asiakkaat, jotta se voi opettaa varsinaiselle ennustemallille, miltä näiden asiakkaiden parametrit näyttävät.

Aineistosta tehtiin Cohenin kappaa-analyysi 100 prosentin sekä 0 prosentin poistumariskissä olevista asiakkaista, jotta saataisiin myös tilastollinen varmuus tilastollisen asiakaspoistuman sekä itse tehdyn analyysin riittävästä samankaltaisuudesta. Analyysin tulokset on esitetty taulukossa 12. Kappa-arvo on 0.789, jota voidaan pitää riittävänä, sillä vahvan korrelaation raja on 0.8 (McHugh, 2012).

Taulukko 12. Cohenin kappa-analyysi

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
	Tilastollinen * oma tulkinta	1837	100.0%	0	0.0%	1837

Tilastollinen * oma tulkinta Crosstabulation

		oma tulkinta				Total	
		0		1			
		N	%	N	%	N	%
Tilastollinen	0	1340	92.7%	33	8.4%	1373	74.7%
	1	106	7.3%	358	91.6%	464	25.3%
Total		1446	100.0%	391	100.0%	1837	100.0%

Symmetric Measures

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.789	.017	34.009	.000
N of Valid Cases		1837			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Tilastollisen asiakaspoistuman määrittelyn jälkeen voidaan todeta sen tulokset toimiviksi, jotta opetusaineistoa voidaan käyttää itse ennustemallin kouluttamiseen. On syytä huomioida, että muuttujia, joita käytettiin opetusaineistossa, ei voida enää käyttää varsinaisessa ennusteessa. Ennusteessa käytetyt muuttujat on lueteltu taulukossa 13.

Taulukko 13. Ennusteessa käytetyt muuttujat

invoice_count	mbv_sum_trend_ind
total_lines	mbv_avg_trend_ind
avg_lines	Asiakastaso
mbv_sum	AsiakasRyhmä
mbv_avg	AsiakasPkunta
asiakas_period	Asiakastoimiala
asiakas_frequency	Henkiloluokka_nimi
invoice_count_segment	Liikevaihtoluokka_nimi
mbv_trend_over_period	hlo_luokka
mbv_segment	Aktiivisuus
monetary_segment	Riskiluokka
monetary_segment_trend	Toimialateksti
mbv_sum_trend	AsiakasActivationAge
mbv_avg_trend	CustomerAge
monetary_segment_trend_ind	

Ennusteen tekemiseen käytettiin kolmea eri algoritmia, joista malli valitsee parhaimman perustuen eri algoritmien ROC-AUC arvoihin. Käytetyt algoritmit olivat Gradient Boost, XGBoost sekä Multi Layer Perceptron (MLP) luokittelija. Näistä tarkimmaksi osoittautui XGBoost. Luokittelijoiden saamat pisteet on esitetty taulukossa 14. Käytettyjen luokittelijoiden saamat pisteet ovat hyvin lähellä toisiaan, joten periaatteessa mallissa olisi voitu käyttää mitä tahansa näistä kolmesta. Malli valitsee automaattisesti käyttöön tarkimman algoritmin.

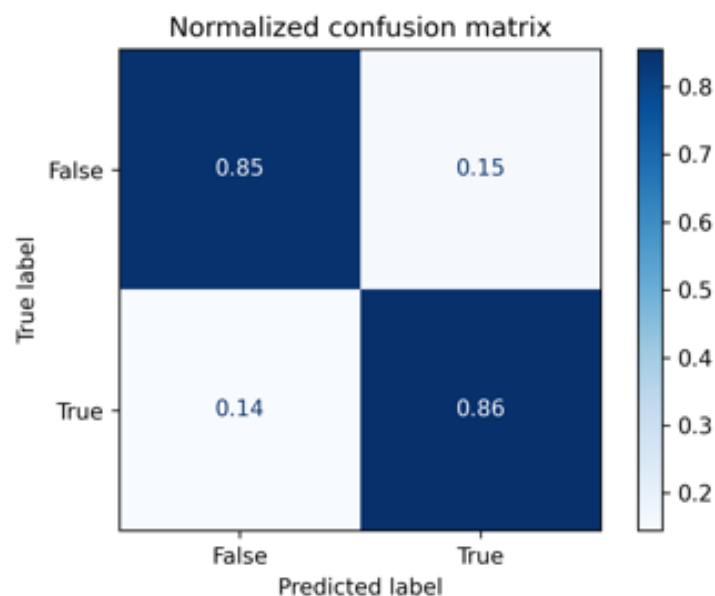
Taulukko 14. Eri algoritmien ROC AUC pisteet

ROC AUC	Gradboost	XGBoost	MLP
validation	0,90051	0,91632	0,91286
test	0,89128	0,89982	0,88936

4. TULOKSET

Tutkimuksessa onnistuttiin määrittelemään tilastollinen asiakaspoistuma riittävän luotettavalla tavalla, jotta sitä voitiin käyttää varsinaisen ennustemallin pohjana. On hyvä muistaa, että malli on vasta ensimmäinen versio ja sen toiminnan perusteella on tarkoitus myöhemmin määrittää, voidaanko sen käyttöä laajentaa yrityksen muihin asiakasryhmiin. Mallin kehittämiseen oli varattu tietty työtuntimäärä, eikä siinä saatu kaikkia haluttuja ominaisuuksia vielä malliin mukaan. Tämä on kuitenkin hyvä alku automatisoidulle asiakaspoistuman määrittämiselle kohdeyrityksessä. Seuraavaksi käydään läpi mallin toimivuutta analysoivia tunnuslukuja.

Kuvassa 19 on esitetty mallin sekaannusmatriisi. Matriisista voidaan tulkita, että mikäli malli käy läpi 200 asiakasta, joista 100 on oikeasti poistunutta ja 100 edelleen voimassa olevaa asiakkuutta, malli tunnistaa 86 asiakasta varmasti poistuneesta 100 asiakkaasta ja antaa väärän negatiivisen arvon 14 asiakkaalle. Toisin sanoen malli luokittelee siis 14 asiakasta väärin, vaikka asiakkaat oikeasti olisivat poistuneita. Näitä poistuneita asiakkaita malli ei siis tunnista. Vastaavasti malli tunnistaa 85 asiakasta edelleen asiakkaina olevista 100 asiakkaasta ja luokittelee 15 asiakasta poistuneiksi, vaikka asiakkaat oikeasti ovat edelleen asiakkaina. Käytännössä näille 15 asiakkaalle lähtisi siis tarkastussoitto turhaan.

