



**TURUN
YLIOPISTO**
Kauppakorkeakoulu

Datalouhinta ja data-analyysi verkkokauppojen työkaluna

Laskentatoimen ja rahoituksen
kandidaatintutkielma

Laatija:
Arttu Myllyrinne

Ohjaaja:
KTT Antti Miihkinen

26.03.2024
Turku

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

Kandidatutkielma

Oppiaine: Laskentatoimi ja rahoitus

Tekijä(t): Arttu Myllyrinne

Otsikko: Datalouhinta ja data-analyysi verkkokauppojen työkaluna

Ohjaaja(t): KTT Antti Miihkinen

Sivumäärä: 28

Päivämäärä: 26.03.2024

Tiivistelmä

Tutkielman tarkoituksena on selvittää, minkälainen rooli ja merkitys datalouhinnalla ja data-analyysillä on yrityksissä painottuen verkkokauppoihin. Tähän pyritään vastaamaan kirjallisuuskatsauksen keinoin.

Nykyajan digitaalisessa ympäristössä, jossa datan määrä kasvaa eksponentiaalisesti, yritykset kohtaavat haasteita datan tehokkaassa hyödyntämisessä. Datalouhinta ja data-analyysi tarjoavat menetelmiä, joiden avulla voidaan muuttaa raakadata hyödylliseksi tiedoksi päätöksenteon tueksi. Tutkimuksessa käydään läpi datalouhinnan ja data-analyysin perusteita, niiden sovelluksia verkkokaupoissa sekä hyötyjä ja haasteita, joita näihin teknologioihin liittyy.

Tutkielma osoittaa, että datalouhinta ja data-analyysi voivat merkittävästi parantaa verkkokauppojen toimintaa esimerkiksi asiakasdatan analysoinnin, markkinoinnin kohdentamisen ja tuotevalikoiman optimoinnin kautta. Haasteista huolimatta nämä teknologiat tarjoavat verkkokaupoille mahdollisuuksia saavuttaa strateginen kilpailuetu digitalisoituvilla markkinoilla.

Avainsanat: Datalouhinta, data-analyysi, verkkokauppa

SISÄLLYS

1	Johdanto	7
1.1	Johdatus aiheeseen	7
1.2	Tutkielman tavoite ja rajaus	7
1.3	Menetelmät ja rakenne	8
2	Datalouhinta ja data-analyysi	9
2.1	Data-analyysin määritelmä ja prosessi	9
2.2	Data-analyysin tyypit	9
2.2.1	Tilastoanalyysi	10
2.2.2	Diagnostinen analyysi	10
2.2.3	Ennustava analyysi	10
2.2.4	Preskriptiivinen analyysi	11
2.3	Datalouhinta	11
2.4	Data-analyysin ja datalouhinnan ero	13
3	Datalouhinnan ja data-analyysin hyödyntäminen	14
3.1	Datalouhinnan hyödyt ja haasteet	14
3.1.1	Hyödyt	14
3.1.2	Haasteet	15
3.2	Datalouhinta verkkokaupoissa	16
3.2.1	Asiakasprofilointi ja Personointi	16
3.2.2	Ostoskorin Analyysi	16
3.2.3	Myynnin Ennustaminen	17
3.2.4	Tuotevalikoiman Suunnittelu	17
3.2.5	Markkinasegmentointi	17
3.3	Data-analyysin hyödyt ja haasteet	18
3.4	Data-analyysi verkkokaupoissa	20
3.4.1	Ymmärrys asiakaskäyttäytymisestä	20
3.4.2	Henkilökohtaisen asiakaskokemuksen luominen	20
3.4.3	Markkinoinnin optimointi ja kohdentaminen	20
3.4.4	Tuotevalikoiman ja hinnoittelustrategian hienosäätö	21
3.4.5	Verkkosivuston käytettävyyden ja suorituskyvyn parantaminen	21
4	Johtopäätökset	22
4.1	Hyödyt ja haasteet	22

4.2 Omat näkemykset ja kriittinen arvio	23
5 Yhteenveto	24
Lähteet	27

1 Johdanto

1.1 Johdatus aiheeseen

Arviolta vuonna 2023 internettiin on yhteydessä 15,14 miljardia laitetta ja tämän ennustetaan kasvavan noin 30 miljardiin vuoteen 2030 mennessä (Statista.com 2023). Tämän seurauksena datan määrä maailmassa on kasvanut räjähdysmäisesti samana ajanjaksona, ja datan määrä onkin paisunut niin suureksi, että monet yritykset eivät kykene käsittelemään koko keräämäänsä dataa tehokkaasti. Vuonna 2012 datanmäärän arvioitiin kasvavan 2,5 miljardia gigatavua päivässä (McAfee & Brynjolfsson 2012), kun vuonna 2023 sen arvioitiin kasvavan 328,77 miljardia gigatavua päivässä (Statista.com 2023).

Tämä nopeasti kasvava datamäärä ei ole ainoastaan taakka ja kustannus yrityksille, vaan se tarjoaa myös mahdollisuuden oppia uusia ja yllättäviä seikkoja asiakkaista ja heidän käyttäytymismalleistaan. Tämän uuden tiedon avulla yritykset voivat tehdä perusteltuja päätöksiä liiketoimintansa parantamiseksi ja strategisen kilpailuedun saavuttamiseksi. Provost & Fawcett (2013, 4-7) ovat määritelleet tällaisen dataan perustuvan päätöksenteon päätöksenteoksi, jonka perustana on datan analysoiminen (data-driven decision-making). Tämä eroaa normaalista siinä, että päätöksiä ei ohjaa päätöstentekijöiden kokemukset ja intuitio.

Yritysten olisi suositeltavaa käyttää kaikkea saatavilla olevaa tietoa, mutta nykypäivänä suurten datamäärien analysointi ja käsittely ovat muodostuneet haastaviksi. Tätä varten on alettu hyödyntämään datalouhintaa ja data-analyysia päätösten tukena.

Vaikka aihe on hyvin ajankohtainen ja digitalisaation seurauksena koskettaa lähes kaikkia aloja, on sen vaikutuksista verkkokauppoihin erittäin vähän tutkimuksia. Sen takia on mielekästä lähteä tekemään kyseisestä aiheesta tutkielmaa

1.2 Tutkielman tavoite ja rajaus

Tutkimuksen tarkoitus on syventyä ja esitellä datalouhinnan ja data-analyysin roolia verkkokaupoissa ja arvioida, millaisia hyötyjä ja haasteita nämä teknologiat tuovat mukanaan. Tutkielmassa pyritään vastaamaan seuraaviin kysymyksiin:

- Minkälaisia hyötyjä/haasteita datalouhinta ja data-analyysi tuovat yrityksille?

- Miten datalouhintaa ja data-analyysia käytetään verkkokaupoissa?

Vaikka datalouhinnan ja data-analytiikan tekniset näkökohdat ovat olennaisia, tämä tutkimus ei syvenny syvällisesti teknisiin yksityiskohtiin tai työkaluihin, joita käytetään näiden menetelmien toteuttamisessa. Sen sijaan keskitytään enemmän tuloksiin ja vaikutuksiin verkkokauppojen liiketoiminnassa. Vaikka tutkielman toisessa luvussa avataan datalouhintaa ja data-analyysia yleisellä tasolla, niin 3 luvussa keskitytään vain niiden vaikutuksiin yrityksissä painottuen verkkokauppoihin.

