

Data-analytiikan hyödyntäminen luottoriskin hallinnassa

Utilizing data-analytics in credit risk management

Kandidaatintyö

TIIVISTELMÄ

Tekijä: Saara Kovanen	
Työn nimi: Data-analytiikan hyödyntäminen luottoriskin hallinnassa	
Vuosi: 2021	Paikka: Lappeenranta
Kandidaatintyö. LUT-yliopisto, Tuotantotalous. 31 sivua, 2 kuvaa ja 5 taulukkoa Tarkastaja: tutkijaopettaja Kalle Elfvengren	
Hakusanat: data-analytiikka, luottoriski, data, luottoriskin hallinta, Z-luku	
Keywords: data analytics, credit risk, data, credit risk management, Z-score	
<p>Tämän kandidaatintyön tavoitteena on selvittää data-analytiikan merkitys luottoriskin hallinnassa. Työssä pyritään selvittämään data-analytiikan käyttö luottoriskin ennustamisessa nykyhetkellä sekä sen hyödyntämismahdollisuudet tulevaisuudessa. Työssä perehdytään ensin sekä data-analytiikkaan että luottoriskeihin erikseen, jolloin avataan muun muassa termeihin läheisesti liittyviä käsitteitä sekä niihin kuuluvia prosesseja. Tämän jälkeen yhdistetään edellisten lukujen aiheet ja keskitytään data-analytiikan käyttöön luottoriskin hallinnassa. Lopussa esitetään myös sovellettu esimerkki luottoriskin ennustamisesta.</p> <p>Työssä saatiin selville data-analytiikalla olevan jo suuri merkitys luottoriskin hallinnassa. Se mahdollistaa tarkkojen ja automatisoitujen mallien luomisen luottotappion ennustamiseen. Jatkuva datan lisääntyminen, kehittyvä teknologia sekä uudet datanlähteet tuovat mahdollisuuksia yhä tehokkaampaan ja tarkempaan luottoriskin hallintaan. Haasteina ovat kuitenkin valtioiden erilaiset rajoitukset luotonmyönnössä sekä analyysien vaatima korkea teknologiaosaaminen.</p>	

SISÄLLYSLUETTELO

1	Johdanto	3
1.1	Työn tausta.....	3
1.2	Työn tavoite, tutkimuskysymykset ja rajaukset	3
1.3	Työnrakenne ja tutkimusmenetelmä	4
2	Data-analytiikka.....	6
2.1	Keskeiset käsitteet.....	6
2.2	Data-analytiikan prosessi	8
2.3	Data-analytiikan ohjelmistotyökalut	10
3	Luottoriski.....	13
3.1	Luottoriskin määritelmä ja keskeiset käsitteet	13
3.2	Luottoriskin tekijät.....	14
3.3	Luottoluokitus	15
4	Data-analytiikan käyttö luottoriskin hallinnassa.....	17
5	Analyysi luottoriskin ennustamisesta	21
6	Johtopäätökset.....	25
	Lähteet	27

1 JOHDANTO

1.1 Työn tausta

Rahan lainaamiseen kuuluu aina tietyn suuruinen riski siitä, että lainattavaa summaa ei makseta sopimuksen mukaisesti takaisin. Yritykset ja pankit pyrkivät erilaisin keinoin minimoimaan tätä riskiä ja hakijan luottokelpoisuuden arvioiminen on siinä avainasemassa. Esimerkiksi asiantuntija-arvioilla ja tilastollisilla menetelmillä pyritään selvittämään kenelle on kannattavaa lainata ja millä hinnalla. Maksuhäiriöiden lisääntyessä tarve yhä tarkemmille ja tehokkaammille menetelmille on syntynyt.

Lainanhakijoista syntyy päivä päivältä yhä enemmän dataa, joka voi antaa merkittävää tietoa hakijan luottokelpoisuudesta. Samaan aikaan teknologian kehittyminen mahdollistaa tehokkaampia keinoja tämän datan muuttamiseen tietoon ja toimintoihin. Yritykset pyrkivät löytämään parhaat mahdolliset keinot - resurssien ja rajoitusten puitteissa - miten hyödyntää näitä uusia menetelmiä luottotappioidensa minimoimiseksi.

