

Onni Vatanen

Tiedonlouhinta ja asiakkuudenhallinta

Tietotekniikan kandidaatintutkielma

30. huhtikuuta 2021

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Onni Vatanen

Yhteystiedot: onni.ei.vatanen@student.jyu.fi

Ohjaaja: Antti-Jussi Lakanen

Työn nimi: Tiedonlouhinta ja asiakkuudenhallinta

Title in English: Data mining and customer relationship management

Työ: Kandidaatintutkielma

Opintosuunta: Tietotekniikka

Sivumäärä: 23+0

Tiivistelmä: Tässä kandidaatintutkielmassa selvitetään, miten tiedonlouhinta on hyödynnetty asiakkuudenhallinnan edistämässä. Tutkielmassa tuodaan esille miten tiedonlouhinta on suoritettu asiakkuudenhallintajärjestelmässä ja minkälaisia tuloksia sen avulla on saatu. Tuloksista saatujen tietojen avulla esitellään keinoja asiakkuudenhallinnan edistämisen suhteen.

Avainsanat: Tiedonlouhinta, Asiakkuudenhallinta, Asiakkuudenhallintajärjestelmä, CRM

Abstract: Subject of this bachelors' thesis is to find out how data mining is utilized in developing customer relationship management practices. This thesis brings up the way data mining has been carried out in a customer relationship management system and the results of that procedure are presented. Ways to develop customer relationship management practices are introduced based on the data gathered from the results.

Keywords: Data mining, Customer relationship management, Customer relationship management system, CRM

Kuviot

Kuvio 1. Asiakkuudenhallinnan eri ulottuvuudet. (Liu ja Zhu 2009)	3
Kuvio 2. Tiedon löytämisen tietokannoista-prosessi (KDD) (Tan, Steinbach ja Kumar 2016).....	7
Kuvio 3. Visualisointia K-means-klusterointialgoritmin toiminnasta (Han, Kamber ja Pei 2011).....	8
Kuvio 4. Graafi lopputuloksista neljällä klusterilla. (Arumawadu, Rathnayaka, Illangarathne ym. 2015)	11

Taulukot

Taulukko 1. Taulukko klusteroinnin tuloksista. (Arumawadu, Rathnayaka, Illangarathne ym. 2015)	12
Taulukko 2. Klustereiden lopulliset arvot. (Arumawadu, Rathnayaka, Illangarathne ym. 2015).....	13

Sisällys

1	JOHDANTO	11
2	ASIAKKUUDENHALLINTA	2
2.1	Mitä asiakkuudenhallinta on	2
2.2	Asiakkuudenhallintajärjestelmä	3
2.3	Analyyttinen asiakkuudenhallinta	4
3	TIEDONLOUHINTA.....	6
3.1	Tiedonlouhinnan määrittelyä	6
3.2	Klusterointi	7
3.3	K-means-algoritmi	8
3.4	RFM-analyysi	9
4	K-MEANS-KLUSTEROINTIALGORITMIN HYÖDYNTÄMINEN ASIAKKUUDENHALLINNASSA	10
4.1	Tutkimuksen läpikäynti	10
4.1.1	RFM-analyysi	10
4.1.2	Klusteroiminen.....	11
4.1.3	Parhaan arvon K määrittely vääristymäkäyrällä.....	12
4.1.4	Painotettujen RFM-arvojen määrittely klustereille.....	12
4.1.5	Tulosten analysointi	12
4.2	Pohdintaa lopputuloksista.....	13
5	YHTEENVETO.....	15
	LÄHTEET	16

1 Johdanto

Tiedon painoarvo yritysmaailmassa kasvaa jatkuvalla tahdilla. Organisaatiot pyrkivät lisäämään tietoa asiakkaistaan, jotta ne pystyvät parantamaan tapojansa asiakkuudenhallinnan toteutuksessa. Asiakkuudenhallintaa on esimerkiksi asiakassuhteiden ylläpitäminen, uusasiakashankinta sekä markkinoinnin kehittäminen (Chalmers [2006](#)).

Tiedon hankkiminen ei ole yksinkertainen prosessi. Jo olemassa olevasta tiedosta voidaan eri menetelmin hankkia uutta tietoa. Digitaalisessa ympäristössä uutta tietoa voidaan hankkia muun muassa tiedonlouhinnan avulla. Tiedonlouhintaa varten tiedon täytyy olla eheää ja sellaisessa muodossa, että sen käsittely on mahdollista. Asiakkuudenhallintajärjestelmä mahdollistaa liiketoimintaan liittyvän tiedon keskittämisen sekä tiedon eheyden ylläpitämisen (Nguyen, Sherif ja Newby [2007](#)).

Tutkielman tarkoituksena on selvittää minkälaista tietoa asiakasdatasta on saatu louhittua. Tätä tarkastellaan valitun K-means-klusterointialgoritmin pohjalta suoritetusta tutkimuksesta. Pyrkimys on selvittää, miten klusteroinnilla saadun tiedon avulla yrityksen harjoittamaa asiakkuudenhallintaa voidaan kehittää.

Tutkielma alkaa ensimmäisestä luvusta eli johdannosta. Toisessa luvussa käydään läpi asiakkuudenhallintaa, sekä siihen liittyviä tutkielman kannalta oleellisia käsitteitä. Tätä seuraa kolmannessa luvussa käsitellään tiedonlouhintaa sekä klusterointia eli tiedonlouhinnan mallia. Luvussa myös esitellään klusterointialgoritmi K-means sekä tietokantamarkkinoinnin työkalu RFM-analyysi. Neljännessä luvussa kerrotaan tutkielman kohdetutkimuksen suorittamisesta, sekä minkälaisia tuloksia K-means-klusterointialgoritmiä käyttämällä saatiin. Luvussa myös pohditaan tulosten merkityksellisyyttä asiakkuudenhallinnan kehittämisen kannalta. Tutkielman viides luku on yhteenveto tutkielmasta.