1.3 Menetelmät ja rakenne

Tutkimus toteutetaan kirjallisuuskatsauksena. Toisessa luvussa tutkielman kannalta keskeiset käsitteen ja termit määritellään yleisellä tasolla. Datalouhinnan ja data-analyysin alaluvuissa tullaan käymään läpi teknologioiden peruspiirteet ja lyhyet kuvaukset erilaisista tekniikoista.

Tutkielman kolmannessa luvussa keskitytään datalouhinnan ja data-analyysin tuomiin hyötyihin ja haasteisiin. Kolmannen luvun viimeinen osio keskittyy näiden teknologioiden hyödyntämiseen verkkokaupoissa ja miten niiden avulla saadaan kilpailuetua.

2 Datalouhinta ja data-analyysi

2.1 Data-analyysin määritelmä ja prosessi

Data-analyysi on prosessi, jossa arvioidaan tietoja analyyttisen ja loogisen päättelyn avulla tutkimalla jokainen datan osa. Se on tutkimuksen suorittamisen monista vaiheista yksi. Tiedot kerätään eri lähteistä, tarkastellaan ja analysoidaan lopulta jonkinlaisen löydöksen tai johtopäätöksen muodostamiseksi. Data-analyysi voidaan määritellä prosessiksi, jossa dataa puhdistetaan, muunnetaan ja mallinnetaan hyödyllisen tiedon löytämiseksi liiketoimintapäätösten tekemiseksi. Sen tarkoituksena on purkaa hyödyllistä tietoa datasta ja tehdä päätöksiä data-analyysin perusteella. Tämä prosessi ei ainoastaan auta ymmärtämään mitä on tapahtunut tai tapahtumassa, vaan myös ennustamaan tulevia tapahtumia ja ehdottamaan toimenpiteitä ongelmien ratkaisemiseksi tai mahdollisuuksien hyödyntämiseksi. (Islam 2020.)

Data-analyysin prosessi koostuu useista osista, kuten datan keräämisestä, datan käsittelystä, datan puhdistamisesta, datan analysoinnista ja viestinnästä. Työkaluja data-analyysiin on monenlaisia, ja niiden avulla voidaan käsitellä dataa tehokkaasti. Esimerkiksi R, Python, SAS, Apache Spark ja Excel ovat yleisesti käytettyjä työkaluja datan analysointiin. Näiden työkalujen avulla voidaan toteuttaa monimutkaisia analyysejä, jotka tukevat liiketoiminnan päätöksentekoa. (Islam 2020.)

Data-analyysin merkitys liiketoiminnalle on valtava. Se ei ainoastaan auta parantamaan yritysten tehokkuutta ja ymmärrystä markkinoista, vaan myös vähentämään kustannuksia, tekemään nopeampia ja parempia päätöksiä sekä kehittämään uusia tuotteita ja palveluita, jotka vastaavat asiakkaiden muuttuviin tarpeisiin. Data-analyysi mahdollistaa yrityksille mahdollisuuden ymmärtää paremmin toimialaansa, tunnistaa uusia liiketoimintamahdollisuuksia ja pysyä kilpailun kärjessä datavetoisessa maailmassa. (Islam 2020.)

2.2 Data-analyysin tyypit

Data-analyysiin sisältyy monia menetelmiä, mutta niistä tärkeimmät ovat, tilastollinen analyysi, diagnostinen analyysi, ennustava analyysi ja preskriptiivinen analyysi. Näiden menetelmien avulla organisaatiot voivat tehdä tietoon perustuvia päätöksiä, parantaa tehokkuutta, ymmärtää markkinoita paremmin, vähentää kustannuksia, tehdä nopeampia

ja parempia päätöksiä sekä kehittää uusia tuotteita tai palveluita, jotka vastaavat asiakkaiden tarpeita ja toiveita paremmin. (Islam 2020.)

2.2.1 Tilastoanalyysi

Tilastollinen analyysi vastaa kysymykseen ”Mitä tapahtui?” käyttämällä menneitä tietoja koottuna visuaalisiksi esityksiksi. Tilastollinen analyysi sisältää datan keräämisen, analysoinnin, tulkinnan, esittämisen ja mallintamisen. Se analysoi koko data-aineiston tai otoksen datasta. Tilastoanalyysi vielä jakautuu kahteen eri menetelmään, jotka ovat kuvaava analyysi ja päättely analyysi. (Islam 2020, 11.)

Kuvaava analyysi keskittyy datan keskeisten ominaisuuksien, kuten keskiarvon, keskihajonnan, prosenttiosuuksien ja frekvenssien, esittämiseen ja ymmärtämiseen. Sen tavoitteena on tiivistää ja kuvata datan kokonaisuutta yksinkertaisten tilastollisten mittareiden ja graafisten esitysten avulla, tarjoten selkeän kuvan siitä, mitä datasetissä on havaittu. (Islam 2020, 11.)

Päättely analyysissä analysoidaan pieni näyte koko datasta. Tällä analyysi menetelmällä voi löytää eri johtopäätöksiä samasta datasta valitsemalla eri otoksia (Islam 2020, 11).

2.2.2 Diagnostinen analyysi

Diagnostinen analyysi kertoo ”Miksi se tapahtui?” löytämällä syyn tilastollisen analyysin havaintojen perusteella. Tämä analyysi on hyödyllinen käyttäytymismallien tunnistamiseen datasta. Kun uusi ongelma ilmenee, niin optimaalisessa tilanteessa analyytikot tunnistavat siitä samankaltaisia käyttäytymismalleja analyysin avulla, kuin menneissä tapauksissa ja tämän pohjalta hyödyntävät näitä havaintoja uusien ongelmien selvittämiseksi. (Islam 2020, 11.)

2.2.3 Ennustava analyysi

Ennustava analyysi näyttää ”mitä todennäköisesti tapahtuu” käyttämällä aiempaa dataa ja sen tarkkuus perustuu siihen, kuinka yksityiskohtaista tietoa sinulla on ja kuinka syvälle siihen paneudut (Islam 2020). Se voi kattaa ennusteet esimerkiksi tuotemyynnistä, kustannuksista, henkilöstömäärästä, avainmittareista, asiakaspoistumisesta, luottokelpoisuudesta, ristimyynnin ja ylennyksen mahdollisuuksista, markkinointikampanjoiden vastauksista, poikkeamista ja mahdollisista petoksista. Sen

avulla voidaan siis saada parempi kuva liiketoiminnasta ja sen ympärillä tapahtuvista muutoksista (MacGregor 2013).

2.2.4 Preskriptiivinen analyysi

Määrävä analyysi yhdistää aiemmista analyysityypeistä saatujen oivallusten hyödyntämisen, tarjoten suosituksia siitä, miten toimia tietyissä tilanteissa. Se käyttää ennakoivaa analyysiä ja muita analyysimenetelmiä päätöksenteon tueksi, ehdottaen toimenpiteitä, jotka voivat parantaa liiketoiminnan suorituskykyä tai ratkaista esiintyneitä ongelmia. Määrävä analyysi on tärkeä työkalu datavetoisessa päätöksenteossa, koska se tarjoaa perusteltuja suosituksia siitä, miten hyödyntää analyysin tuottamaa tietoa käytännössä. (Islam 2020, 11.)