1.2 Työn tavoite, tutkimuskysymykset ja rajaukset

Työn tavoitteena on esittää lukijalle, miten data-analytiikkaa käytetään luottoriskin hallinnassa. Tavoitteena on selvittää data-analytiikan käyttöä luottoriskin hallinnassa nykyhetkellä sekä tutkia sen hyödyntämismahdollisuuksia tulevaisuudessa. Tavoitteiden mukaisesti työn päätutkimuskysymykseksi muodostui:

1. Miten data-analytiikkaa voidaan hyödyntää luottoriskin hallinnassa?

Päätutkimuskysymykseen vastaamiseksi tarvitaan tietoa niin data-analytiikasta kuin luottoriskeistä. Seuraavien alatutkimuskysymysten avulla pyritään esittämään tarkempaa tietoa työn eri osa-alueista:

2. Mitä data-analytiikka on ja miten data-analytiikka prosessi etenee?

3. Mitä luottoriskit ovat?

Työssä yhdistyy kirjallisuuskatsaus sekä esimerkinmukainen luottoriskianalyysi valituille yrityksille. Työ rajataan koskemaan vain luottoriskejä eikä muita rahallisia riskejä oteta työssä

huomioon. Luottoriskien hallinnan kaikkia menetelmiä ei esitetä. Esimerkinmukainen prosessi on rajattu käytetyn mallin mukaisesti koskemaan vain yhdysvaltalaisia julkisesti noteerattuja teollisuusyrityksiä.

1.3 Työn rakenne ja tutkimusmenetelmä

Työ koostuu kuudesta luvusta, joista ensimmäinen toimii johdantona. Toinen luku käsittelee data-analytiikkaa ja sen prosessia. Luvussa esitetään muun muassa data-analytiikan ja siihen läheisesti liittyvien termien käsitteet sekä analytiikassa yleisimmin käytetyt ohjelmistotyökalut. Kolmas luku käsittelee luottoriskejä sekä niihin vaikuttavia tekijöitä. Luvussa esitetään myös luottoluokituksen ja -pisteityksen käsitteet sekä yleisimmin käytetyt luottoluokat. Neljännessä luvussa yhdistetään edellisten lukujen aiheet ja käsitellään data-analytiikan käyttöä luottoriskien ennustamisessa. Nämä luvut suoritetaan kirjallisuuskatsauksena erilaiseen aiheeseen liittyvään kirjallisuuteen viitaten. Viidennessä luvussa suoritetaan esimerkinmukainen luottoriskianalyysi 50:lle yhdysvaltalaiselle yritykselle ja esitetään tulokset erilaisten taulukoiden ja kuvaajien avulla. Kuudennessa luvussa esitetään työn johtopäätökset.

Työ toteutetaan kirjallisuuskatsauksena. Pääasiallisesti lähteinä käytetään aiheeseen liittyviä tieteellisiä artikkeleita. Muita työssä käytettäviä lähteitä ovat muun muassa yritysten verkkosivut sekä kirjat. Tieteelliset lähteet ovat haettu erilaisista tietokannoista, kuten LUT-Primosta, Scopuksesta ja Science Directistä. Hakukone Googlestä on haettu ajankohtaista informaatiota muun muassa työssä käsiteltävistä yrityksistä. Hakusanoina on käytetty erilaisia luottoriskeihin ja data-analytiikkaan liittyviä käsitteitä. Lähteet ovat pääasiassa englanninkielisiä laajemman kattavuuden vuoksi, mutta osa lähteistä on myös suomeksi. Hakusanat ja tietokannat on esitetty taulukossa 1.