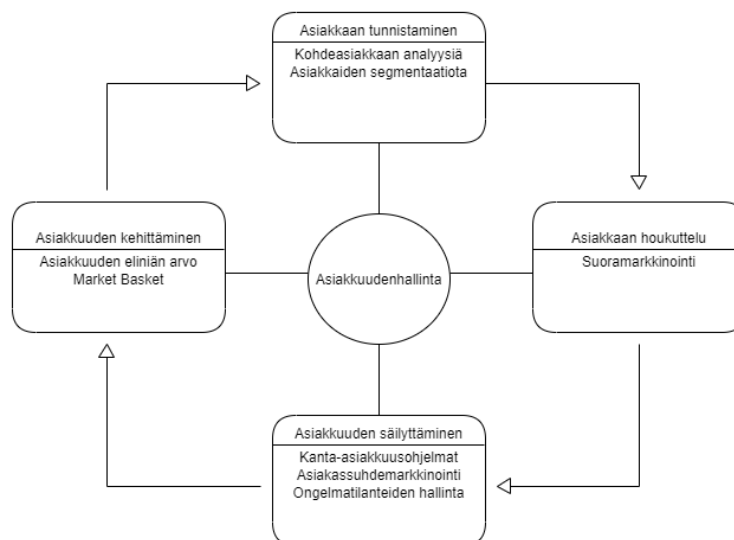
2 Asiakkuudenhallinta

Tässä luvussa käsitellään asiakkuudenhallintaa yleisellä tasolla. Luvussa tuodaan esille asiakkuudenhallinnan määrittelytapoja, millaisin menetelmin asiakkuudenhallintaa toteutetaan sekä tarkastellaan asiakkuudenhallintajärjestelmiä ja niiden merkitystä asiakkuudenhallinnan toteutuksessa.

2.1 Mitä asiakkuudenhallinta on

Asiakkuudenhallintaa (engl. Customer relationship management, CRM) voidaan tarkastella informaatioteknologian näkökulmasta sekä hallinnollisesta näkökulmasta (Buttle ja Maklan 2019). Hallinnollisessa näkökulmassa asiakkuudenhallinta nähdään liiketoimintastrategiana, jolla pyritään maksimoimaan kannattavuutta, tuottoa sekä asiakastyytyväisyyttä. Näitä pyrkimyksiä toteutetaan asiakaslähtöisillä prosesseilla ja luomalla kokonaisvaltaisen kuvan asiakkuuksista. Informaatioteknologiaan nojaavat määritelmät taas kuvaavat asiakkuudenhallinnan olevan termi ohjelmistoille, asiakkuudenhallintajärjestelmille sekä muille ominaisuuksille, joiden avulla yritykset pystyvät hallitsemaan asiakassuhteitaan organisoidusti (Buttle ja Maklan 2019). Swiftin (2001) näkemyksen mukaan asiakkuudenhallinnan strategia pitää sisällään neljä eri ulottuvuutta: Asiakkaan tunnistuksen, asiakkaan houkuttelun, asiakkuuden säilyttämisen sekä asiakkuuden kehittymisen (ks. Kuvio 1).

Asiakkuudenhallinta alkaa asiakkuuden tunnistamisella. Tämä vaihe sisältää keskittymistä siihen massaan, joista tulee todennäköisimmin asiakkaita tai ovat tuottoisimpia yritykselle. Tunnistamisvaiheessa suoritetaan asiakkaiden segmentointia, eli potentiaalisen asiakasmassan jakamista pienempiin ryhmiin tai segmentteihin siten, että jonkin tietyn ryhmän tai segmentin keskinäiset jäsenet ovat määritellyn kriteerin mukaisesti samankaltaisia keskenään (Woo, Bae ja Park 2005). Seuraava vaihe on asiakkaan houkuttelu. Kun potentiaalisten asiakkaiden segmentit on muodostettu, voidaan niiden houkuttelua varten kohdentaa resursseja suoraan. Yksi esimerkki tällaisten resurssien kohdentamisesta on suoramarkkinointi, jossa motivoidaan asiakkaita ostamaan palveluita tai tuotteita eri kanavien kautta (Cheung ym. 2003). Kun asiakkuussuhde on muodostunut, seuraa asiakkuuden säilyttämisen vaihe.



Kuvio 1. Asiakkuudenhallinnan eri ulottuvuudet. (Liu ja Zhu [2009](#))

Asiakastyytyväisyys on asiakkuuden säilyvyyden keskeinen elementti. Asiakkuuden säilyvyyttä edistetäänkin kanta-asiakkuusohjelmilla, asiakassuhdemarkkinoinnilla sekä ongelmatilanteiden asiakaspalvelulla. Viimeiseen vaiheeseen eli asiakkuuden kehittämiseen liittyy asiakkuuden eliniän arvo, ylä- ja ristimyynti sekä market basket-analyysi (Liu ja Zhu [2009](#)).

2.2 Asiakkuudenhallintajärjestelmä

Asiakkuudenhallintajärjestelmällä (engl. customer relationship management software) tarkoitetaan tietojärjestelmiä, joiden pääasiallisina käyttötarkoituksina on toteuttaa asiakkuudenhallintaa käytännön tasolla, eli hallinnoida nykyisten sekä potentiaalisten tulevien asiakkaiden tietoja eri tavoin (Rygielski, Wang ja Yen [2002](#), ss. 491-492). Yleisesti käytettyjä asiakkuudenhallintajärjestelmiä ovat muun muassa Salesforce, Microsoft Dynamics, HubSpot sekä Pipedrive.