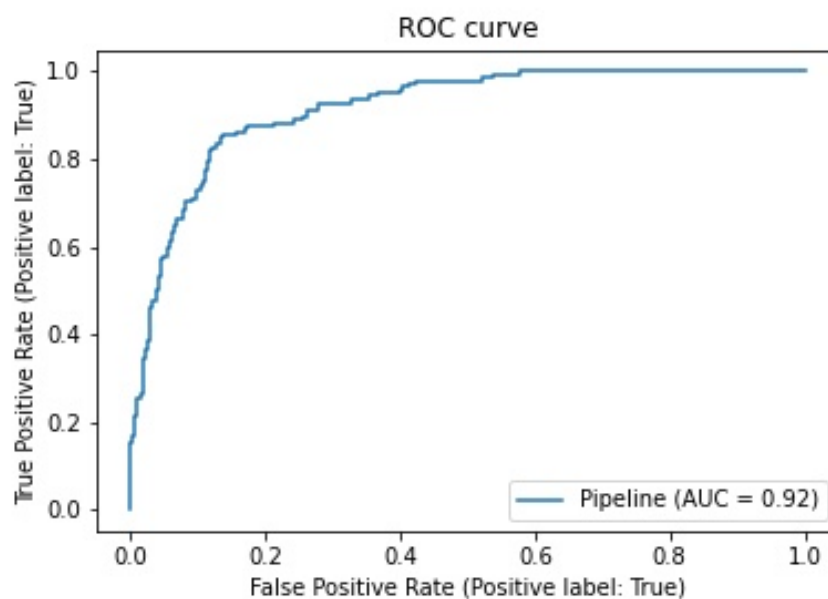


Kuva 19. XGboost-algoritmin sekaannusmatriisi

Taulukossa 15 sekä kuvassa 20 on esitetty luokittelijan toimivuuden analysointiin käytettävät luvut. Taulukon luvut on laskettu opetusaineistolle (training set) sekä validointi- ja testiaineistoille (validation & test set) erikseen. Mallin kouluttamisessa osia opetusdatasta ajetaan mallin läpi useamman kerran, jotta malli oppii, millaiset poistuneen asiakkaan parametrit ovat. Validointiaineiston avulla varmistetaan, että malli toimii halutulla tavalla. Testidata on aina uutta dataa, jota malli ei ole käsitellyt aikaisemmin. Siitä syystä testidatan arvot ovat alhaisemmat kuin opetusaineiston tai validointiaineiston. Mikäli testiaineiston arvot olisivat korkeammat kuin validointiaineiston, se voisi indikoida mahdollista mallin ylikouluttamista.

Taulukko 15. XGBoost-algoritmin metriikka

	precision score	recall score	f1 score	accuracy score	roc auc score
training set	0,93716	0,91796	0,92096	0,91796	0,98496
validation set	0,86964	0,85102	0,85586	0,85102	0,91603
testing set	0,83687	0,81564	0,82182	0,81564	0,89013

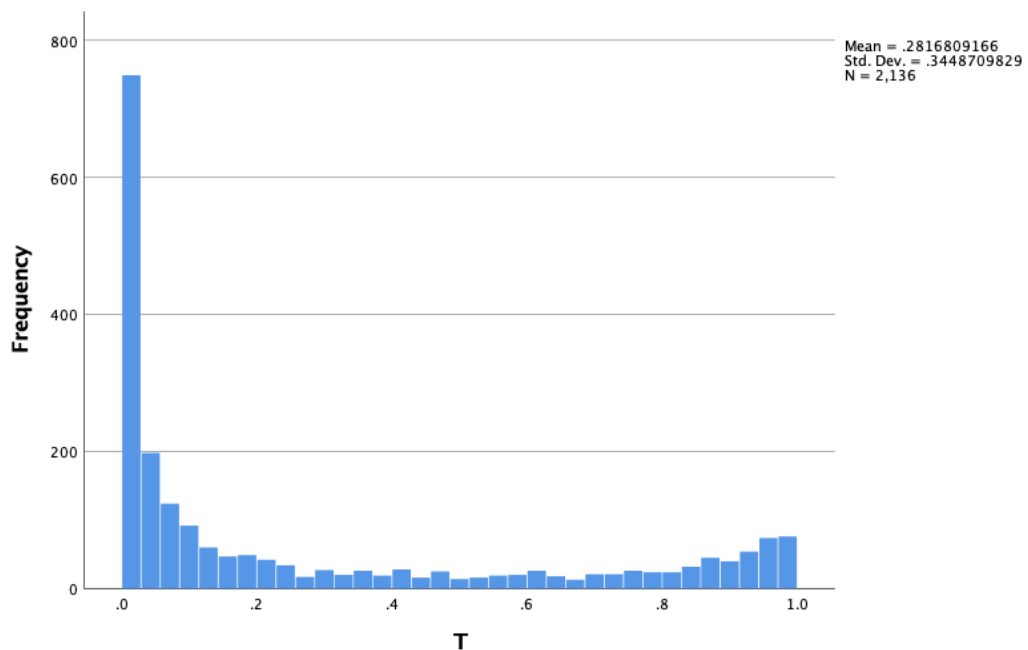


Kuva 20. ROC-kuvaaja

Koneoppimista käytetään sen verran laajasti, ettei ole olemassa kaikille dataseteille sopivia yleispäteviä rajoja mittaamaan mitkä arvot ovat riittävän hyviä. Tuloksia täytyy arvioida tapauskohtaisesti. (Fawcett, 2006; Ribeiro et al., 2016.) Yksinkertaistettuna yli 0,5 menevät arvot todistavat, ettei kahden välistä (0 tai 1) arvoa ole heitetty kolikolla. Jotta mallin tarkkuusarvoja voidaan arvioida, on ensin tarkasteltava tulosten jakaumaa. Kuvan

21 perusteella voidaan todeta, että tulokset eivät ole tasaisesti jakautuneet, vaan ne ovat selvästi painottuneet nollan tuntumaan. Tulosten tarkastelun kannalta tämä tarkoittaa, että täsmällisyyttä (accuracy) tärkeämpi mittari on f1-score, koska se toimii paremmin epätasaisesti jakautuneilla tuloksilla. Yli 0,8 olevaa arvoa voidaan pitää hyvänä asiakaspoistuman tutkimisessa (Holmes, 2022).

Herkkyden (recall) ja tarkkuuden (precision) arvot ovat keskenään riippuvaisia ja toisen merkittävät parannus vaikuttaa heikentävästi toiseen. Onko tärkeämpää, kuinka luotettava positiivinen havainto on (korkea precision-arvo) vai saadaanko varmasti kaikki positiiviset havainnot kiinni (korkea recall-arvo)? Tähän ei ole yksiselitteistä vastausta ja yli asiakaspoistuman tutkimisessa 0,8 meneviä arvoja pidetään riittävän hyvinä (Holmes, 2022). Kun mallista saadaan enemmän käytännön kokemusta, voidaan tehdä johtopäätös, kumpaa halutaan painottaa enemmän. Kuvassa 20 näkyvä ROC-kuvaaja kertoo siis mallin kyvystä erotella tulokset eri luokkiin. ROC-käyrän alle jäävää aluetta nimitetään AUC-arvoksi. XGboost-mallin AUC-arvo on 0,92. Yli 0,9 arvoja voidaan pitää hyvänä asiakaspoistuman tutkimisessa (Holmes, 2022).