Taulukko 1. Hakusanat ja -palvelut

Hakusanat	Hakupalvelut
<ul style="list-style-type: none">- Data-analytics- Credit risk (managment)- Big data- Business Intelligence- Credit rating- Altman Z- Z-score- Credit risk analysis- Business Analytics	<ul style="list-style-type: none">- LUT-Primo- Scopus- ScienceDirect- Google Scholar- Google

2 DATA-ANALYTIikka

2.1 Keskeiset käsitteet

Data-analytiikan käyttö on yleistynyt 2000-luvun alkupuolella (Runkler 2016, s. 2), vaikka käsitteenä se on ollut olemassa ainakin 1960-luvulta asti. Tuolloin Tukey (1962, s. 2) määritteli data-analytiikan tilastotieteeseen pohjautuvaksi prosessiksi, jossa data kerätään, prosessoidaan ja tulkitaan. Nykyään data-analytiikka on kansainvälisesti tunnettu käsite, jolle löytyy monia samankaltaisia määritelmiä. Esimerkiksi Kwon et al. (2014) määrittelee data-analytiikan teknologioihin ja tekniikoihin, joita yritys voi käyttää laajan ja monimutkaisen tietoaaineiston analysoimiseen lisätäkseen yrityksen suorituskykyä. Käytännössä nykyaikainen data-analytiikka tarkoittaa päätöksenteon tueksi, suurista tietoaaineistoista tehtävää analyysia erilaisia tietokonejärjestelmiä hyväksi käyttäen. Alana data-analytiikka on monitieteinen ja se on saanut vaikutteita muun muassa tilastotieteestä, koneoppimisesta sekä tekoälystä. (Runkler 2016, s. 2)

Data-analytiikka pyrkii tunnistamaan kaavoja ja yhtäläisyyksiä, jotka voivat edesauttaa organisaatioita. Data-analytiikka on laajasti käytetty esimerkiksi kaupallisissa organisaatioissa, jotka voivat tehdä analyysista saatavan tiedon pohjalta valistuneempia liiketoimintapäätöksiä. Data-analytiikan läheinen termi **liiketoiminta-analytiikka** (Business Analytics) kattaa myös laajasti datan erilaiset hyödyntämismahdollisuudet. Seddon et al. (2017, s. 237) määrittelee liiketoiminta-analytiikan datan käytöksi, jotta voidaan tehdä vakaampia, näyttöön perustuvia liiketoimintapäätöksiä. Liiketoiminta-analytiikkaa käytetään organisaatioiden datan analysoimisessa ja se pyrkii selittämään miksi jotain tapahtuu. Niin kuin data-analytiikan, myös liiketoiminta-analytiikan tarkoitus on auttaa organisaatioiden päätöstentekoa. (Goodwyn 2019, s. 33)

Business Intelligence eli **liiketoimintatiedon hallinta** on toinen data-analytiikkaan läheisesti liittyvä termi. Liiketoimintatiedon hallinta on nykypäivänä laajasti käytetty käsite kirjallisuudessa, mutta sille ei silti ole yhtä yleistä määritelmää (Mach-Krol et al. 2015, s. 90). Saxena ja Srinivasan (2013, s. 85) määrittelee liiketoimintatiedon hallinnan työkalujen ja tekniikoiden kokoelmaksi datan hallintaa, analyysia ja päätöksenteon tukemista varten. Rausch et al. (2013, s. 4) puolestaan määrittelee liiketoimintatiedon hallinnan integroiduksi, yrityskohtaiseksi, IT-pohjaiseksi lähestymistavaksi johdon päätöksenteon tukea varten. Käytännössä liiketoimintatiedon hallinta käsittää datan visualisoinnin ja raportoinnin, ja siihen

liittyvän kaavioiden, taulukoiden ja ohjausnäkyvien tekemisen, joiden tarkoituksena on pyrkiä selittämään mitä on tapahtunut ja mitä tulee tapahtumaan (Shmueli et al. 2016, s. 3). Liiketoimintatiedon hallintaa pidetään usein 'sateenvarjoterminä', joka kattaa ohjelmistotuotejoukkojen käytön informaation keräämiseen, kokoamiseen ja analysoimiseen organisaatioiden tehokkaamman päätöksenteon tueksi (Adelman & Moss 2000). Liiketoimintatiedon hallinnan ja liiketoiminta-analytiikan määritelmät ovat hyvin samankaltaiset ja usein alan harjoittajatkään eivät näe merkittävää eroa niiden välillä (Kowalczyk 2017, s. 5). Chen et al. (2012, s. 1166) ehdottaa Davenportin määritelmän mukaisesti, että termi liiketoiminta-analytiikka edustaa keskeistä analyttistä osaa liiketoimintatiedon hallinnassa. Termejä käytetään myös usein yhdistelmänä BI&A (Business Intelligence & Analytics), joka käsittää laajemmin tekniset valmiudet datan keräämiseen, integrointiin ja analysointiin tiedon laadun parantamiseksi päätöksenteon tukea varten (Kowalczyk 2017, s. 1).