Asiakkuudenhallintajärjestelmät voidaan toimintojensa perusteella jakaa kolmeen eri kategoriaan: Yhdistäviin toimintoihin, operatiivisiin toimintoihin sekä analyttisiin toimintoihin (Torggler [2009](#)). Yhdistävät toiminnot pitävät sisällään kaikki yrityksen ja asiakkaan väliset kommunikaatiokanavat. Eri kommunikaatiokanavien käyttö mahdollistaa informaationvaihdon optimoinnin yrityksen ja asiakkaan kesken. Asiakkuudenhallintajärjestelmän operatiivi-

siin toimintoihin kuuluu asiakkuudenhallintaa ylläpitävien toimien suunnittelu sekä implementaatio. Tällaisia toimintoja ovat esimerkiksi markkinointiautomaatio sekä myynnin automaatio. Analyttiset toiminnot ovat keskitetysti asiakaskeskeisten päätöksentekoprosessien valmistelua, simulointia, analysointia sekä optimointia. Nämä toiminnot pohjautuvat asiakasdatasta saatuun tietoon, jota kerätään markkinoinnin, myynnin ja asiakaspalvelun prosesseista (Torggler 2009). Nämä kategoriat eivät ole toisiaan poissulkevia, vaan yhdessä ja samassa järjestelmässä on yleensä eri kategorioiden toimintoja.

Asiakkuudenhallintajärjestelmä tarjoaa alustan tiedon hallinnoinnille ja sen käsittelylle. Tämä hallinnoitava tieto täytyy kerätä jostakin lähteestä. Tämä tieto voidaan kerätä muista tietojärjestelmistä. Ahn, Kim ja Han (2003) määrittelevät nämä kerättävät tiedot kolmeen kategoriaan: lokitietoon, käyttäjätietoon sekä markkinatietoon. Lokitiedot ovat palvelimilta saatuja tietoja, jotka sisältävät tietoa käyttäjän toiminnasta verkkosivulla. Käyttäjätiedot pitävät sisällään yksilöiviä tietoja käyttäjästä itsestään, esimerkiksi käyttäjätunnuksen, nimen, iän sekä sukupuolen. Käyttäjätietojen kerääminen mahdollistaa tarkan käyttäjäkuvauksen muodostamisen. Viimeisen kategoria eli markkinatieto sisältää tietoa esimerkiksi yrityksen tuotteista sekä niiden menekistä verkkoympäristössä. Asiakkuudenhallintajärjestelmä mahdollistaa eri tietojärjestelmien tuottaman tiedon yhdistämisen saman järjestelmän alle. Tämä avulla tiedon esikäsittelyn voi keskittää. Famili ym. (1997) toteavat, että lähteet joissa tietoa käsitellään ja sen jälkeen saadaan on hajontaa sisältävää tietoa, joka vaikuttaa sen esikäsittelyyn.

2.3 Analyttinen asiakkuudenhallinta

Analyttinen asiakkuudenhallinta (engl. Analytical CRM) on olemassa olevan datan analysoimista eri työkaluilla ja prosesseilla, jotta datasta voidaan generoida esimerkiksi asiakasprofieja, tunnistaa käyttäytymismalleja, tyytyväisyystasoja sekä tukea asiakassegmentointia (Xu ja Walton 2005, sivu 961). Nämä työkalut mahdollistavat suurtenkin datamassojen analysoinnin (Chen ja Popovich 2003).

Analyttistä asiakkuudenhallintaa suoritettaessa käytetään erilaisia tekniikoita, esimerkiksi luokittelua, klusterointia, sekvenssien louhimista (engl. sequential pattern) sekä mallin löy-

tämisen assosiaatiosääntöjä (engl. pattern discovery association rules). Analyttisellä asiakkuudenhallinnalla pyritään saavuttamaan selkeämpää asiakasryhmien segmentaatiota, jonka myötä tuotteita ja palveluita voidaan kohdentaa paremmin (Xu ja Walton [2005](#), sivu 961).

3 Tiedonlouhinta

Seuraavassa luvussa käsitellään tiedonlouhintaa. Luvussa esitellään tiedonlouhinnan kategoriaita, sekä avataan tiedonlouhinnan mallin, klusteroinnin, käsitettä. Luvussa myös esitellään K-means-klusterointialgoritmi ja käydään läpi sen toimintaa. Luvun viimeisessä kappaleessa käsitellään RFM-analyysia.

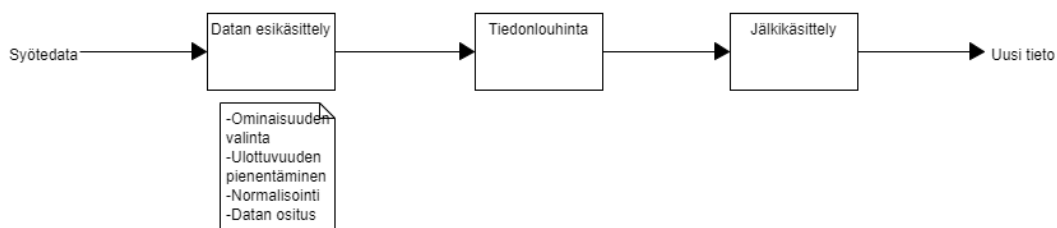
3.1 Tiedonlouhinnan määrittelyä

Tiedonlouhinta (engl. Data mining) tarkoittaa prosessia, jonka avulla kerätään kiinnostavia elementtejä suuresta määrästä tietoa. Tällaista tietoa on esimerkiksi tiedon muodostamat erilaiset kuviot, johdonmukaisuudet sekä eri tietueiden mahdolliset keskinäiset suhteet. Tiedonlouhinta yhdistää prosessissaan eri työkaluja tilastotieteistä, koneoppimisesta sekä tiedonhallintajärjestelmistä (Chen, Han ja Yu [1996](#), ss. 886-887). Tiedonlouhinta voidaan jakaa kolmeen suurpiirteiseen kategoriaan prosessinkuvauksen kannalta.