2.3 Datalouhinta

Datalouhinta, joskus myös tiedonlouhinta, kuuluu tietojenkäsittelytieteen alaluokkaan, ja se voidaan määritellä prosessiksi, jonka tavoitteena on automaattisesti tai osittain automaattisesti etsiä piilotettuja malleja, suhteita ja trendejä suuresta datajoukosta. Sillä onkin kasvava rooli kaupassa ja tieteessä. Datalouhinnan sovellusalueet ovatkin laajat ja sen tyypillisiä tieteellisiä sovelluksia ovat tietojen jälkikäsittely lääketieteessä (esim. CT-tiedot), tietojen arviointi astrofysiikassa (esim. teleskooppi- ja observatoriotiedot), seismisten tietojen ryhmittely tai satelliittikuvien arviointi (esim. NASA:n maan havainnointijärjestelmä). Ehkä vielä tärkeämpiä ovat rahoitus- ja kaupalliset sovellukset. Internetin ja verkkokaupan kehittymisen myötä kerätään valtavia datajoukkoja lähes automaattisesti, joita voidaan käyttää liiketoimintapäätöksiin ja strategiseen suunnitteluun. Siellä sovellukset ulottuvat sopimusten päättämisestä ja peruuttamisesta, luottoriskien arvioinnista asiakkaiden ryhmittelyyn, petosten havaitsemiseen, osakeanalyysiin ja liikevaihdon ennustamiseen. (Garcke ym. 2001, 225–227.)

Datalouhinnan historia ulottuu vain 1970-luvun vaihteeseen. Automaattisen päätöksenteon mahdollisuudet tunnistettiin jo ennen 1970-lukua, mutta silloiset tekoälyyn perustuvat ohjelmistot eivät tarjonneet juurikaan apua varsinaiseen päätöksentekoon vaan keskittyivät ongelmien etsimiseen (Gill 1995). Datalouhinnan historia ulottuu vain 1970-luvun vaihteeseen. Automaattisen päätöksenteon mahdollisuudet tunnistettiin jo ennen 1970-lukua, mutta silloiset tekoälyyn perustuvat ohjelmistot eivät tarjonneet juurikaan apua varsinaiseen päätöksentekoon vaan

keskittyivät ongelmien etsimiseen. Päätöksenteon tukijärjestelmät olivat kuitenkin kehittyneet merkittävästi 1970-luvulle mentäessä, mahdollistaen analyytikoiden ja yritysjohtajien tehokkaan nopeuttamisen datan raportoinnissa, analysoinnissa ja tulkinassa (Gorry & Scott Morton 1971). Kun tietokoneiden tallennuskapasiteetti kasvoi 1980-luvulla, monet yritykset alkoivat tallentaa enemmän tapahtumatietoja. Tuloksena saadut datavarastot, olivat liian suuria analysoitavaksi perinteisillä tilastollisilla lähestymistavoilla. Useita tietojenkäsittelytieteen konferensseja ja työpajoja pohdittiin, kuinka tekoälyn alalla viimeaikaisia edistysaskeleita kuten asiantuntijajärjestelmien löydöksiä, geneettisiä algoritmeja, koneoppimista ja hermoverkkoja voitaisiin mukauttaa tiedon löytämiseen. Prosessi johti vuonna 1995 Montrealissa pidettyyn ensimmäiseen kansainväliseen Knowledge Discovery and Data Mining konferenssiin ja *Data Mining and Knowledge Discovery* lehden julkaisuun vuonna 1997. (Britannica 2024.)

Fayyad ym. (1996) ja hänen kollegansa määrittelevät datalouhinnan algoritmien avulla tapahtuvaksi prosessiksi, jossa etsitään kaavoja datajoukosta. Heidän mukaansa datalouhinta on osa laajempaa prosessia, jonka tarkoituksena on löytää informaatiota datasta, ja siten datalouhinta voidaan sijoittaa KDD-termin (Knowledge Discovery in Databases) alle. Datalouhintaa kuitenkin on kuvailtu eri konteksteissa ja hyvin eri laajuuksilla. Yksi yleisimmistä prosesseista on CRISP-DM, jossa datalouhinta nähdään monivaiheiseksi prosessiksi, samankaltaisesti kuin Fayyad kuvaa KDD-prosessia.

DM ja KD prosessimallien näkökulmasta vuosi 2000 merkitsi tärkeää virstanpylvästä, CRISP-DM (Cross-Industry Standard Process for Data Mining) julkaistiin. CRISP-DM on käytetyin menetelmä datan louhinta- ja tiedonkaivuuprojektien kehittämiseen, ja sitä voidaankin pitää alan standardina. Prosessi on jaettavissa kuuteen päävaiheeseen, jotka ovat liiketoiminnan ymmärrys, dataymmärrys, datan valmistelu, mallintaminen, arviointi ja käyttöönotto. (Chapman ym. 2000.)

Ensimmäisessä vaiheessa, liiketoiminnan ymmärtämisessä, keskitytään liiketoiminnan tavoitteiden ja vaatimusten ymmärtämiseen ja dokumentointiin, tilanteen arviointiin, datanlouhinnan tavoitteiden määrittämiseen liiketoiminnan tavoitteiden pohjalta ja yksityiskohtaisen projekti suunnitelman laatimiseen datanlouhinnan tavoitteiden saavuttamiseksi. Datan ymmärtämisen vaiheessa hankitaan tarvittava data, tutkitaan datan karkeita ominaisuuksia, analysoidaan dataa suhteiden ja alustavien oivallusten paljastamiseksi sekä arvioidaan datan laatua. Datan valmisteluvaiheessa päätetään

analyysiin käytettävästä datasta, parannetaan datan laatua analyysitekniikoiden vaatimusten mukaiseksi, luodaan tarvittaessa uusia datatietueita tai attribuutteja, yhdistetään tietoja useista lähteistä ja suoritetaan tarvittavat syntaktiset muutokset mallinnustyökalun vaatimusten täyttämiseksi. Mallinnusvaiheessa valitaan käytettävä mallinnustekniikka, luodaan testisuunnitelma mallin laadun arvioimiseksi, suoritetaan mallinnustyökalu datasetillä luodakseen yhden tai useamman mallin ja arvioidaan mallit parhaan valitsemiseksi. Arvioinnissa arvioidaan, miten hyvin malli vastaa liiketoiminnan tavoitteita, tarkistetaan projekti tunnistaakseen mahdollisesti huomioimatta jääneet tekijät tai tehtävät ja päätetään projektin tulevaisuudesta arvioinnin perusteella. Käyttöönoton vaiheessa kehitetään strategia mallin viemiseksi osaksi liiketoimintaa, laaditaan suunnitelma jatkuvaa seuranta ja ylläpitoa varten, dokumentoidaan projektin tulokset ja kokemukset loppuraportissa ja arvioidaan projektin onnistumista sekä dokumentoidaan opitut läksyt. Nämä askeleet ovat iteratiivisia ja syklisiä, mikä tarkoittaa, että on tavallista liikkua edestakaisin niiden välillä uusien oivallusten saadessa ja haasteiden ilmetessä. (Chapman ym. 2000.)

2.4 Data-analyysin ja datalouhinnan ero

Vaikka data-analyysi ja datalouhinta jakavat yhteisen tavoitteen – tiedon ja oivallusten löytämisen datan avulla – ne eroavat lähestymistavassaan ja painopistealueissaan. Data-analyysi on laajempi käsite, joka voi sisältää datalouhinnan osana prosessiaan ja keskittyä tiettyjen kysymysten tai hypoteesien tutkimiseen. Datalouhinta puolestaan keskittyy automatisoituun tiedon etsimiseen ja uusien, odottamattomien mallien löytämiseen suurista ja monimutkaisista tietoaaineistoista, usein ilman etukäteen määriteltyä tavoitetta tai kysymystä.