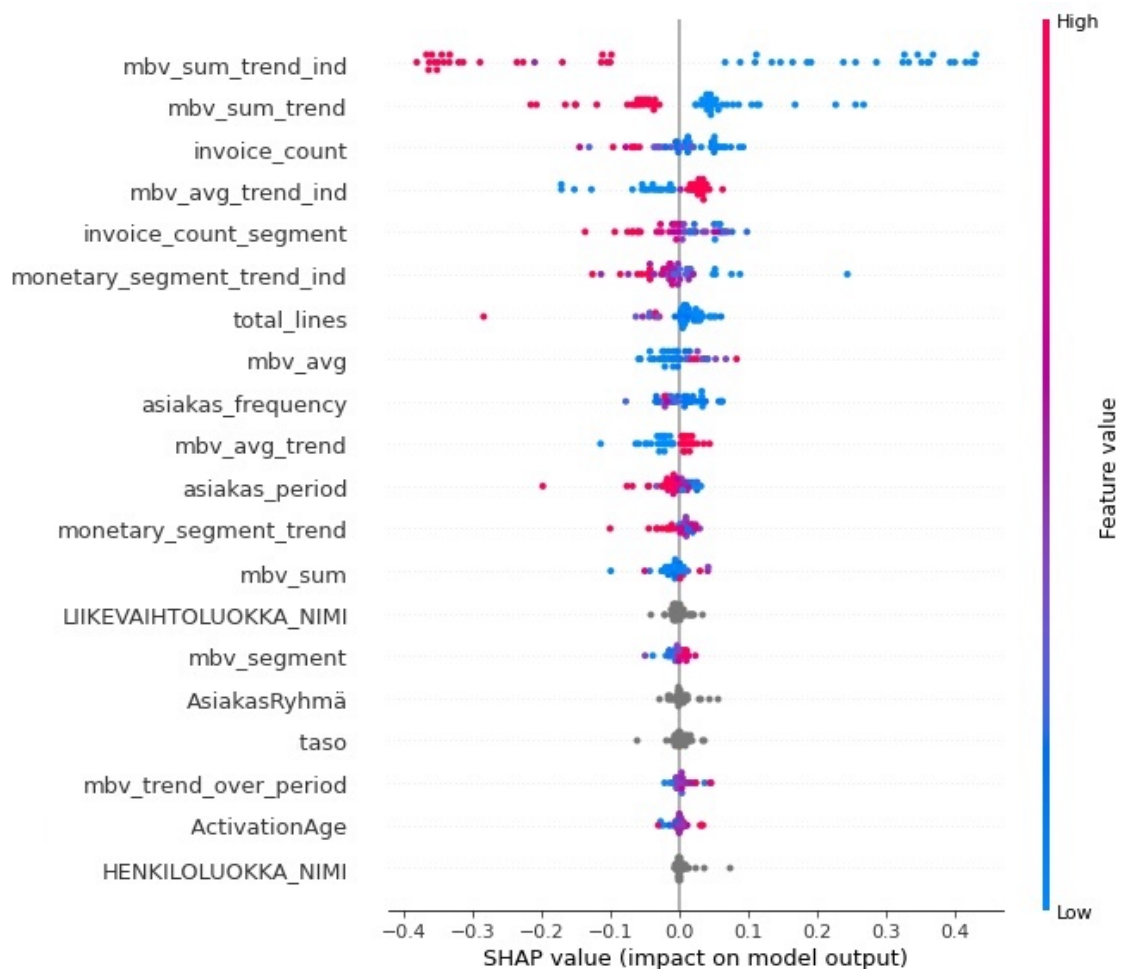


Kuva 21. Asiakaspoistumatodennäköisyyksien jakauma

Kuvasta 21 voidaan havaita, että tulosten jakauma on vahvasti keskittynyt nollan tuntumaan, mikä tarkoittaa, että suurimmalla osalla kohdeyrityksen asiakkaista poistumisriski on pieni. Tämä on kohdeyrityksen toiminnan kannalta tietysti positiivinen asia. Tässä

vaiheessa asiakkaista, joiden poistumatodennäköisyys on yli 0,6, lähtee ilmoitus myyjälle CRM-järjestelmään.

Tärkeimmät asiakaspoistumatodennäköisyyteen vaikuttavat muuttujat on esitetty kuvassa 22. Muuttujilla on joko positiivinen tai negatiivinen vaikutus poistumatodennäköisyyteen. Käytännössä tämä tarkoittaa, että jos SHAP-arvo on positiivinen, sillä on nostava vaikutus poistumatodennäköisyyteen ja päinvastoin. Tärkeimmät asiakaspoistumaan vaikuttavat muuttujat ovat ostoskorin kokonaisarvon kehitys edelliseen vuoteen tarkasteltuna (mbv_sum_trend_ind), ostoskorin kokonaisarvon kehitys tarkastelujaksoa edeltävään yhtä pitkään jaksoon (mbv_sum_trend) sekä laskujen lukumäärä tarkastelujaksolla (invoice count). Muita merkittäviä muuttujia ovat ostoskorin keskiarvon muutos edelliseen vuoteen (mbv_avg_trend_ind), laskujen määrän muutos tarkastelujaksolla (invoice_count_segment) sekä luokkamuuotos kokonaisostoissa edelliseen vuoteen verrattuna (monetary_segment_trend_ind).



Kuva 22. XGBoost algoritmin SHAP-arvot

Seuraavaksi mallin tuottamat tulokset viedään kohdeyrityksen CRM-järjestelmään ja poistumariskissä olevista asiakkaista alkaa tulla ilmoituksia myyjille. Alkuvaiheessa asiakaspoistumamalli otetaan käyttöön vain yhdelle yrityksen asiakasryhmistä. Malli on tarkoitus päivittää aluksi kerran kuussa ja mallin päivityksen yhteydessä tiedot päivitetään myös CRM-järjestelmään. Mallin toimivuutta on tarkoitus tarkastella syksyn 2022 ja kevään 2023 aikana, jonka jälkeen tehdään päätökset mallin mahdollisesta käytön laajentamisesta sekä jatkokehittämisestä.

Kun myyjä saa tiedon poistumariskissä olevasta asiakkaasta, on tarkoitus, että hän soittaa asiakkaalle ja yrittää saada paremman käsityksen tilanteesta. Myyjän tulkinta asiakkaan tilanteesta sekä sovitut jatkotoimenpiteet kirjataan CRM-järjestelmään. Mikäli myyjä havaitsee, että asiakas on jo kääntänyt ostot muualle, yritetään aina sopia asiakaskäynti asiakkaan voittamiseksi takaisin. Myynnin johto saa ajettua koontiraportin poistumariskin alla olevista asiakkaista ja pystyy jatkossa tarkastelemaan, pieneneekö poistumariski myyjän suorittamien toimenpiteiden jälkeen. Jatkossa kannattaa myös ottaa tarkasteluun takaisin voitettujen asiakkaiden asiakaskannattavuus, jotta voidaan tarkastella, onko takaisin voittaminen tehty kokonaiskannattavalla tavalla.