Löytävän (engl. Discovery) kategorian alle kuuluu prosessit, joiden avulla tietokannasta etsitään havaitsemattomia kuvioita ilman, että niistä oltaisiin muodostettu ennalta määritettyjä hypoteesejä, mitä nämä kuviot voisivat olla. Toinen kategoria on ennakoiva mallinnus (engl. Predictive modeling). Tämän alle kuuluvien prosessien tarkoituksena on etsiä olemassa olevia kaavoja tietokannasta, joiden avulla voidaan ennustaa tulevia tapahtumia. Kolmas tiedonlouhinnan kategoria on tutkiva analyysi (engl. Forensic analysis), jonka prosessien tarkoitus on käyttää tiedosta saatuja kuvioita anomalioiden ja epätavallisten dataelementtien löytämiseen.

Tiedonlouhinta on keskeinen osa KDD:ta, eli tiedon löytämistä tietokannoista. Se on kokonaisvaltainen prosessi raa'an datan muuttamisesta käytettävään muotoon (ks. Kuvio 2). Tämä prosessi pitää sisällään useita eri vaiheita, joissa dataa esikäsitellään ja jälkikäsitellään (Tan, Steinbach ja Kumar [2016](#)).

Tiedonlouhinta on kuitenkin vain yksi työkalu tiedon kartuttamista varten. Se ei poista tarvetta ymmärtää tiedonlouhinnan kohteena olevan tiedon taustoja. Tiedonlouhinnan avulla



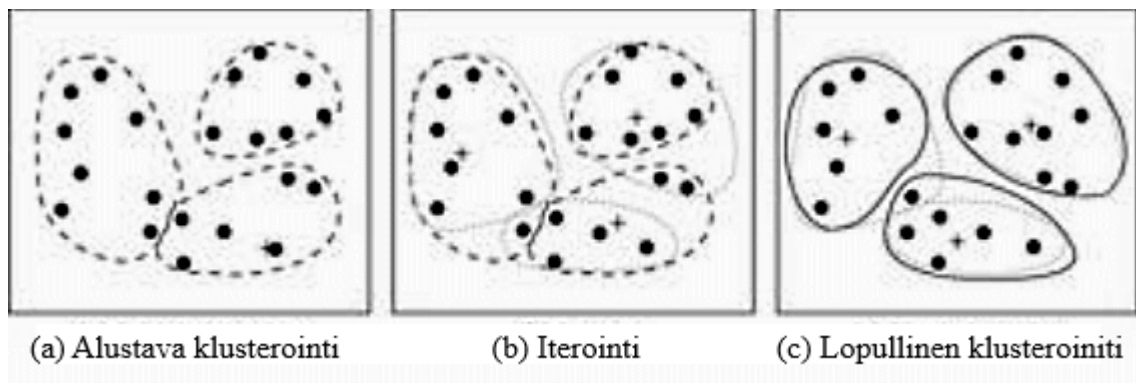
Kuvio 2. Tiedon löytämisen tietokannoista-prosessi (KDD) (Tan, Steinbach ja Kumar [2016](#)).

saatu tieto vaatii syvällistä perehtymistä, jotta se voidaan todentaa relevantiksi. Tiedonlouhinta auttaa hypoteesien luomisessa eri skenaarioissa, mutta sen ei tule toimia hypoteesin pohjana (Rygielski, Wang ja Yen [2002](#)).

3.2 Klusterointi

Klusterointi on tiedonlouhinnan malli, jonka tarkoituksena on tunnistaa kuvioita tai klustereita, eli niin sanottuja ryppäitä olemassa olevan tiedon pohjalta. Yhdessä klusterissa oleva tieto on keskenään samanlaista jollakin yhdistävällä tekijällä. Tämä tieto on poikkeavaa toisesta klusterista löytyvän tiedon kanssa. Klusterointia suoritetaan usein pohjatyönä jotakin muuta tiedonlouhinnan prosessia tai mallinnusta varten (Berry ja Linoff [2004](#), sivu 11).

Klusterointialgoritmit voidaan jakaa neljään eri metodiin: Hierarkkisiin metodeihin, osittaviin metodeihin, tiheyspohjaisiin metodeihin sekä ruudukkopohjaisiin metodeihin. Hierarkiset menetit voidaan jaotella joko kasaaviksi tai jakaviksi metodeiksi riippuen siitä, miten hierarkinen hajonta on muodostettu. Kasaavassa metodissa jokainen dataobjekti muodostaa aluksi oman ryhmänsä ja ne alkavat yhdistymään muodostaen lopulta yhden ryhmän. Jakavassa metodissa dataobjektit taas muodostavat yhden ryhmän, joka aloittaa jakaantumisen. Jakaantuminen jatkuu, kunnes jokainen dataobjekti muodostaa oman ryhmänsä. Osittavat menetit jakavat dataobjektit annettuun määrään klustereita jonkin metodin mukaan. Yleinen metodi etäisyyden määrittelyyn on Euklidinen etäisyys (Saxena ym. [2017](#)). Tiheyspohjaiset menetit voivat jakaa dataobjektit useaan eksklusiivisiin klustereihin tai hierarkisiin klustereihin. Eksklusiivisen klusterin dataobjektit eivät voi kuulua useampaan klusteriin. Ruudukkopohjaiset menetit muodostavat määritellyn määrän soluja, jotka muodostavat ruudukkorakenteen. Kaikki klusterointioperaatiot toteutetaan tällä ruudukkorakenteella. Ruudukko-



Kuvio 3. Visualisointia K-means-klusterointialgoritmin toiminnasta (Han, Kamber ja Pei 2011).

rakenteen etuna on nopea prosessointiaika. Ruudukkopohjaisia metodeita voidaan yhdistää muihin klusterointimethodeihin, esimerkiksi tiheyspohjaisiin metodeihin (Han, Kamber ja Pei 2011).