3 Datalouhinnan ja data-analyysin hyödyntäminen

3.1 Datalouhinnan hyödyt ja haasteet

3.1.1 Hyödyt

Data louhinnan tarjoamat hyödyt yrityksille ovat moninaiset ja voivat merkittävästi muuttaa tapaa, jolla yritykset ymmärtävät markkinoita, asiakkaitaan ja omia sisäisiä prosessejaan. Tämän tekniikan avulla yritykset voivat tunnistaa piilotettuja malleja suurista ja monimutkaisista tietokannoista, mikä tarjoaa uudenlaisen näkemyksen päätöksenteon tueksi. (D’Haen ym. 2013.)

Data louhinta mahdollistaa päätöksenteon tehostamisen tarjoamalla syvällistä tietoa yrityksen suorituskyvystä, asiakaskäyttäytymisestä ja markkinatrendeistä. Tämä tieto auttaa yrityksiä tunnistamaan uusia mahdollisuuksia, optimoimaan toimintaansa ja välttämään mahdollisia riskejä. Esimerkiksi, analysoimalla asiakasdataa yritykset voivat kehittää räätälöityjä markkinointikampanjoita, jotka resonoivat kohderyhmänsä kanssa paremmin, parantaen siten asiakaskokemusta ja kasvattaen asiakasuskollisuutta. Data louhinnan avulla yritykset voivat parantaa asiakassuhteitaan tunnistamalla ja ennakoimalla asiakkaiden tarpeita ja käyttäytymistä. Tämä ymmärrys mahdollistaa yrityksille kohdennetumman kommunikaation ja personoidun palvelun tarjoamisen, mikä vahvistaa asiakassuhteita ja edistää pitkäaikaista asiakasuskollisuutta. Lisäksi asiakasdatan analysointi voi paljastaa uusia asiakassegmenttejä tai markkinoita, joita yritys ei ole aiemmin tunnistanut. (Chen et al., 2012; D’Haen et al., 2013)

Riskien hallinta on keskeinen hyöty data louhinnasta. Sen avulla yritykset voivat tunnistaa ja arvioida liiketoimintaansa kohdistuvia riskejä, kuten petoksia, luottotappioita tai toimitusketjun häiriöitä. Analysoimalla historiallista dataa ja tunnistamalla epätavallisia poikkeamia, yritykset voivat ryhtyä ennaltaehkäiseviin toimiin riskien minimoimiseksi. Tämä ei ainoastaan vähennä taloudellisia menetyksiä, vaan myös suojaa yrityksen mainetta. (Hand 1998.)

Lopuksi data louhinta inspiroi innovaatioita ja uusien liiketoimintamahdollisuuksien tunnistamista. Analysoimalla markkinatrendejä, kilpailijoiden toimintaa ja asiakaskäyttäytymistä, yritykset voivat kehittää uusia tuotteita, palveluita ja

liiketoimintamalleja, jotka vastaavat markkinoiden muuttuviin vaatimuksiin. Tämä edistää yrityksen kasvua ja kilpailukykyä pitkällä tähtäimellä. (Chen ym. 2012.)

3.1.2 Haasteet

Data louhinnan käyttöönotto yrityksissä tuo mukanaan paitsi merkittäviä hyötyjä myös useita haasteita ja riskejä, jotka vaativat huolellista harkintaa ja strategista lähestymistapaa. Yksi keskeisimmistä haasteista on tietosuojan ja tietoturvan varmistaminen. Kun yritykset keräävät, tallentavat ja analysoivat suuria määriä henkilökohtaisia ja arkaluonteisia tietoja, on äärimmäisen tärkeää noudattaa tietosuojalakeja ja -määräyksiä sekä varmistaa, että asiakkaiden tiedot ovat turvassa kaikissa tilanteissa. Tämä edellyttää vahvoja tietoturvakäytäntöjä, kuten salauksen ja pääsynvalvonnan käyttöä, sekä jatkuvaa valppautta tietoturvahkien ja haavoittuvuuksien suhteen. (Chen ym. 2012.)

Toinen merkittävä haaste on datan laadun ja eheyden varmistaminen. Epätäydelliset, virheelliset tai vanhentuneet tiedot voivat johtaa harhaanjohtaviin analyysituloksiin ja päätöksiin, jotka saattavat vaikuttaa haitallisesti yrityksen toimintaan. Tämän estämiseksi on tärkeää investoida datan puhdistus- ja validointiprosesseihin sekä kehittää menetelmiä datan jatkuvaa laadunvalvontaa varten. (Bose & Mahapatra 2001; D’Haen ym. 2013.) Lisäksi data louhinnan tehokas hyödyntäminen vaatii sekä teknistä osaamista että resursseja. Kehittyneiden analytiikka- ja koneoppimistyökalujen käyttöönotto, ylläpito ja päivitys edellyttävät merkittäviä investointeja sekä koulutettua henkilöstöä, jolla on tarvittava asiantuntemus datan käsittelystä ja analysoinnista. Tämä voi olla erityisen haasteellista pienille ja keskisuurille yrityksille, joilla on rajalliset resurssit. (Bose & Mahapatra 2001.)

Tulosten tulkinnan monimutkaisuus on myös keskeinen haaste data louhinnassa. Monimutkaiset algoritmit ja mallit tuottavat usein tuloksia, jotka eivät ole suoraan ymmärrettävissä ilman syvällistä analytiikkaosaamista. Tämän vuoksi on tärkeää, että yritykset kehittävät kykyä tulkita ja kommunikoida analyysituloksia selkeästi ja ymmärrettävästi, jotta ne voivat tehdä tietoon perustuvia päätöksiä tehokkaasti. (Chen ym. 2012.)

Lopuksi data louhinnan projekteihin liittyy aina riski, että löydetyt mallit ja säännönmukaisuudet eivät ole toistettavissa tai eivät ole yleistettävissä uusiin

datamääriin. Tämä voi johtua ylikuottamisesta mallin ennustekykyyne tai datan muuttumiseen ajan myötä. Siksi on olennaista suhtautua kriittisesti ja varovaisesti data louhinnan tuloksiin, ja varmistaa, että analyysit perustuvat vankkaan metodologiaan ja kattavaan datamäärään. (Chen ym. 2012.)

Kaiken kaikkiaan data louhinnan hyödyntäminen yrityksissä edellyttää tasapainoilua hyötyjen ja haasteiden välillä, ja se vaatii jatkuvaa sitoutumista datan laadun, tietosuojan, tietoturvan ja osaamisen kehittämiseen.

3.2 Datalouhinta verkkokaupoissa

Data louhinnan hyödyntäminen verkkokaupoissa on monipuolinen ja strategisesti tärkeä osa liiketoiminnan kehittämistä. Se kattaa laajan kirjon menetelmiä ja prosesseja, jotka auttavat parantamaan asiakaskokemusta, tehostamaan markkinointia ja optimoimaan tuotevalikoimaa.