5. PÄÄTELMÄT JA KEHITYSEHDOTUKSET

5.1 Päätelmät tuloksista

Tulososiossa läpikäytiin asiakaspoistumamallin tunnuslukuja. Vaikka luvuilla ei menetelmän uutuuden vuoksi olekaan yleispätevää asteikkoa, voidaan todeta niiden olevan hyvällä tasolla mallin käyttöönottoa ajatellen. Vasta käytännön kokemus kuitenkin vahvistaa toimiiko malli kunnolla vai ei. Siksi onkin erityisen tärkeää, että mallin hyödyntäminen otetaan osaksi jokapäiväistä tekemistä ja havaitut tulokset raportoidaan CRM-järjestelmään. Kohdeyrityksessä on ollut CRM-järjestelmä käytössä useita vuosia, mutta sen käyttö ei ole ollut tehokasta. Viimeisen parin vuoden aikana tässä on kuitenkin otettu aimo harppaus eteenpäin ja CRM-järjestelmää ollaan juurruttamassa osaksi myyjien tärkeimpiä työkaluja. Aktiivinen CRM:n käyttö on tärkeä osa tehokasta asiakassuhteen hallintaa. Erityisesti myynnin johdon tulee kiinnittää tähän huomiota ja seurata, että kaikki myyjät käyttävät aktiivisesti CRM-järjestelmää.

Kohdeyrityksen myyjillä voi olla hoidossaan 20–200 asiakkuutta per myyjä. Isoimpien asiakkaiden asioista myyjät ovat yleisesti ottaen hyvin perillä, mutta jos asiakkaita on hoidettavana useita kymmeniä, kaikkien asioista ei millään pysty pysymään ajan tasalla ilman CRM-järjestelmän aktiivista käyttöä. Asiakaspoistumamalli havaitsee, mikäli asiakkaan ostokäyttäytyminen muuttuu aikaisemmasta, ja mikäli muutos on riittävän iso, järjestelmä varoittaa myyjää. Poistumamalli helpottaa siis myös myyjien työtä erityisesti niiden myyjien osalta, joilla on paljon asiakkuuksia hoidettavana.

Asiakaspoistuman määrittely antaa myynnin johdolle myös uudenlaisia työkaluja seurata myyjien suoritustasoa. Mikäli myyjälle alkaa kertyä useita hiipumisvaarassa olevia asiakkaita, se voi indikoida tarvetta muuttaa myyjän työtapoja. Asiakaspoistumaa on tähän asti seurattu manuaalisesti raporteista, joten uudistus helpottaa myös myynnin johdon arkea, koska tieto hiipumisesta menee suoraan myyjälle.

Kun mallin toiminnasta saadaan enemmän kokemusta, on mielenkiintoista nähdä, onko myyjän puhelinsoitolla ja/tai tapaamisella positiivinen vaikutus asiakaspoistumaan. B2B-puolella tavara ostetaan melkein pä poikkeuksetta tarpeeseen, joten kohdennetuilla kampanjoilla voi olla vaikea saada asiakkaan ostoja käännettyä takaisin, ellei kampanja osu juuri asiakkaan tarvitsemaan tuotteeseen.

5.2 Kehitysehdotuksia

5.2.1 Uudet muuttajat

Kohdeyrityksessä ei ole aikaisemmin ollut käytössä asiakaspoistumamallia, joten jo mallin suunnittelun alussa päätettiin luoda perusmalli, jotta nähdään, onko siitä yritykselle hyötyä. Kaikkia toivottuja ominaisuuksia ei saatu heti ensimmäiseen versioon mukaan. Mikäli malli todetaan toimivaksi ja sitä halutaan jatkokehittää, malliin olisi hyvä tuoda lisää muuttujia, joiden avulla sen toimintaa saisi vielä tarkemmaksi.

Ensimmäinen kehityskohta olisi vastuumyyjän mahdollisen vaihtumisen huomioon ottaminen mallissa. Aikaisemmissa asiakaspoistumatutkimuksissa on todettu, että myyjän vaihdolla voi olla jopa 13,2–17,6 prosentin negatiivinen vaikutus myyntiin (Shi et al., 2016). Rakennusalalla suhteet henkilöityvät ja asiakkaat seuraavat helposti myyjää, mikäli tämä vaihtaa työpaikkaa kilpailijalle. Myös myyjän eläköitymisen yhteydessä on hyvä huomioida, että asiakkaalle seuraava tuttu myyjä ei välttämättä olekaan oman yrityksen palveluksessa. Poistumamalliin olisi hyvä saada ominaisuus, joka tarkkailisi asiakkaan ostokäyttäytymistä erityisen tarkasti myyjän vaihdoksen yhteydessä.

Toinen kehityskohta olisi huomioida aika viimeisestä myyjän yhteydenotosta. Tällä muuttujalla on myös todettu olevan merkitystä aikaisemmissa asiakaspoistumatutkimuksissa erityisesti aloilla, joilla asiakassuhteet henkilöityvät vahvasti (Coussement & Van Den Poel, 2008; Gatterman-Itschert & Thonemann, 2021). Tämän muuttujan tärkeys korostuu erityisesti myyjillä, joilla on paljon pieniä asiakkaita. Tällöin säännöllinen yhteydenpito kaikkiin asiakkaisiin voi helposti jäädä kiireen keskellä. Mikäli järjestelmä muistuttaisi automaattisesti asiasta, tämä parantaisi varmasti asiakassuhteen laatua ja johtaisi parempaan asiakaskokemukseen ja -uskollisuuteen.

Mielenkiintoista olisi lisäksi saada malli ottamaan huomioon, mikäli asiakas siirtää jonkin tietyn tuoteryhmän ostot muualle. Osittaista asiakaspoistumaa on tutkittu aikaisemmin useissa tutkimuksissa (Buckinx & Van Den Poel, 2005; Clemente-Císcar et al., 2014; Gatterman-Itschert & Thonemann, 2021; Miguéis et al., 2013; Ringbeck et al., 2019). Osittaisen asiakaspoistuman on todettu usein indikoivan asiakkaan halukkuutta siirtää muutkin ostot muualle. Tämän muuttujan tuomisesta osaksi mallia olisi myös paljon apua myyjän päivittäisessä työssä. Myyjät eivät ehdi analysoimaan, mitä kaikkea he asiakkaalle myyvät, kun asiakkaita on paljon ja tuoterepertuaari iso. Järjestelmä voisi varoittaa myyjää, mikäli asiakas ei vaikkapa ole ostanut puutavaraa ollenkaan kuukauteen muun kaupan käydessä normaalisti. Tämä toimisi samalla myyjälle indikaationa, että kyseinen tuoteryhmä on hinnoiteltu liian ylös markkinahintaan nähden.

Myös asiakaskannattavuus olisi hyödyllistä saada integroitua osaksi mallia. Teoriaosassa käsiteltiin myös valaskäyrää (Kuva 11) eli miten erilaisten asiakkaiden kannattavuus vaikuttaa lopulta yrityksen tulokseen. Kannattamattomia asiakkaita ei kannata houkutella takaisin. Lisäksi voidaan nimetä muuttujia, kuten asiakastyytyväisyys (NPS), toimitusvarmuus (OTIF), tuotteiden saatavuus sekä esimerkiksi yksikön läheisyys asiakkaan työmaahan verrattuna. Näitä voisi tuoda osaksi mallia. Kaikki muuttujat eivät välttämättä ole oleellisia mallin toimivuuden kannalta, mutta SHAP-arvoja tarkastelemalla voidaan nähdä merkitykselliset muuttujat.