3.3 K-means-algoritmi

K-means on osittava klusterointialgoritmi. Se toimii iteratiivisesti, eli se toistaa samaa työvaihetta useampaan kertaan kunnes sille asetetut odotukset täyttyvät. K-means on suoritusajaltaan nopea ja sen on huomattu tuottavan hyviä lopputuloksia monissa käytännönläheisissä applikaatioissa (Na, Xumin ja Yong 2010, sivu 63). K-means osittaa määritellyt dataobjektit haluttuun määrään k klustereita. Peruseriaate on muodostaa aluksi satunnaisesti klusterit, joita parannetaan päivittämällä prototyyppejä, kunnes saavutetaan lokaali optimi.

Algoritmin ensimmäisessä osassa valitaan k painopisteet sattumanvaraisesti, joissa arvo k eli klustereiden määrä on vakio. Euklidista etäisyyttä käytetään määrittelemään etäisyyksiä dataobjektien ja painopisteiden välillä. Dataobjekti määritellään kuuluvan siihen klusteriin, minkä painopisteeseen sillä on lyhin etäisyys. Kun jokainen dataobjekti on sisällytetty johonkin klusteriin, on ensimmäinen osa valmis ja alustavat klusterit ovat muodostettu. Tämän jälkeen lasketaan alustavien klustereiden keskiarvo, joista muodostetaan uudet painopisteet. Tätä kaavaa toistetaan, kunnes klustereiden painopisteissä ei tapahdu enää muutoksia (ks. Kuvio 3) (Na, Xumin ja Yong 2010, sivu 63).

3.4 RFM-analyysi

RFM-analyysi on tietokantamarkkinoinnin työkalu, jota käytetään yleisesti mittaamaan asiakkuuksien arvoa olemassa olevan tiedon pohjalta (Hu ja Yeh [2014](#)). RFM-akronyymi muodostuu seuraavista englanninkielisistä sanoista: Recency (suom. viimeaikaisuus), frequency (suom. toisteisuus) ja monetary (suom. rahallinen arvo). Viimeaikaisuus mittaa aikaväliä kahden viimeaikaisen tapahtuman välillä, esimerkiksi kahden suoritetun ostotoimenpiteen välisien päivien lukumäärää. Toisteisuus tarkoittaa tapahtumien numeerista määrää jollakin annetulla aikaikkunalla, esimerkiksi edellä mainittujen ostotoimenpiteiden määrää viimeisen kuukauden aikana. Rahallinen arvo mittaa näiden esimerkkitapausten kokonaisarvoa määritellyllä aikavälillä. RFM-analyysin vahvuuksiin lukeutuu sen indikaattorien sovellettavuus eri toimialoille (Dursun ja Caber [2016](#)).

RFM-analyysia on hyödynnetty onnistuneesti viime vuosina osana tiedonlouhinnan menetelmiä (Hsieh [2004](#)). RFM-analyysin indikaattorien soveltaminen klusteroinnissa tuottaa enemmän tietoa asiakkaiden markkinointitason sijainnista verrattuna muihin klusterointianalyysiin (Birant [2011](#)). RFM-analyysin integroiminen tiedonlouhinnan eri menetelmien kanssa lisää relevantin tiedon määrää ja täten tuo lisäarvoa tiedonlouhintaa sekä markkinointia suorittaville osapuolille.

4 K-means-klusterointialgoritmin hyödyntäminen asiakkuudenhallinnassa

Tässä luvussa tarkastellaan tutkimusta, jossa on käytetty K-means-klusterointialgoritmia asiakkuudenhallintajärjestelmässä asiakassegmenttien tuottavuuden kartoitukseen. Luvussa käydään läpi tutkimuksen tarpeet, jonka jälkeen kerrotaan kuinka klusterointi toteutettiin. Lopuksi vielä käydään läpi tutkimuksesta saadut tulokset (Arumawadu, Rathnayaka, Illangarathne ym. [2015](#)).