3.2.1 Asiakasprofilointi ja Personointi

Asiakasdatan louhinnan avulla verkkokaupat voivat tunnistaa yksittäisten asiakkaiden tai asiakasryhmien käyttäytymismalleja ja mieltymyksiä. Tämä mahdollistaa henkilökohtaisemman asiakaskokemuksen tarjoamisen räätälöityjen suositusten, kohdennettujen tarjousten ja personoitujen markkinointiviestien muodossa. Asiakasprofiloinnin ja personoinnin hyödyt ilmenevät parempana asiakastyytyväisyytenä, uskollisuutena ja sitoutumisena, mikä johtaa kasvaneeseen myyntiin ja asiakaspidon parantumiseen. (Ismail ym. 2015.)

3.2.2 Ostoskorin Analyysi

Ostoskorin analyysi on voimakas työkalu ymmärtää asiakkaiden ostokäyttäytymistä syvällisemmin. Analysoimalla, mitä tuotteita asiakkaat ostavat yhdessä, verkkokaupat voivat suunnitella tehokkaampia ristiinmyynti- ja lisämyyntistrategioita, optimoida tuotesijoittelua ja kehittää houkuttelevia tuotepaketteja. Tämä ei ainoastaan kasvata myyntiä vaan myös parantaa asiakaskokemusta tarjoamalla asiakkaille relevantteja tuotteita ja tarjouksia. (Ismail ym. 2015.)

3.2.3 Myynnin Ennustaminen

Data louhinnan avulla voidaan ennustaa tulevia myyntitrendejä ja kysynnän muutoksia analysoimalla historiallista myyntidataa, asiakaskäyttäytymistä ja muita markkinoihin vaikuttavia tekijöitä. Tämä auttaa verkkokauppoja suunnittelemaan varastotasot optimaalisesti, optimoimaan hinnoittelustrategioita ja kohdentamaan markkinointitoimiaan tehokkaammin. Ennakoiva näkemys auttaa myös varautumaan paremmin kausivaihteluihin ja trendimuutoksiin, mikä tukee liiketoiminnan kasvua ja kannattavuutta. (Ismail ym. 2015.)

3.2.4 Tuotevalikoiman Suunnittelu

Analytiikan ja data louhinnan avulla verkkokaupat voivat optimoida tuotevalikoimaansa vastaamaan asiakkaiden tarpeita ja toiveita. Tämä käsittää tuotteiden valinnan, hinnoittelun ja varastoinnin strategisen suunnittelun. Data louhinta auttaa tunnistamaan suosituimmat tuotteet, tuoteryhmät, joilla on parhaat myyntimahdollisuudet, ja tuotteet, jotka eivät täytä asiakkaiden odotuksia. Näin ollen verkkokaupat voivat keskittyä resurssiensa ja markkinointiponnistelujaan tehokkaampaan kohdentamiseen. (Ismail ym. 2015.)

3.2.5 Markkinasegmentointi

Data louhinnan avulla voidaan segmentoida asiakaskuntaa demografisten, geografisten ja käyttäytymiseen perustuvien tietojen perusteella. Tämä mahdollistaa kohdennetumman ja tehokkaamman markkinoinnin, sillä viestit, tarjoukset ja kampanjat voidaan räätälöidä vastaamaan kunkin segmentin erityistarpeita ja -toiveita. Markkinasegmentoinnin avulla verkkokaupat voivat myös tunnistaa uusia markkinamahdollisuuksia ja kohderyhmiä, joiden tarpeisiin ei ole vielä vastattu. Tämä auttaa luomaan entistä tarkempia ja tuloksellisempia markkinointistrategioita, jotka lisäävät asiakastyytyväisyyttä, sitoutumista ja lopulta myyntiä. Kun verkkokaupat ymmärtävät paremmin eri segmenttien käyttäytymistä ja odotuksia, ne voivat kehittää segmenttikohtaisia tarjouksia, jotka resonovat kohderyhmän kanssa paremmin, näin ollen lisäten myyntiä. (Ismail ym. 2015.)

3.3 Data-analyysin hyödyt ja haasteet

Data-analytiikan integroiminen yritysten toimintaan on avannut uusia ulottuvuuksia liiketoiminnan kehittämisessä ja strategisessa suunnittelussa. Sen sijaan, että yritykset luottaisivat pelkästään perinteisiin markkinatutkimuksiin tai historiallisiin suorituskykytietoihin, data-analytiikka tarjoaa mahdollisuuden syvälliseen ymmärrykseen reaaliajassa tapahtuvista ilmiöistä, asiakaskäyttäytymisen muutoksista ja markkinoiden dynamiikasta. Yksi merkittävimmistä data-analytiikan sovellusalueista on asiakaskokemuksen räätälöinti. Yritykset keräävät ja analysoivat suuria määriä asiakasdataa, ostohistoriaa, verkkokäyttäytymistä, asiakaspalautteita ja sosiaalisen median interaktioita ymmärtääkseen paremmin asiakkaidensa tarpeita ja toiveita. Tämän tiedon avulla yritykset voivat kehittää kohdennettuja markkinointikampanjoita, räätälöityjä tuote- ja palvelutarjouksia sekä henkilökohtaisia asiakaskokemuksia, jotka lisäävät asiakastyytyväisyyttä ja -sitoutumista. (McAfee & Brynjolfsson 2012.)

Toiminnallisella tasolla data-analytiikka mahdollistaa prosessien tehostamisen ja kustannustehokkuuden parantamisen. Reaaliaikaiset data-analyysit antavat yrityksille arvokasta tietoa toimintojensa suorituskyvystä, paljastaen pullonkaulat, hukkaan menevät resurssit ja mahdollisuudet prosessien optimointiin. (McAfee & Brynjolfsson 2012.) Tämä ei ainoastaan vähennä kustannuksia vaan myös parantaa tuotteen tai palvelun laatua ja toimitusnopeutta. Lisäksi data-analytiikka tukee tuotekehitystä ja innovaatioita. Analysoimalla trendejä, kuluttajien mieltymyksiä ja markkinoiden kehityssuuntia yritykset voivat tunnistaa uusia liiketoimintamahdollisuuksia ja kehittää kilpailuetua tuottavia innovaatioita. Ennakoiva analytiikka, joka käyttää koneoppimista ja tekoälyä, mahdollistaa tulevien trendien ennustamisen ja auttaa yrityksiä olemaan askeleen edellä kilpailijoitaan markkinoilla. (Medeiros ym. 2020.)

Riskienhallinta on toinen alue, jossa data-analytiikka näyttelee keskeistä roolia. Yritykset käyttävät kehittyneitä analytiikkatyökaluja tunnistaa mahdolliset riskit ja epävarmuudet liiketoiminnassaan. Tämä kattaa taloudelliset riskit, toimitusketjun häiriöt, tietoturvariskit ja mainehaitat. Data-analytiikan avulla yritykset voivat toteuttaa proaktiivisia toimenpiteitä riskien minimointiin ja varautua mahdollisiin kriisitilanteisiin paremmin. (Cerchiello & Giudici 2016.)