Kaiken kaikkiaan mallin kehittäminen vaatii, että kaikki yrityksen järjestelmien tuottama data on riittävän laadukasta ja kattavaa. Esimerkiksi asiakastyytyväisyyden (NPS) integroiminen osaksi mallia vaatii, että kaikki asiakkaat ovat vastanneet asiakastyytyväisyyskyselyyn. Vastaavasti muuttuja aika viimeisestä myyjän yhteydenotosta vaatii, että myyjä on ylipäättään soittanut asiakkaalle ja tehnyt siitä asianmukaisen merkinnän CRM-järjestelmään. Jotta näitä ominaisuuksia voitaisiin ottaa osaksi asiakaspoistumamallia, mainittujen tietojen on löydyttävä kaikista asiakkaista. Tämä asettaa sekä organisaatiolle että myynnin johdolle vaatimuksen luoda myyjille riittävät kannustimet ylläpitää tarpeeksi laadukasta dataa.

5.2.2 Laajemmat kehityskohteet

Kohdeyrityksen on tärkeä ymmärtää datan arvo ja merkitys sekä panostaa riittävästi data-analysoivan henkilöstön osaamiseen. Koneoppimisen mallin kehittäminen ja ylläpitäminen vaatii aikaa ja osaamista. Palvelu voidaan tietysti ostaa ulkopuolelta, mutta rautakauppa-alan ominaispiirteet vaativat, että kehitystyössä on mukana myös henkilöitä, joilla on liiketoimintaosaamista ja -ymmärrystä alalta. Mikäli mallia halutaan kehittää ja ylläpitää itse tarvitaan lisäksi henkilö, jolla on kokemusta Azuren Databricks-ympäristöstä sekä koodausosaamista. Malli on koodattu Python-ohjelmointikielellä ja uusien ominaisuuksien integroiminen malliin vaatii syvää osaamista.

Asiakaspoistumaa ehkäisevinä toimenpiteitä on yleisesti käytetty paljon erilaisia markkinointitoimenpiteitä, kuten kampanjoita ja suoramarkkinointia. Rautakaupan B2B-puolella tuotekampanjat ovat hieman kyseenalaisia, koska tavara ostetaan yleensä tarpeeseen, joten kampanjat vaatisivat paljon asiakaskohtaista räätälöintiä. Sen sijaan tarjouksena voisi olla esimerkiksi rahtivapaa seuraava toimitus asiakkaan takaisin houkuttelemiseksi. Teoriaosassa esiteltiin lyhyesti myös CRM:n seuraava kehitysversio; tekoälyyn pohjautuva CRM. Tämänkaltainen järjestelmä mahdollistaisi paremmin automatisoidun

tekoälyyn pohjatuvan asiakaskohtaisten kampanjoiden ja palveluiden kustomoinnin. Tulevaisuudessa järjestelmä voisi lähettää automaattisesti räätälöityjä tarjouksia asiakkaalle, mikäli se havaitsisi jonkin tuotteen tai tuoteryhmän myynnin hiipuvan.

Helpoin asiakaspoistumaa ennaltaehkäisevä toimi on kuitenkin säännöllinen yhteydenpito asiakkaaseen. Myyjien asiakasmääriin tulisi kiinnittää huomiota, jotta jokaisella myyjällä olisi mahdollisuus hoitaa kunnolla koko asiakaspiirinsä. Asiakaspoistumamallia jatkokehittämällä myyjän työtä voidaan helpottaa integroimalla malliin muuttujiksi vastuumyyjän vaihto, aika viimeisestä myyjän yhteydenotosta sekä tieto, mikäli asiakas siirtää tietyn tuoteryhmän ostot muualle. Näiden muuttujien avulla malli muuttuisi herkemmäksi ja osittaiseenkin asiakaspoistumaan päästäisiin kiinni mahdollisimman varhaisessa vaiheessa. Asiakas olisi tällöin helpompi voittaa takaisin verrattuna tilanteeseen, jossa ostot ovat ehtineet loppua kokonaan.

Asiakaspoistuman hallitseminen edesauttaa yrityksen asiakassuhteiden hallintaa sekä parantaa kiistatta yrityksen kannattavuutta. Se on lähtölaukaus kohti tehokkaampaa ja johdonmukaisempaa asiakassuhteiden hoitoa. Isommassa kuvassa kannattaakin huomiota kiinnittää asiakaspoistuman hallinnan lisäksi myös koko asiakassuhteen hallintaprosessiin. Tekoälyn avulla on mahdollista alkaa ennakoida asiakassuhteen elinkaarta ja elinkaaren arvoa. Tämä mahdollistaa tulevaisuudessa keskittymisen oikeiden asiakkaiden hoitamiseen.

LÄHTEET

- Ahvonen, K., Bremer, O., Djakonoff, V., Koponen, J., Mikkonen, J., & Toivonen, M. (2022). *Suomen vahvuudet, haasteet ja mahdollisuudet datatalouden rakentamisessa*. Sitra. <https://www.sitra.fi/julkaisut/suomen-vahvuudet-haasteet-ja-mahdollisuudet-datatalouden-rakentamisessa/#esipuhe>
- Alpaydin, E. (2021). *Koneoppiminen*. Terra Cognita.
- Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing & Informatics*, 15(2), 94–101. <https://doi.org/10.1016/j.aci.2018.05.004>
- Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schrift, R. (2018). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. *Customer Needs and Solutions*, 5(1–2), 65–81. <https://doi.org/10.1007/S40547-017-0080-0>
- Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2), 245–271. [https://doi.org/10.1016/S0004-3702\(97\)00063-5](https://doi.org/10.1016/S0004-3702(97)00063-5)
- Braun, M., & Schweidel, D. A. (2011). Modeling Customer Lifetimes with Multiple Causes of Churn. *Marketing Science*, 30(5), 757–944. <https://doi.org/10.1287/MKSC.1110.0665>
- Buckinx, W., & Van Den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1), 252–268. <https://doi.org/10.1016/J.EJOR.2003.12.010>
- Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2), 277–288. <https://doi.org/10.1016/j.eswa.2005.11.037>
- Buttle, F., & Maklan, S. (2019). Customer relationship management: Concepts and technologies: Fourth edition. *Customer Relationship Management: Concepts and Technologies: Fourth Edition*, 1–444. <https://doi.org/10.4324/9781351016551>