Datan jatkuva lisääntyminen on saanut yritykset kiinnostumaan niin sanotusta **big datasta** (Schuster 2017, s. 167). Big data termin suuresta suosiosta huolimatta, sille ei silti ole yhtä yleispätevää määritelmää (Buyya et al. 2016, s. 1). Vaikka big data mielletäänkin usein vain valtavaksi määräksi dataa, kuuluu siihen koon lisäksi myös muita ominaisuuksia. Yksi suosittu big datan tulkinta on Douglas Laney'n kehittämä 3V:n malli. Siinä big data on määritelty kolmen ominaispiirteen (3V:n) avulla, joita ovat määrä (volume), monimuotoisuus (variety) sekä nopeus (velocity) (Gandomi & Haider 2015, s. 138). Sittemmin 3V mallia on laajennettu ja määritelmään on lisätty uusia ominaispiirteitä. Esimerkiksi Microsoft on lisännyt malliin kolme uutta ominaisuutta: vaihtelevuuden (variability), epävarmuuden (veracity) sekä näkyvyyden (visibility). (Buyya et al. 2016, s. 10) Eri ominaisuudet on kuvattu taulukossa 2.

Taulukko 2. 6V:n ominaispiirteet (Gandomi & Haider 2015, s. 138; Buyya et al. 2016, s. 10)

Ominaispiirre	Merkitys
Määrä (Volume)	Viittaa datan suuruuteen. Koolle ei ole tiettyä pistettä, jolloin voidaan puhua big datasta, sillä datan suuruus on suhteellista ja se vaihtelee ajan, datatyypin ja alan mukaan.
Monimuotoisuus (Variety)	Viittaa datan erilaisiin muotoihin. Data voi olla strukturoitua, osittain strukturoitua tai strukturoimatonta.
Nopeus (Velocity)	Tarkoittaa millä nopeudella uutta dataa syntyy, ja millä nopeudella sitä tulisi analysoida. Uuden datan nopea syntyminen voi aiheuttaa haasteita aloilla, jossa tarvitaan reaaliaikaista datan analysointia.
Vaihtelevuus (Variability)	Viittaa datavirran nopeuden vaihteluun. Usein big datan nopeus (velocity) ei ole johdonmukaista, vaan siinä on ajoittaisia huippuja ja syvennyksiä.
Epävarmuus (Veracity)	Viittaa datan luotettavuuteen. Epäluotettavia tietoaineistoja täytyy käsitellä ja louhia, jotta datasta saadaan oikeanmukaista informaatiota.
Näkyvyys (Visibility)	Datasta täytyy hahmottaa kokonaiskuva, jotta siitä voi tehdä johtopäätöksiä.