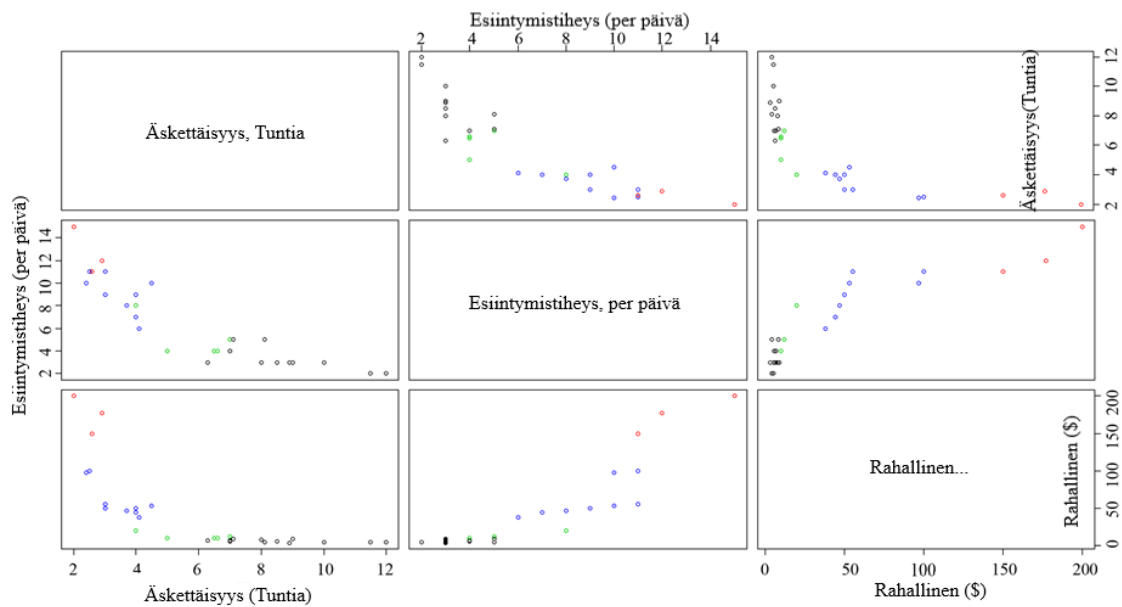
4.1 Tutkimuksen läpikäynti

Arumawadu, Rathnayaka, Illangarathne ym. [\(2015\)](#) käyttävät tutkimuksessaan K-means-klusterointialgoritmia, jonka avulla he jaottelevat asiakkaat tuottavuuden perusteella eri segmentteihin. Tutkimus suoritettiin telekommunikaatioalan yrityksen asiakasdatan avulla. Asiakkaiden tuottavuudella on suora vaikutus yrityksen liikevaihtoon ja tuottoon. Arumawadu, Rathnayaka, Illangarathne ym. [\(2015\)](#) mukaan on tärkeää pystyä kategorisoimaan asiakkaat tuottavuuden mukaan. Tämä asiakkuuksien segmentaatio auttaa markkinoinnin suunnittelussa eri asiakasryhmien kesken (McDonald [2015](#)).

Tutkimuksen metodologia on jaoteltu viiden eri vaiheen alle. Ensimmäisessä vaiheessa suoritetaan RFM-analyysi. Toisessa vaiheessa jaetaan data RFM-analyysin avulla määriteltyihin segmentteihin K-means-klusterointialgoritmia käyttämällä. Kolmannessa vaiheessa määritellään paras klustereiden määrä k vääristymäkäyrää käyttäen. Neljännessä vaiheessa määritellään painotetut RFM-arvot jokaiselle klusterille, jonka jälkeen viimeisessä vaiheessa tarkasteellaan jokaisen klusterin asiakasryhmän tuottavuutta.

4.1.1 RFM-analyysi

Tutkimus alkaa RFM-analyysillä, jonka avulla voidaan määritellä ja mitata muuttujia asiakkaiden ostokäyttäytymisestä. Analyysia sovelletaan telekommunikaation alalla tutkimuksen mukaan seuraavasti: R vastaa keskimääräistä aikaikkunaa kahden puhelun välillä kuukauden



Kuvio 4. Graafi lopputuloksista neljällä klusterilla. (Arumawadu, Rathnayaka, Illangarathne ym. 2015)

aikajaksolla, F vastaa keskimääräistä puhelujen määrää vuorokauden aikana kuukauden aikajaksolla ja M vastaa asiakkaiden puhelinlaskun arvoa kuukauden aikajaksolla. Tutkimuksessa on käytetty sadan asiakkaan tietoja tutkimusaineistona.

4.1.2 Klusteroiminen

Seuraavassa vaiheessa sovelletaan K-Means-algoritmia tutkimusainestoon. Valitaan alustavat klustereiden painopisteet, joiden pohjalta klustereita aletaan muodostamaan. Klustereiden määrä on aluksi sattumanvaraisesti valittu. Tämän jälkeen aletaan käymään läpi dataobjektien listausta. Jokainen dataobjekti siirretään siihen klusteriin, minkä klusterin painopiste on sitä Euklidean etäisyydellä mitattuna lähimpänä. Tämän jälkeen lasketaan vastaanottavan klusterin uusi painopiste tai dataobjektin menettäneen klusterin uusi painopiste. Tätä prosessia toistetaan, kunnes painopisteissä ei tapahdu enää muutoksia.

Kuviossa 4 nähdään graafi klusteroinnin lopputuloksista, kun klustereita on neljä kappaletta. Kuvio on yhdistelmä kuudesta eri graafista. Koordinaatit x ja y edustavat äskettäisyyttä, esiintymistiheyttä sekä rahallista arvoa kuten graafeissa on esitetty.

Taulukko 1. Taulukko klusteroinnin tuloksista. (Arumawadu, Rathnayaka, Illangarathne ym. 2015)

Klusteri	Keskiarvo R	Keskiarvo F	Keskiarvo M
1	8.62	3.33	5.98
2	2.50	12.67	175.67
3	5.82	5.00	12.38
4	3.47	9.00	59.33

4.1.3 Parhaan arvon K määrittely vääristymäkäyrällä

Seuraavaksi määritellään paras arvo K , jotta saataisiin selville optimaalisin klustereiden alustava määrä. Arvoa lähdettiin ratkaisemaan ajamalla standardeja K-means-operaatioita kaikilla k arvoilla välillä 1-15. Tämän jälkeen saaduista tuloksista laskettiin klusterointien vääristymät eri klusterimäärillä, joista tietyllä hajontavälillä etsittiin minimaalista laskua keskiarvoisessa määrässä. Tutkimuksessa todetaan optimaalisen määrän klustereita klusterointialgoritmin alustuksessa olevan kolme klusteria.