Kokonaisuudessaan data-analytiikan hyödyntäminen yrityksissä on monipuolinen ja strategisesti merkittävä toimenpide, joka mahdollistaa liiketoiminnan kasvun, tehostaa

toimintaa, parantaa asiakaskokemusta ja vahvistaa kilpailuetua dynaamisesti muuttuvassa liiketoimintaympäristössä. Se edellyttää kuitenkin jatkuvia investointeja datan keräämiseen, analysointikykyihin ja teknologiaan sekä organisaation kulttuurin muokkaamiseen datalähtöisemmäksi. Jotta yritykset voivat täysin hyödyntää data-analytiikan tarjoamia mahdollisuuksia, on välttämätöntä rakentaa osaava tiimi, joka ymmärtää sekä datan että liiketoiminnan dynamiikan. Tämä tiimi ei vain analysoi tietoa, vaan myös tulkitsee sitä tavalla, joka tukee strategista päätöksentekoa ja innovaatiota. Datan keräämiseen ja analysointiin liittyvät investoinnit kattavat laajan kirjon, mukaan lukien edistyneet IT-infrastruktuurit, pilvipalvelut, tietoturvaratkaisut ja analytiikkaohjelmistot. Lisäksi organisaatioiden on sijoitettava koulutukseen ja kehitykseen, jotta työntekijät voivat pysyä ajan tasalla uusimmista analytiikkatyökaluista ja -menetelmistä. (Medeiros ym. 2020; McAfee & Brynjolfsson 2012.)

Organisaation kulttuurin muuttaminen datalähtöisemmäksi ei ole yksinkertainen tehtävä. Se vaatii sitoutumista yrityksen johdolta ja avointa viestintää organisaation sisällä. Johdon on näytettävä esimerkkiä arvostamalla dataa päätöksenteon perustana ja kannustamalla työntekijöitä hyödyntämään data-analytiikkaa omassa työssään. Samalla on tärkeää luoda ympäristö, jossa epäonnistumisia pidetään oppimismahdollisuuksina, rohkaistaan innovointia ja tuetaan luovia ratkaisuja. (Frisk & Bannister 2017.)

Data-analyysin suurimmat haasteet kuitenkin keskittyvät itse dataan. Dataa on yleensä suuria määriä, sitä kertyy nopeasti, erityyppiset ja rakenteiset data-aineistot, datan epävarmuus ja epätarkkuus, hyödyllisen datan tunnistaminen, esittäminen ymmärrettävässä muodossa ja datan jatkuva muutos (Sivarajah ym., 2017).

Lopulta data-analytiikan hyödyntämisen tulee olla jatkuva prosessi, joka mukautuu liiketoiminnan tarpeiden ja markkinoiden muutosten mukaan. Tämä tarkoittaa jatkuvaa datan arvioimista, analysointimenetelmien päivittämistä ja strategioiden säätämistä uusien oivallusten perusteella. Tällainen dynaaminen lähestymistapa varmistaa, että yritys ei vain reagoi markkinamuutoksiin, vaan myös ennustaa tulevaisuuden suuntauksia ja mukautuu niihin proaktiivisesti. (Medeiros ym. 2020.)

Kaiken kaikkiaan data-analytiikan hyödyntäminen yrityksissä tarjoaa merkittäviä etuja, mutta vaatii myös merkittäviä investointeja, jatkuvaa oppimista ja organisaatiokulttuurin muutosta. Kun nämä elementit yhdistyvät ja ne selätetään yritykset voivat saavuttaa kestävän kilpailuedun ja vahvistaa asemaansa markkinoilla.

3.4 Data-analyysi verkkokaupoissa

Data-analytiikan hyödyntäminen verkkokaupoissa avaa monia mahdollisuuksia parantaa liiketoimintaa, tehostaa markkinointia ja tarjota räätälöityä asiakaskokemusta. Käyttämällä data-analytiikkaa, verkkokaupat voivat tunnistaa asiakkaiden käyttäytymismalleja, optimoida verkkosivujen suunnittelua, parantaa asiakaspalvelua ja kasvattaa myyntiä. Yleisimmistä keinoista on muutama esitelty tässä luvussa, miten data-analytiikkaa voidaan hyödyntää verkkokaupoissa sen parantamiseksi.

3.4.1 Ymmärrys asiakaskäyttäytymisestä

Data-analytiikan avulla verkkokaupat voivat kerätä ja analysoida tarkkaa tietoa siitä, miten asiakkaat käyttäytyvät heidän sivustoillaan: mitä sivuja he katsovat, kuinka kauan viiptyvät tietyillä sivuilla, ja millaisia toimintoja suorittavat. Tämä tieto auttaa tunnistamaan suosituimmat tuotteet, tehokkaimmat markkinointikanavat ja asiakkaiden mahdolliset esteet ostoprosessissa. Syvälinen ymmärrys asiakkaan polusta mahdollistaa kohdennettujen toimenpiteiden suunnittelun, jolla voidaan parantaa asiakaskokemusta ja lisätä konversioita eli ostojen määrää per kävijät sivulla. (Zineb ym. 2021; Octoparse.com.)

3.4.2 Henkilökohtaisen asiakaskokemuksen luominen

Hyödyntämällä kerättyä dataa verkkokaupat voivat luoda henkilökohtaisia asiakaskokemuksia, jotka vastaavat yksittäisen asiakkaan mieltymyksiä ja ostohistoriaa. Esimerkiksi, suosittelujärjestelmät voivat tarjota tuotteita, jotka vastaavat asiakkaan aiempia hankintoja tai selailukäyttäytymistä, parantaen näin asiakkaan kokemusta ja lisäten mahdollisuuksia ristiin myyntiin ja uusintamyyntiin. (Zineb ym. 2021; Octoparse.com.)

3.4.3 Markkinoinnin optimointi ja kohdentaminen

Data-analytiikan avulla voidaan seurata eri markkinointikanavien ja kampanjoiden suorituskykyä reaaliajassa. Tämä mahdollistaa markkinointibudjetin kohdentamisen tehokkaimpiin toimenpiteisiin, parantaa mainonnan ROI:ta ja vähentää hukkaan menevää markkinointirahaa. Lisäksi datan avulla voidaan tunnistaa potentiaalisia uusia kohderyhmiä ja kohdentaa mainontaa entistä tarkemmin. (Zineb ym. 2021; Octoparse.com.)

3.4.4 Tuotevalikoiman ja hinnoittelustrategian hienosäätö

Analysoimalla asiakasdataa, verkkokaupat voivat optimoida tuotevalikoimansa vastaamaan kysyntää ja ennustaa tulevia trendejä. Hinnoittelustrategian hienosäätö datan perusteella auttaa määrittämään kilpailukykyiset hinnat, jotka maksimoivat tuottoa samalla kun pitävät asiakkaat tyytyväisinä. Lisäksi varastonhallinnan tehostaminen vähentää ylimääräisen varaston kustannuksia ja parantaa toimitusketjun tehokkuutta. (Zineb ym. 2021; Octoparse.com.)

3.4.5 Verkkosivuston käytettävyyden ja suorituskyvyn parantaminen

Verkkokaupan sivuston suorituskyvyn parantaminen on kriittistä asiakaskokemuksen kannalta. Data-analytiikan avulla voidaan tunnistaa sivuston tekniset ongelmat, kuten pitkät latausajat tai virhesivut, ja ymmärtää, kuinka kävijät liikkuvat sivustolla. Tämä tieto auttaa tekemään tarvittavat muutokset sivuston suunnitteluun, parantaen navigointia, nopeutta ja lopulta konversioasteita. (Zineb ym. 2021; Octoparse.com.)