- Cao, G., Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2021). Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. *Technovation*, *106*, 102312. <https://doi.org/10.1016/j.technovation.2021.102312>
- Charlet Annie, L. M., & Kumar, A. D. (2012). *Market Basket Analysis for a Supermarket based on Frequent Itemset Mining*. International Journal of Computer Science Issues. <https://www.proquest.com/docview/1270319005/fulltext/21C54DB946494BC3PQ/1?accountid=14242>
- Chatterjee, S., Rana, N. P., Tamilmani, K., & Sharma, A. (2021). The effect of AI-based CRM on organization performance and competitive advantage: An empirical analysis in the B2B context. *Industrial Marketing Management*, *97*, 205–219. <https://doi.org/10.1016/J.INDMARMAN.2021.07.013>
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly: Management Information Systems*, *36*(4), 1165–1188. <https://doi.org/10.2307/41703503>
- Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM) People, process and technology. *Business Process Management Journal*, *9*(5), 672–688. <https://doi.org/10.1108/14637150310496758>
- Chen, K., Hu, Y.-H., & Hsieh, Y.-C. (2015). Predicting customer churn from valuable B2B customers in the logistics industry: a case study. *Information Systems and E-Business Management*, *13*, 475–494. <https://doi.org/10.1007/s10257-014-0264-1>
- Chen, P. Y., & Hitt, L. M. (2002). Measuring switching costs and the determinants of customer retention in internet-enabled businesses: A study of the online brokerage industry. *Information Systems Research*, *13*(3), 255–274. <https://doi.org/10.1287/ISRE.13.3.255.78>
- Chiang, R. H. L., Grover, V., Liang, T.-P., Dongsong, & Guest, Z., & Zhang, D. (2018). Special Issue: Strategic Value of Big Data and Business Analytics. *Journal of Management Information Systems*, *35*(2), 383–387. <https://doi.org/10.1080/07421222.2018.1451950>
- Clemente-Císcar, M., San Matías, S., & Giner-Bosch, V. (2014). Interfaces with Other Disciplines A methodology based on profitability criteria for defining the partial defection of customers in non-contractual settings. *European Journal of Operational Research*, *239*(1), 276–285. <https://doi.org/10.1016/j.ejor.2014.04.029>

- Coussement, K., & Van Den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), 313–327. <https://doi.org/10.1016/j.eswa.2006.09.038>
- Davenport, T. H., Barth, P., & Bean, R. (2012). *How “Big Data” is Different*. MIT Sloan Management Review. [https://www.hbs.edu/ris/Publication Files/SMR-How-Big-Data-Is-Different_782ad61f-8e5f-4b1e-b79f-83f33c903455.pdf](https://www.hbs.edu/ris/Publication%20Files/SMR-How-Big-Data-Is-Different_782ad61f-8e5f-4b1e-b79f-83f33c903455.pdf)
- De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760–772. <https://doi.org/10.1016/J.EJOR.2018.02.009>
- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings*, 1644, 97–104. <https://doi.org/10.1063/1.4907823>
- Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in Scientific Data Infrastructure. *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*, 48–55. <https://doi.org/10.1109/CTS.2013.6567203>
- Dick, A. S., & Basu, K. (1994). Customer loyalty: Toward an integrated conceptual framework. *Journal of the Academy of Marketing Science*, 22(2), 99–113. <https://doi.org/10.1177/00920703942222001>
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/J.IJINFOMGT.2019.01.021>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/J.PATREC.2005.10.010>
- Figalist, I., Elsner, C., Bosch, J., & Olsson, H. H. (2019). Customer churn prediction in B2B contexts. *Lecture Notes in Business Information Processing*, 370 LNBIP, 378–386. https://doi.org/10.1007/978-3-030-33742-1_30
- Gallo, A. (2014). *The Value of Keeping the Right Customers*. Harvard Business Review. <https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>
- Gatterman-Itschert, T., & Thonemann, U. W. (2021). (PDF) *Proactive customer retention management in a non-contractual B2B setting based on churn prediction*

with random forests.

https://www.researchgate.net/publication/353794359_Proactive_customer_retention_management_in_a_non-contractual_B2B_setting_based_on_churn_prediction_with_random_forests

Gollapudi, S. (2016). *Practical machine learning : tackle the real-world complexities of modern machine learning with innovative and cutting-edge techniques* (1st editio). Packt Publishing.

Gordini, N., & Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*, 62, 100–107. <https://doi.org/10.1016/j.indmarman.2016.08.003>

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, 26(3), 191–209. <https://doi.org/10.1016/J.JSIS.2017.07.003>

Gupta, S., Leszkiewicz, A., Kumar, V., Bijmolt, T., & Potapov, D. (2020). Digital Analytics: Modeling for Insights and New Methods. *Journal of Interactive Marketing*, 51, 26–43. <https://doi.org/10.1016/J.INTMAR.2020.04.003>

Gustafsson, A., Johnson, M. D., Roos, I., Maynard, D., & Ross, S. M. (2005). The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention. *Journal of Marketing*, 69(4), 210–218. <https://doi.org/10.1509/jmkg.2005.69.4.210>

Hallikainen, H., Savimäki, E., & Laukkanen, T. (2020). Fostering B2B sales with customer big data analytics. *Industrial Marketing Management*, 86, 90–98. <https://doi.org/10.1016/J.INDMARMAN.2019.12.005>

Harbert, T. (2021). *Tapping the power of unstructured data | MIT Sloan*. MIT Sloan Management Review. <https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data>

Haver, J. Van. (2016). *Benchmarking analytical techniques for churn modelling in a B2B context*. Gent University. https://libstore.ugent.be/fulltxt/RUG01/002/351/241/RUG01-002351241_2017_0001_AC.pdf

Hills, W., Daniel, W., Lu, M. Y., Schaer, O., & Adams, S. (2020). Modeling Client Churn for Small Business-to-Business Firms. *2020 Systems and Information Engineering*

Design Symposium (SIEDS), 1–7.

<https://doi.org/10.1109/SIEDS49339.2020.9106673>

Holmes, N. (2022). *Sähköpostihaastattelu 16.6.2022 Nicola Holmes, Senior Data Scientist, Kaito Insight Oy.*

Idris, A., & Khan, A. (2016). Churn Prediction System for Telecom using Filter–Wrapper and Ensemble Classification. *The Computer Journal*, 60(3), 410–430.

<https://doi.org/10.1093/comjnl/bxv123>

Ishwarappa, & Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and hadoop technology. *Procedia Computer Science*, 48(C), 319–324.

<https://doi.org/10.1016/J.PROCS.2015.04.188>

Jabbar, A., Akhtar, P., & Dani, S. (2020). Real-time big data processing for instantaneous marketing decisions: A problematization approach. *Industrial Marketing Management*, 90, 558–569.

<https://doi.org/10.1016/J.INDMARMAN.2019.09.001>

Jahromi, A. T., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 43(7), 1258–1268.

<https://doi.org/10.1016/J.INDMARMAN.2014.06.016>

Johnson, M. D., Herrmann, A., & Huber, F. (2006). The Evolution of Loyalty Intentions. *Journal of Marketing*, 70(2), 122–132. <https://doi.org/10.1509/jmkg.70.2.122>

Jokiniemi, S. (2020). *B2B-myyntin organisoituminen ja uudistuvat roolit digijajalla. Myyntin Ja Markkinoinnin Ammattilaiset.*

<https://mma.fi/ajankohtaista/artikkelit/b2b-myyntin-organisoituminen-ja-uudistuvat-roolit-digijajalla/>

Jones, M. (2019). What we talk about when we talk about (big) data. *The Journal of Strategic Information Systems*, 28(1), 3–16.

<https://doi.org/10.1016/J.JSIS.2018.10.005>

Jones, M. A., Mothersbaugh, D. L., & Beatty, S. E. (2002). Why customers stay: measuring the underlying dimensions of services switching costs and managing their differential strategic outcomes. *Journal of Business Research*, 55(6), 441–450. [https://doi.org/10.1016/S0148-2963\(00\)00168-5](https://doi.org/10.1016/S0148-2963(00)00168-5)

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.