4.1.4 Painotettujen RFM-arvojen määrittely klustereille

Taulukko 1 näyttää klusteroinnin tulokset RFM-arvojen mukaan. Arvojen keskinäinen painotus on kuitenkin vielä laskematta, eikä RFM-arvoja voi laskea yhteen samalla painotuksella. Jokaisen klusterin arvo täytyy laskea seuraavanlaisella kaavalla:

$$V_k = W_R(1/\overline{R}_k) + W_F\overline{F}_k + W_M\overline{M}_k \quad (4.1)$$

jossa V_k on klusterin k arvo ja $\overline{R}_k, \overline{F}_k, \overline{M}_k$ ovat RFM-arvojen keskiarvoja. Tässä tutkimuksessa määritellään $W_R\overline{R}_k + W_F\overline{F}_k + W_M\overline{M}_k$ arvot olevan yhtä kuin $W_R = 0.2, W_F = 0.2, W_M = 0.6$.

4.1.5 Tulosten analysointi

Edellisessä kappaleessa esitellyn kaavan pohjalta saadaan laskettua jokaisen klusterin arvo ja sen pohjalta määriteltyä, minkä klusterin asiakassegmentti on kaikista tuottavin. Taulukko 2 osoittaa, että klusteri numero 2 on se asiakassegmentti, joiden asiakkaat ovat kaikista tuottoisampia yritykselle. Seuraavaksi tuottoisin oli klusteri numero 4, sitten klusteri numero 3

Taulukko 2. Klustereiden lopulliset arvot. (Arumawadu, Rathnayaka, Illangarathne ym. 2015)

Klusteri	V_k
1	4.277
2	108.016
3	8.462
4	37.456

ja viimeisenä klusteri numero 1.

4.2 Pohdintaa lopputuloksista

Tutkimuksessa käytettiin K-means-klusterointialgoritmia segmentoimaan telekommunikaatioalan yrityksen asiakkaita eri klustereihin sen mukaan, miten tuottoisia nämä ovat yritykselle. Tämän tiedon myötä yritys voi kehittää harjoittamaansa asiakkuudenhallintaa.

Tutkimuksen tuloksista nähdään, mitkä asiakasryhmät ovat yritykselle tuottoisimpia. Tuottoisimpien asiakkaiden asiakastyytyväisyyteen kannattaa panostaa, sillä se vaikuttaa suoraan asiakassuhteiden pysyvyyteen. Näin yritys pystyy allokoimaan resurssejaan ja vähentämään täten kulujaan (Tsiptsis ja Chorianopoulos 2011). Tuottoisimpien asiakkuuksien suhdetta kannattaa ylläpitää, sillä uusasiakashankinta on paljon vaikeampaa ja kalliimpaa, kuin olemassa olevien asiakssuhteiden ylläpitäminen (Reinartz, Thomas ja Kumar 2005).

Myös vähemmän tuottavien asiakassegmenttien tilannetta on syytä tarkastella. Reinartz, Krafft ja Hoyer (2004) toteavat, että on olemassa asiakkaita joihin käytetään paljon resursseja verrattuna niistä saatavaan tuottoon, sekä asiakkaita joihin panostetaan liian vähän suhteessa potentiaaliin tuottaa yritykselle enemmän. On siis perusteltua tarkastella tutkimuksessa esille nousseita ryhmiä niiden mahdollisesti olemassa olevan potentiaalın näkökulmasta.

Asiakasryhmien tuottoisuuden segmentaatiota voidaan vielä jatkokäsitellä tiedonlouhinnan muilla menetelmillä, esimerkiksi päätöspuu-metodeilla. Muodostettujen asiakassegmenttien ei voida olettaa olevan vakioita ajan kuluessa, vaan niistä voidaan ottaa pidemmällä aikajaksolla uusia tuloksia ja niiden pohjalta ennustaa segmenttien muutosta tai tarkentaa olemassa

olevia segmenttejä (Arumawadu, Rathnayaka, Illangarathne ym. 2015).

5 Yhteenveto

Tutkielmassa tarkasteltiin asiakkuudenhallintaa ja miten sen toimintaa voidaan viedä eteenpäin tiedonlouhinnan klusterointi-menetelmällä. Esitellyn tutkimuksen ja luetun kirjallisuuden perusteella löytyi näyttöjä siitä, miten tiedonlouhinnan avulla voidaan viedä yritystason asiakkuudenhallintaa eteenpäin.

Tutkielmaa voidaan jatkaa analysoimalla muita tiedonlouhinnan menetelmiä ja miten niitä on käytetty asiakkuudenhallinnan edistämiseksi. Tutkielman uskottavuutta heikentää juuri-kin se, että siinä on käytetty vain yhtä tutkimusta tulosten perusteena. Mikäli vastaavanlaisia tutkimuksia olisi useampi joita voisi verrata keskenään, voisi lopputuloksista saada totuudenmukaisemman kuvan.

Lähteet

Ahn, Jeong Yong, Seok Ki Kim ja Kyung Soo Han. 2003. "On the design concepts for CRM system". *Industrial Management & Data Systems*, <https://doi-org.ezproxy.jyu.fi/10.1108/02635570310477370>.

Arumawadu, Hasitha Indika, RM Kapila Tharanga Rathnayaka, SK Illangarathne ym. 2015. "Mining profitability of telecommunication customers using k-means clustering". *Journal of Data Analysis and Information Processing* 3 (03): 63. <https://doi.org/10.4236/jdaip.2015.33008>.

Berry, Michael JA, ja Gordon S Linoff. 2004. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons. ISBN: 9780471470649. <https://books.google.fi/books?id=Ni5nMDO1OfEC>.

Birant, Derya. 2011. "Data mining using RFM analysis". Teoksessa *Knowledge-oriented applications in data mining*. IntechOpen. https://openresearchlibrary.org/ext/api/media/149979dc-cbb9-414b-9cf0-18d6f3348f7f/assets/external_content.pdf.

Buttle, Francis, ja Stan Maklan. 2019. *Customer relationship management: concepts and technologies*. Routledge. ISBN: 9781351016544. <https://books.google.fi/books?id=I5DUDwAAQBAJ>.