4 Johtopäätökset

4.1 Hyödyt ja haasteet

Data-analyysin ja datalouhinnan strateginen käyttö verkkokaupoissa tarjoaa useita merkittäviä etuja, jotka voivat parantaa liiketoiminnan suorituskykyä, asiakaskokemusta ja kilpailukykyä markkinoilla. Näiden teknologioiden hyödyntäminen ulottuu monille liiketoiminnan osa-alueille, mukaan lukien markkinointi, asiakassuhdehallinta, varastohallinta ja strateginen päätöksenteko. Yksi datalouhinnan ja data-analyysin keskeisimmistä eduista on kyky kerätä ja analysoida suuria määriä asiakastietoja reaaliajassa. Tämä mahdollistaa syvällisen ymmärryksen asiakkaiden käyttäytymisestä, mieltymyksistä ja tarpeista. Verkkokaupat voivat hyödyntää tätä tietoa personoidakseen asiakaskokemusta, räätälöidäkseen markkinointiviestintää ja kehittääkseen tuotevalikoimaa, mikä parantaa asiakastyytyväisyyttä ja sitoutumista. Kuten Provost ja Fawcett (2013) korostavat, data-analytiikan avulla yritykset voivat tehdä tietoon perustuvia päätöksiä, jotka ovat tarkempia ja tehokkaampia kuin intuitioon perustuvat päätökset. Lopulta, kaikki hyödyt kumuloituvat merkittäväksi kilpailueduksi. Yritykset, jotka pystyvät tehokkaasti hyödyntämään datalouhintaa ja analyysiä, ovat paremmin varustautuneita vastaamaan markkinoiden vaatimuksiin, ymmärtämään asiakkaitaan ja optimoimaan toimintaansa. Tämä johtaa kestävään kilpailuetuun, joka voi erottaa yrityksen kilpailijoistaan dynaamisessa ja nopeasti muuttuvassa digitaalisessa kaupankäynnin ympäristössä.

Data-analyysin ja datalouhinnan hyödyntämisen myötä verkkokaupat kohtaavat useita merkittäviä haasteita, jotka voivat vaikuttaa niiden kykyyn käyttää teknologiaa tehokkaasti. Nämä haasteet voidaan jakaa useisiin keskeisiin kategorioihin, jotka vaativat huomattavia resursseja ja strategista lähestymistapaa niiden ratkaisemiseksi. Suurimpana haasteena, joka mainitaan lähes kaikissa lähteissä, voidaan pitää datan laatua ja eheyttä. Kuten Chen ym. (2012) totesivat, analytiikan tehokkuus riippuu suuresti käytettävän datan laadusta. Verkkokaupat keräävät valtavia määriä dataa, joka voi olla epätäydellistä, virheellistä tai vanhentunutta. Virheellinen data voi johtaa harhaanjohtaviin analyysituloksiin ja päätöksentekoon, mikä saattaa vaikuttaa negatiivisesti liiketoiminnan tuloksiin ja asiakassuhteisiin. Datan laadun varmistaminen edellyttää tehokkaita puhdistus-, validointi- ja hallintamenetelmiä, joiden kehittäminen ja ylläpito vaatii sekä aikaa että teknistä osaamista. Toinen suurimmista haasteista on yritysten tekninen

osaaminen ja resurssit. Tehokkaan datalouhinnan ja data-analyysin toteuttaminen edellyttää osaavaa henkilöstöä, joka ymmärtää sekä algoritmeja että liiketoimintaprosesseja. Pätevien data-analyttikoiden ja tietotekniikan asiantuntijoiden palkkaaminen voi olla haaste, erityisesti pienille ja keskisuurille yrityksille, joilla on rajoitetut resurssit. Lisäksi teknologian nopea kehitys tarkoittaa, että yritysten on sijoitettava jatkuvasti koulutukseen ja uusiin teknologioihin pysyäkseen kilpailukykyisinä.

4.2 Omat näkemykset ja kriittinen arvio

Tutkielman analyysi ja tulokset vahvistavat kirjallisuudessa esitettyjä teorioita, mutta tuovat myös esiin joitakin ristiriitaisuuksia ja haasteita. Esimerkiksi data-analyysin tehokkuus riippuu suuresti datan laadusta ja relevanssista, mikä voi vaihdella suuresti eri verkkokauppojen ja markkinasegmenttien välillä. Mistä voidaan siis tietää, että datan pohjalta tehdyt johtopäätökset ovat hyödyllisiä, jos ei omata resursseja tutkimaan datan laatua ja relevanssia? Ongelma ilmenee monessa lähteessä, joten uskon, että se on yksi alan keskeisistä asioista, johon tullaan tulevaisuudessa keskittymään teknologian kehittyessä.

Kirjallisuudesta saa myös sellaisen käsitteen, että verkkokaupan alalla datan hyödyntäminen on vasta alkutaipaleella. Uskoisin, että tulevaisuudessa voimme odottaa näkevämme entistä edistyneempiä tekniikoita datan hyödyntämiseen, kuten tekoälyn ja koneoppimisen laajamittaisemman käytön. Nämä teknologiat tarjoaisivat mahdollisuuden ei ainoastaan ymmärtää asiakaskäyttäytymistä syvemmin, vaan myös ennustaa tulevia trendejä ja automatisoida monia nykyisiä manuaalisia prosesseja. Uskon näin, sillä monessa lähteessä käsitellään tai ainakin mainitaan tekoäly ja sen luomat mahdollisuudet.

Kirjallisuuteen perehtyessä tulee selvästi ilmi, että vaikka datalouhinta ja data-analyysi tuovat verkkokaupoille lukuisia etuja, niiden käyttö sisältää vielä monia rajoitteita, joita kaikki yritykset eivät onnistu kukistamaan. On välttämätöntä, että alalla toimivat yritykset investoivat paitsi teknologian kehittämiseen myös työntekijöidensä koulutukseen.

5 Yhteenveto

Tutkielma keskittyi tutkimaan datalouhinnan ja data-analyysin rooleja verkkokaupan liiketoiminnan kehittämisessä. Tutkimuksen menetelmänä toimi kirjallisuuskatsaus, jonka avulla analysoitiin olemassa olevaa tutkimusta aiheesta ja tarkasteltiin, kuinka yritykset voivat hyödyntää näitä teknologioita toimintansa tehostamiseksi, asiakaskokemuksen parantamiseksi ja kilpailuedun saavuttamiseksi. Tutkielma alkoi johdatuksella aiheeseen, korostaen datan merkityksen kasvua digitaalisessa ympäristössä ja miten tämä kasvava datamäärä tarjoaa sekä haasteita että mahdollisuuksia yrityksille. Sen jälkeen tutkielmassa määriteltiin keskeiset käsitteet, kuten datalouhinta ja data-analyysi, ja selitettiin niiden prosesseja. Loppu osio tutkielmasta keskittyi erityisesti siihen, miten näitä menetelmiä sovelletaan verkkokaupoissa, millaisia hyötyjä niistä voidaan saada ja mitä haasteita niiden käyttöön liittyy. Tutkielmassa esitettiin myös konkreettisia esimerkkejä siitä, miten datan hyödyntäminen voi auttaa verkkokauppoja ymmärtämään asiakaskäyttäytymistä, optimoimaan markkinointitoimiaan ja kohdentamaan tuotevalikoimaansa tehokkaammin.