<https://doi.org/10.1126/SCIENCE.AAA8415>

- Kandeil, D. A., Saad, A. A., & Youssef, S. M. (2014). A Two-Phase Clustering Analysis for B2B Customer Segmentation [Proceeding]. *2014 International Conference on Intelligent Networking and Collaborative Systems*, 221–228.
<https://doi.org/10.1109/INCoS.2014.49>
- Khodabandehlou, S., & Zivari Rahman, M. (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*, 19(1–2), 65–93.
<https://doi.org/10.1108/JSIT-10-2016-0061>
- Kim, M., Eun Park, J., Dubinsky, A. J., Dubinsky, A. J., Dubinsky, A. J., & Chaiy, S. (2012). Frequency of CRM implementation activities: A customer-centric view. *Journal of Services Marketing*, 26(2), 83–93.
<https://doi.org/10.1108/08876041211215248>
- King, S. F., & Burgess, T. F. (2008). Understanding success and failure in customer relationship management. *Industrial Marketing Management*, 37(4), 421–431.
<https://doi.org/10.1016/J.INDMARMAN.2007.02.005>
- Knuuti, J. (2021, February 19). *Monigeeniset riskimittarit eivät ratkaise sepelvaltimotaudin seulontaa – TS Blogit*. Turun Sanomat.
<https://blogit.ts.fi/terveys-tiede/monigeeniset-riskimittarit-eivat-ratkaise-sepelvaltimotaudin-seulontaa/>
- Kortelainen, M. (2022, March 7). *”On tämä melkoista hintarallia” – Katso, miten rakennusmateriaalien hinnat ja työntekijöiden palkat ovat muuttuneet | Rakennuslehti*. Rakennuslehti. <https://www.rakennuslehti.fi/2022/03/on-tama-melkoista-hintarallia-katso-miten-rakennusmateriaalien-hinnat-ja-tyontekijoiden-palkat-ovat-muuttuneet/>
- Kotu, V. (2019). *Data science : concepts and practice* (Second edi). Morgan Kaufmann Publishers, an imprint of Elsevier.
- Larsson, A., & Broström, E. (2020). Ensuring customer retention: insurers’ perception of customer loyalty. *Marketing Intelligence and Planning*, 38(2), 151–166.
<https://doi.org/10.1108/MIP-02-2019-0106>
- Libai, B., Bart, Y., Gensler, S., Hofacker, C. F., Kaplan, A., Kötterheinrich, K., & Kroll, E. B. (2020). Brave New World? On AI and the Management of Customer Relationships. *Journal of Interactive Marketing*, 51, 44–56.
<https://doi.org/10.1016/J.INTMAR.2020.04.002>
- Lilien, G. L. (2016). The B2B Knowledge Gap. *International Journal of Research in*

- Marketing*, 33(3), 543–556. <https://doi.org/10.1016/j.ijresmar.2016.01.003>
- Lipiäinen, H. S. M. (2015). CRM in the digital age: Implementation of CRM in three contemporary B2B firms. *Journal of Systems and Information Technology*, 17(1), 2–19. <https://doi.org/10.1108/JSIT-06-2014-0044>
- Lu, N., Lin, H., Lu, J., & Zhang, G. (2014). A Customer Churn Prediction Model in Telecom Industry Using Boosting. *IEEE Transactions on Industrial Informatics*, 10(2), 1659–1665. <https://doi.org/10.1109/TII.2012.2224355>
- Makhija, P. (2021). *RFM Analysis for Customer Segmentation | CleverTap*. <https://clevertap.com/blog/rfm-analysis/>
- Mangalathu, S., Hwang, S. H., & Jeon, J. S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures*, 219, 110927. <https://doi.org/10.1016/J.ENGSTRUCT.2020.110927>
- Marcos Cuevas, J. (2018). The transformation of professional selling: Implications for leading the modern sales organization. *Industrial Marketing Management*, 69, 198–208. <https://doi.org/10.1016/J.INDMARMAN.2017.12.017>
- Marr, B. (2019). *What Is Unstructured Data And Why Is It So Important To Businesses? An Easy Explanation For Anyone*. Forbes. <https://www.forbes.com/sites/bernardmarr/2019/10/16/what-is-unstructured-data-and-why-is-it-so-important-to-businesses-an-easy-explanation-for-anyone/?sh=528e9fe315f6>
- Martínez, A., Schmuck, C., Pereverzyev, S. B., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, 281, 588–596. <https://doi.org/10.1016/j.ejor.2018.04.034>
- Mattila, M., Hautamäki, P., Yrjölä, M., & Aarikka-Stenroos, L. (2020). *Business-to-Business Selling in Transition: A Digital Dynamic Managerial Capability Framework*. Journal of Finnish Studies. [https://www.shsu.edu/eng_ira/finnishstudies/Finnish Tables of Content/JoFs_Vol 23.2.pdf#page=162](https://www.shsu.edu/eng_ira/finnishstudies/Finnish%20Tables%20of%20Content/JoFs_Vol%2023.2.pdf#page=162)
- Mattila, M., Yrjölä, M., & Hautamäki, P. (2021). Digital transformation of business-to-business sales: what needs to be unlearned? *Journal of Personal Selling & Sales Management*, 41(2), 113–129. <https://doi.org/10.1080/08853134.2021.1916396>
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*,

22(3), 276. <https://doi.org/10.11613/bm.2012.031>

- Meyer, C., & Schwager, A. (2007). *Understanding Customer Experience*. Harvard Business Review. <https://hbr.org/2007/02/understanding-customer-experience>
- Miguéis, V., Camanho, A., & Falcão Cunha, J. (2013). Customer attrition in retailing: An application of Multivariate Adaptive Regression Splines. *Expert Systems with Applications*, 40(16), 6225–6232. <https://doi.org/10.1016/j.eswa.2013.05.069>
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98, 261–276. <https://doi.org/10.1016/J.JBUSRES.2019.01.044>
- Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, 57(2), 103169. <https://doi.org/10.1016/J.IM.2019.05.004>
- Mustak, M., Salminen, J., Plé, L., & Wirtz, J. (2021). Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda [Article]. *Journal of Business Research*, 124, 389–404. <https://doi.org/10.1016/j.jbusres.2020.10.044>
- Mutanen, T., Nousiainen, S., & Ahola, J. (2010). Customer churn prediction – a case study in retail banking. *Frontiers in Artificial Intelligence and Applications*, 218, 77–83. <https://doi.org/10.3233/978-1-60750-633-1-77>
- Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2008). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592–2602. <https://doi.org/10.1016/j.eswa.2008.02.021>
- Óskarsdóttir, M., Van Calster, T., Baesens, B., Lemahieu, W., Vanthienen, J., & Leuven, K. U. (2018). Time series for early churn detection: Using similarity based classification for dynamic networks. *Expert Systems With Applications*, 106, 55–65. <https://doi.org/10.1016/j.eswa.2018.04.003>
- Paschen, J., Kietzmann Peter, J. B., & Christian Kietzmann, T. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business and Industrial Marketing*, 34(7), 1410–1419. <https://doi.org/10.1108/JBIM-10-2018-0295>
- Patil, A. P., Deepshika, M. P., Mittal, S., Shetty, S., Hiremath, S. S., & Patil, Y. E. (2018). Customer churn prediction for retail business. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing*,