Chalmeta, Ricardo. 2006. "Methodology for customer relationship management". *Journal of systems and software* 79 (7): 1015–1024. ISSN: 0164-1212. <https://doi.org/https://doi.org/10.1016/j.jss.2005.10.018>.

Chen, Injazz J, ja Karen Popovich. 2003. "Understanding customer relationship management (CRM)". *Business process management journal*, <https://doi-org.ezproxy.jyu.fi/10.1108/14637150310496758>.

Chen, Ming-Syan, Jiawei Han ja Philip S. Yu. 1996. "Data mining: an overview from a database perspective". *IEEE Transactions on Knowledge and data Engineering* 8 (6): 866–883. <https://doi.org/10.1109/69.553155>.

- Cheung, Kwok-Wai, James T Kwok, Martin H Law ja Kwok-Ching Tsui. 2003. "Mining customer product ratings for personalized marketing". *Decision Support Systems* 35 (2): 231–243. [https://doi.org/https://doi.org/10.1016/S0167-9236\(02\)00108-2](https://doi.org/https://doi.org/10.1016/S0167-9236(02)00108-2).
- Dursun, Aslihan, ja Meltem Caber. 2016. "Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis". *Tourism management perspectives* 18:153–160. <https://doi.org/https://doi.org/10.1016/j.tmp.2016.03.001>.
- Famili, A, Wei-Min Shen, Richard Weber ja Evangelos Simoudis. 1997. "Data preprocessing and intelligent data analysis". *Intelligent data analysis* 1 (1): 3–23. <https://doi.org/10.3233/IDA-1997-1102>.
- Han, Jiawei, Micheline Kamber ja Jian Pei. 2011. "Data mining concepts and techniques third edition". *The Morgan Kaufmann Series in Data Management Systems* 5 (4): 83–124. https://www.academia.edu/download/43034828/Data_Mining_Concepts_And_Techniques_3rd_Edition.pdf.
- Hsieh, Nan-Chen. 2004. "An integrated data mining and behavioral scoring model for analyzing bank customers". *Expert systems with applications* 27 (4): 623–633. ISSN: 0957-4174. <https://doi.org/https://doi.org/10.1016/j.eswa.2004.06.007>.
- Hu, Ya-Han, ja Tzu-Wei Yeh. 2014. "Discovering valuable frequent patterns based on RFM analysis without customer identification information". *Knowledge-Based Systems* 61:76–88. ISSN: 0950-7051. <https://doi.org/https://doi.org/10.1016/j.knosys.2014.02.009>.
- Liu, ChunNian, ja XiaoWen Zhu. 2009. "A study on CRM technology implementation and application practices". *Teoksessa 2009 International Conference on Computational Intelligence and Natural Computing*, 2:367–370. IEEE. <https://doi.org/10.1109/CINC.2009.120>.
- McDonald, Malcolm. 2015. "Market segmentation". *Wiley Encyclopedia of Management*, 1–10. <https://doi.org/10.1002/9781118785317>.
- Na, Shi, Liu Xumin ja Guan Yong. 2010. "Research on k-means clustering algorithm: An improved k-means clustering algorithm". *Teoksessa 2010 Third International Symposium on intelligent information technology and security informatics*, 63–67. Ieee. <https://doi.org/10.1109/IITSI.2010.74>.

- Nguyen, ThuyUyen H, Joseph S Sherif ja Michael Newby. 2007. “Strategies for successful CRM implementation”. *Information Management & Computer Security*, <https://doi.org/10.1108/09685220710748001>.
- Reinartz, Werner, Manfred Krafft ja Wayne D Hoyer. 2004. “The customer relationship management process: Its measurement and impact on performance”. *Journal of marketing research* 41 (3): 293–305. <https://doi.org/10.1509/jmkr.41.3.293.35991>.
- Reinartz, Werner, Jacquelyn S Thomas ja Viswanathan Kumar. 2005. “Balancing acquisition and retention resources to maximize customer profitability”. *Journal of marketing* 69 (1): 63–79. <https://doi.org/10.1509/jmkg.69.1.63.55511>.
- Rygielski, Chris, Jyun-Cheng Wang ja David C Yen. 2002. “Data mining techniques for customer relationship management”. *Technology in society* 24 (4): 483–502. [https://doi.org/10.1016/S0160-791X\(02\)00038-6](https://doi.org/10.1016/S0160-791X(02)00038-6). <https://www.sciencedirect.com/science/article/pii/S0160791X02000386>.
- Saxena, Amit, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding ja Chin-Teng Lin. 2017. “A review of clustering techniques and developments”. *Neurocomputing* 267:664–681. <https://doi.org/https://doi.org/10.1016/j.neucom.2017.06.053>.
- Tan, Pang-Ning, Michael Steinbach ja Vipin Kumar. 2016. *Introduction to data mining*. Pearson Education India. <https://www-users.cs.umn.edu/~kumar001/dmbook/sol.pdf>.
- Torggler, Michael. 2009. “The functionality and usage of CRM systems”. *environment* 41:30–47. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.193.5110&rep=rep1&type=pdf>.
- Tsiptsis, Konstantinos K, ja Antonios Chorianopoulos. 2011. *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons. ISBN: 9781119965459. <https://books.google.fi/books?id=t4ZIKY7sMRsC>.
- Woo, Ji Young, Sung Min Bae ja Sang Chan Park. 2005. “Visualization method for customer targeting using customer map”. *Expert Systems with Applications* 28 (4): 763–772. <https://doi.org/https://doi.org/10.1016/j.eswa.2004.12.041>.

Xu, Mark, ja John Walton. 2005. "Gaining customer knowledge through analytical CRM".
Industrial management & data systems, <https://doi.org/10.1108/02635570510616139>.