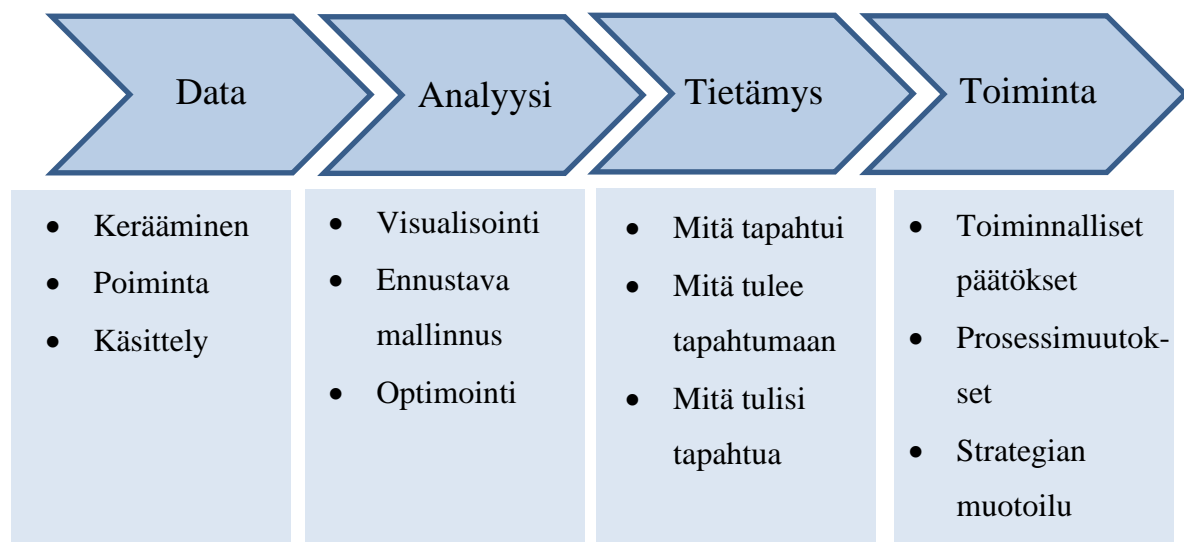
2.2 Data-analytiikan prosessi

Kerätty raakadata on usein sellaisenaan hyvin käyttökeltvotonta. Jotta datalla olisi jotain arvoa, täytyy se jalostaa ihmiselle ymmärrettävään muotoon. Data-informaatio-tieto-viisus-hierarkia (eng. Data to Wisdom (DIKW) -hierarchy) pyrkii kuvaamaan tiedon jalostusastetta. DIKW on yhdysvaltalaisen teoretikon Russel Ackoffin vuonna 1989 esittämä hierarkia, josta on sittemmin esitetty variaatiota ja sen tasoille on tehty erilaisia määrittelyjä. DIKW-hierarkian tasot ovat seuraavat (Martinsuo & Kärri 2017, s. 19-21):

- **Data** – Tarkasteltavasta kohteesta kerättyä aineistoa, joka voi olla sekä numeerista että ei-numeerista. Päätöksenteon näkökulmasta raakadata ei ole sellaisenaan kovin käyttökelpoista.
- **Informaatio (Information)** – Ihmisaivoille ymmärrettävään muotoon jalostettua dataa. Esimerkiksi tunnusluvut, kuvaajat ja keskiarvot.

- **Tietämys (Knowledge)** – Kyky tulkita informaatiota ja tunnistaa mahdollisesti tarvittavat toimenpiteet.
- **Viisaus (Wisdom)** – Kyky yhdistää tietoa eri lähteistä sekä tunnistaa vaihtoehtoiset toimintatavat.

DIKW-hierarkia kuvaa datan jalostusastetta, mutta ei ota kantaa tiedon jalostamisen prosessiin, jolla tiettyyn jalostusasteeseen päästään (Martinsuo & Kärri 2017, s. 19-21). Datan muuntaminen tietämykseen ja toimintoihin on analyysin ja havaintojen kautta syntyvä prosessi, jolla pyritään vaikuttamaan organisaation päätöksentekoon ja ongelmien ratkaisuun (Liberatore ja Luo 2010, s. 314). Data-analytiikan prosessi on havainnollistettu kuvassa 1.



Kuva 1. Analytiikan prosessi (Liberatore & Luo 2010, s. 314)

Analytiikan prosessi alkaa raakadatan keräämisellä, poiminnalla ja käsittelyllä. Koska kerätty data on tallennettu erilaisia tarkoituksia varten, on se usein myös vaikeasti analysoitavassa muodossa. Tämän vuoksi olennaisen tiedon poimiminen ja käsittely on välttämätöntä, jotta oikeanlainen analyysi pystytään suorittamaan. (Liberatore & Luo 2010, s. 314) Prosessia, jossa tavoitteena on poimia olennainen informaatio datasta, kutsutaan datan louhinnaksi (Runkler 2016, s. 2). Datan keräys ja käsittely on usein kriittisin ja aikaa vievin osuus analytiikan prosessissa. Datan louhinnan ja käsittelyn jälkeen prosessissa siirrytään analyysivaiheeseen. Analyysivaiheessa erilaisia tekniikoita käytetään datan tutkimiseen ja arvioimiseen. Vaihe