- ICECDS 2017, 845–851. <https://doi.org/10.1109/ICECDS.2017.8389557>
- Paukkeri, R. (2013). *Tilastolliset oppimisyhdistelmät asiakasvaihtuvuuden ennustamisessa*. Tampereen Yliopisto. <https://urn.fi/URN:NBN:fi:uta-201401171046>
- Payne, A., Frow, P., & Marketing, R. (2005). A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, 69, 167–176. <https://doi.org/10.1509/jmkg.2005.69.4.167>
- Peppers, D., & Rogers, M. (2016). *Managing Customer Relationships: A Strategic Framework: Third Edition*. John Wiley & Sons, Incorporated. <https://doi.org/10.1002/9781119239833>
- Pfeifer, P., Haskins, M., & Conroy, R. (2005). *Customer lifetime value, customer profitability, and the treatment of acquisition spending*. *Journal of Managerial Issues*. <https://www.jstor.org/stable/40604472>
- Raab, G., Ajami, R. A., & Goddard, G. J. (2016). *Customer Relationship Management*. <https://doi.org/10.4324/9781315575636>
- Rakennusteollisuus. (2020). *Rakennusteollisuus suhdannekuvaajat*. <https://www.rakennusteollisuus.fi/Tietoa-alasta/Talous-tilastot-ja-suhdanteet/Kuviopankki/Keskeiset-suhdannekuvaajat/>
- Ramadan, R. (2017). Big Data Tools-An Overview. *International Journal of Computer & Software Engineering*, 2(2), 125–140. <https://doi.org/10.15344/2456-4451/2017/125>
- RaSi Ry. (2022). <https://www.rasi.fi/uutiset.html>
- Rautio, A. (2019). *Churn Prediction in SaaS using Machine Learning*. Tampereen Yliopisto. <https://urn.fi/URN:NBN:fi:tty-201905211686>
- Reinartz, W. J., & Kumar, V. (2000). On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. *Journal of Marketing*, 64, 17–35. <https://doi.org/https://doi.org/10.1509/jmkg.64.4.17.180>
- Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, 41, 293–305. <https://doi.org/10.1509/jmkr.41.3.293.35991>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?” Explaining the Predictions of Any Classifier. *KDD '16: Proceedings of the 22nd ACM SIGKDD*

- International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
<https://doi.org/10.1145/2939672.2939778>
- Richards, K. A., & Jones, E. (2008). Customer relationship management: Finding value drivers. *Industrial Marketing Management*, 37(2), 120–130.
<https://doi.org/10.1016/J.INDMARMAN.2006.08.005>
- Ringbeck, D., Smirnov, D., & Huchzermeier, A. (2019). Proactive Retention Management in Retail: Field Experiment Evidence for Lasting Effects. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.3378498>
- Sagioglu, S., & Sinanc, D. (2013). Big data: A review. *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*, 42–47. <https://doi.org/10.1109/CTS.2013.6567202>
- Schoenherr, T., Logistics, C. S.-J. of B., & 2015, U. (2005). Data science, predictive analytics, and big data in supply chain management: Current state and future potential. *Journal of Business Logistics*, 36(1), 120–132.
<https://doi.org/10.1111/jbl.12082>
- Shi, H., Sridhar, S., Grewal, R., & Lilien, G. (2016). Sales representative departures and customer reassignment strategies in business-to-business markets. *Journal of Marketing*, 81(2), 25–44. <https://doi.org/10.1509/jm.15.0358>
- Stein, A. D., Smith, M. F., & Lancioni, R. A. (2013). The development and diffusion of customer relationship management (CRM) intelligence in business-to-business environments. *Industrial Marketing Management*, 42(6), 855–861.
<https://doi.org/10.1016/J.INDMARMAN.2013.06.004>
- Stein, A., & Smith, M. (2009). CRM systems and organizational learning: An exploration of the relationship between CRM effectiveness and the customer information orientation of the firm in industrial markets. *Industrial Marketing Management*, 38(2), 198–206.
<https://doi.org/10.1016/J.INDMARMAN.2008.12.013>
- Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, 57(1), 103146. <https://doi.org/10.1016/J.IM.2019.02.001>
- Tsai, M. T., Tsai, C. L., & Chang, H. C. (2010). The effect of customer value, customer satisfaction, and switching costs on customer loyalty: An empirical study of hypermarkets in Taiwan. *Social Behavior and Personality*, 38(6), 729–740.
<https://doi.org/10.2224/SBP.2010.38.6.729>

- Tullis, T., & Albert, B. (2013). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics: Second Edition. In *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics: Second Edition*. Elsevier Inc. <https://doi.org/10.1016/C2011-0-00016-9>
- Tuominen, H., Neittaanmäki, P., Niinimäki, E., Pölönen, I., Rautiainen, I., Äyrämö, S., Ruohonen, T., Nyrhinen, R., Ojalainen, A., Vähäkainu, P., & Äyrämö, S.-M. (2019). *Tekoälyn perusteita ja sovelluksia*. Jyväskylän Yliopisto. <https://jyx.jyu.fi/handle/123456789/64975>
- Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9. <https://doi.org/10.1016/J.SIMPAT.2015.03.003>
- Valluri, C. (2019). *The Many Types of Churn and Their Predictive Models*. <https://www.proquest.com/central/docview/2314065359/fulltextPDF/97042EF07F024DF4PQ/2?accountid=14242>
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211–229. <https://doi.org/10.1016/J.EJOR.2011.09.031>
- Verhoef, P. C. (2003). Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development. *Journal of Marketing*, 30, 30–45. <https://doi.org/10.1509/jmkg.67.4.30.18685>
- Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, 261(2), 626–639. <https://doi.org/10.1016/J.EJOR.2017.02.023>
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. fan, Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365. <https://doi.org/10.1016/J.JBUSRES.2016.08.009>
- Wengler, S., Überwimmer, M., Hautamäki, P., Hildmann, G., Vossebein, U., & Fueederer, R. (2020). *Digital transformation in B2B sales: differences and best practices in three different european countries*. Marketing Science & Inspirations. <https://msijournal.com/digital-transformation-in-b2b-sales-differences-and-best-practices-in-three-different-european-countries/>

- Wiersema, F. (2012). The B2B Agenda: The current state of B2B marketing and a look ahead ☆. *Industrial Marketing Management*, 42(4), 470–488.
<https://doi.org/10.1016/j.indmarman.2013.02.015>
- Zablah, A. R., Bellenger, D. N., & Johnston, W. J. (2004). An evaluation of divergent perspectives on customer relationship management: Towards a common understanding of an emerging phenomenon. *Industrial Marketing Management*, 33(6), 475–489. <https://doi.org/10.1016/J.INDMARMAN.2004.01.006>
- Zerbino, P., Aloini, D., Dulmin, R., & Mininno, V. (2017). Big Data-enabled Customer Relationship Management: A holistic approach. *Information Processing & Management*, 54(5), 818–846. <https://doi.org/10.1016/j.ipm.2017.10.005>
- Zhang, C., Wang, X., Cui, A. P., & Han, S. (2020). Linking big data analytical intelligence to customer relationship management performance. *Industrial Marketing Management*, 91, 483–494.
<https://doi.org/10.1016/j.indmarman.2020.10.012>