Tutkielman tavoitteena oli tarjota kattava tarkastelu siitä, miten datalouhinta ja data-analyysi ovat muuttaneet tapaa, jolla yritykset, erityisesti verkkokaupat, ymmärtävät markkinoita, asiakaskäyttäytymistä ja omia sisäisiä prosessejaan. Tutkimuksen keskeinen tavoite oli selvittää, kuinka nämä teknologiat auttavat yrityksiä tehostamaan toimintaansa ja saavuttamaan kilpailuetua nopeasti digitalisoituvassa ympäristössä.

Tutkielman keskeisempänä löydöksenä voidaan pitää sitä, että datalouhinta ja data-analyysi mahdollistavat yrityksille uudenlaisen näkemyksen liiketoimintaan, joka perustuu suurten datamassojen systemaattiseen tutkimiseen ja analysointiin. Nämä teknologiat mahdollistavat asiakaskäyttäytymisen, markkinatrendien ja liiketoiminnan sisäisten prosessien ymmärtämisen tavalla, joka ei aiemmin ollut mahdollista. Vaikka datalouhinta ja data-analyysi pureutuvat tähän ongelmaan lähtien samoista kysymyksistä, ne kuitenkin löytävät vastaukset eri tavalla ja näin ollen voivat täydentää toisiaan. Erityisesti verkkokaupoille tämä tarkoittaa mahdollisuutta ymmärtää asiakkaita yksilöllisellä tasolla, ennustaa heidän käyttäytymistään ja optimoida tarjontaa vastaamaan asiakkaiden tarpeita ja mieltymyksiä.

Vaikka datalouhinta ja data-analyysi tuovat mukanaan lukuisia etuja, ei voida sivuuta niihin liittyviä haasteita. Tietosuojan, datan laadun varmistamisen ja teknisen osaamisen tarve ovat keskeisiä haasteita, jotka yritysten on ratkaistava hyödyntäessään näitä teknologioita. Lisäksi tutkielmassa korostuu, että vaikka teknologia tarjoaa työkalut datan hyödyntämiseen, yritysten on myös kehitettävä organisaatiokulttuuriaan tukeakseen datavetoista päätöksentekoa.

Tutkielmassa esitellyt tulokset tuovat ilmi useita tärkeitä näkökohtia datan hyödyntämisestä, mutta tutkimukseen liittyy myös rajoitteita, jotka vaikuttavat tulosten yleistettävyyteen ja tulkintaan. Ensinnäkin tutkimus nojautuu kirjallisuuskatsaukseen, joka rajoittaa tutkimuksen empiriaa ja käytännön soveltuvuutta. Vaikka kirjallisuuskatsaus antaa laajan yleiskuvan aiheesta ja sen teoreettisista perusteista, se ei välttämättä kata kaikkia alueellisia eroja tai eri kokoisten verkkokauppojen kokemuksia. Lisäksi datalouhinnan ja data-analyysin nopea kehitys voi johtaa siihen, että tutkielmassa käytetty tieto vanhenee nopeasti, mikä voi vaikuttaa tutkimuksen ajankohtaisuuteen ja sovellettavuuteen eri konteksteissa. Myös lähdemateriaalin vähäinen määrä loi rajoitteita tutkimukselle. Lähteitä löytyy kohtalaisesti, mutta niissä ei hirveästi käsitellä konkreettisia esimerkkejä ja tuloksia, joita yritykset ovat saavuttaneet kyseisillä teknologioilla.

Jatkotutkimusaiheita ajatellen on monia mahdollisuuksia syventää ymmärrystä datalouhinnasta ja data-analyysistä verkkokaupan alalla. Jatkotutkimukset voivat esimerkiksi keskittyä empiirisiin tapaustutkimuksiin, jotka tutkivat konkreettisia sovellusesimerkkejä ja niiden vaikutuksia yritysten suorituskykyyn. Tämä lähestymistapa voisi tarjota syvällisempää tietoa siitä, miten teoria toimii käytännössä ja millaisia erityisiä haasteita ja ratkaisuja liittyy datan hyödyntämiseen tietyissä verkkokauppakohtaisissa tilanteissa. Lisäksi olisi hyödyllistä tutkia, miten pienet ja keskisuuret verkkokaupat voivat ylittää resurssihaasteensa ja soveltaa datalouhinta ja data-analyysiä kilpailuetunsa parantamiseksi. Toinen merkittävä jatkotutkimuksen alue voisi keskittyä datalouhinnan eettisiin ja yksityisyydensuojakysymyksiin, erityisesti kun otetaan huomioon asiakastietojen käsittelyn lisääntyvät sääntelyvaatimukset maailmanlaajuisesti. Näiden aiheiden tutkiminen voisi tarjota kattavampaa tietoa siitä, miten yritykset voivat navigoida monimutkaisessa ja jatkuvasti muuttuvassa datan maailmassa vastuullisesti ja tehokkaasti.

Lähteet

- Bose, I., & Mahapatra, R. K. (2001). Business data mining—a machine learning perspective. *Information & Management*, 39(3), 211–225.
- Cerchiello, P., & Giudici, P. (2016). Big data analysis for financial risk management. *Journal of Big Data*, 3, 1–12.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS Inc*, 9(13), 1–73.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188. <https://doi.org/10.2307/41703503>
- D’Haen, J., Van den Poel, D., & Thorleuchter, D. (2013). Predicting customer profitability during acquisition: Finding the optimal combination of data source and data mining technique. *Expert Systems with Applications*, 40(6), 2007–2012. <https://doi.org/10.1016/j.eswa.2012.10.023>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37.
- Frisk, J. E., & Bannister, F. (2017). Improving the use of analytics and big data by changing the decision-making culture: A design approach. *Management Decision*, 55(10), 2074–2088.
- Garcke, J., Griebel, M., & Thess, M. (2001). Data mining with sparse grids. *Computing*, 67, 225–253.
- Gill, T. G. (1995). Early expert systems: Where are they now? *MIS Quarterly*, 51–81.
- Gorry, G. A., & Scott Morton, M. S. (1971). *A framework for management information systems*.
- Hand, D. J. (1998). Data Mining: Statistics and More? *The American Statistician*, 52(2), 112–118. <https://doi.org/10.2307/2685468>
- Islam, M. (2020). Data analysis: types, process, methods, techniques and tools. *International Journal on Data Science and Technology*, 6(10).
- Ismail, M., Ibrahim, M. M., Sanusi, Z. M., & Nat, M. (2015). Data mining in electronic commerce: benefits and challenges. *International Journal of Communications, Network and System Sciences*, 8(12), 501.

- MacGregor, J. (2013). *Predictive Analysis with SAP*. Bonn: Galileo Press.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60–68.
- Medeiros, M. M. de, Hoppen, N., & Maçada, A. C. G. (2020). Data science for business: benefits, challenges and opportunities. *The Bottom Line*, 33(2), 149–163.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. “ O’Reilly Media, Inc.”
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286.
- Zineb, E. L. F., Najat, R., & Jaafar, A. (2021). An intelligent approach for data analysis and decision making in big data: a case study on e-commerce industry. *International Journal of Advanced Computer Science and Applications*, 12(7).
- Statista.com (2024) Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2023, with forecasts from 2022 to 2030. <https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/>, haettu 27.1.2024
- Statista.com (2024) Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. <https://www.statista.com/statistics/871513/worldwide-data-created/> , haettu 27.1.2024
- Octoparse.com (2022) 6 Benefits of Big Data Analytics for E-Commerce <https://www.octoparse.com/blog/benefits-of-big-data-analytics-for-e-commerce/> , haettu 22.3.2024