jaetaan usein kolmeen eri kategoriaan: visualisointiin, ennustavaan mallintamiseen sekä optimointiin. Visualisointivaiheessa data analysoidaan ja esitetään erilaisten kuvaajien, taulukoiden ja ohjausnäkymien (dashboards) avulla. Ennustavia malleja käytetään datan luokitteluun sekä erilaisten suhteiden ja yhtäläisyyksien hahmottamiseen. Ennustavan mallinnuksen tekniikoita ovat muun muassa lineaarinen regressio, klusterianalyysi sekä neuroverkot. Analyysin viimeisessä vaiheessa pyritään löytämään optimaalinen vaihtoehto erilaisten mallien avulla. Esimerkkejä optimointimalleista ovat muun muassa erilaiset matemaattiset ohjelmointimallit sekä tekoälyn heuristiikka, kuten geneettiset algoritmit. Nämä kaikki analyysin vaiheet antavat päätöksentekijöille erilaista informaatiota. Visualisointi vastaa etenkin kysymykseen mitä on jo tapahtunut, kun taas ennustava mallintaminen indikoi mitä tulevaisuudessa tulee tapahtumaan. Optimointivaihe puolestaan pyrkii näyttämään suunnan, mitä tiettyjen olosuhteiden puitteissa tulisi tapahtua. Viimeisessä vaiheessa datasta saatu informaatio muutetaan toiminnaksi. Tällöin havaintojen pohjalta voidaan tehdä valistuneempia päätöksiä esimerkiksi yrityksen strategiaan tai prosesseihin liittyen. (Liberatore & Luo 2010, s. 314-315)

2.3 Data-analytiikan ohjelmistotyökalut

Jotta yritykset voivat onnistua raakadatan jalostamisessa tietämykseen ja toimintaan, tarvitsevat ne sitä varten oikeanlaiset työkalut, jotka mahdollistavat data-analytiikan prosessin eri vaiheet. Analyysityökaluilla tarkoitetaan ohjelmia, joilla voi suorittaa tilastollisia ja kvantitatiivisia analyysejä, selittäviä ja ennustavia malleja, sekä datan laajaa hyväksikäyttöä. Ne mahdollistavat toiminnan, päätöksenteon ja johtamisen perustumisen tosiasioihin. (Davenport & Harris 2007, s. 26) Analytiikan työkalut voivat erikoistua analytiikan eri osa-alueisiin, ja niitä voi olla ilmaisia, avoimen lähdekoodin ohjelmistoja sekä kaupallisia, maksullisia työkaluja.

R ja Python ovat molemmat suosittuja avoimen lähdekoodin ohjelmointikieliä. Python on yksi suosituimmista ohjelmointikielistä maailmanlaajuisesti, ja sen kaikki versiot ovat ladattavissa ilmaiseksi macOS-, Linux- ja Windows-käyttöjärjestelmille. (Romano et al. 2016, s. 7) Python ei ole erikoistunut data-analytiikkaan, vaan sitä käytetään moniin eri käyttötarkoituksiin, kuten yleiseen ohjelmointiin, verkkokehitykseen sekä komentosarjojen tekemiseen. Siitä huolimatta se on laajasti käytetty data-analytiikan työkalu, ja sen avulla on mahdollista esimerkiksi käsitellä, louhia ja prosessoida dataa. Python sisältää data-analyysia varten tehtyjä kirjastoja,

kuten kirjaston 'pandas', joka on erikoistunut data-analyysiin ja kirjaston 'matplotlib', joka on tehty datan visualisointia varten. (Nelli 2018, s. 17, 87-88, 231) Kuten Python, myös R-ohjelmointikieli on ladattavissa Windows-, Linux- ja macOS-käyttöjärjestelmiin. R on tilastotieteilijöiden kehittämä ohjelmointikieli, joka sopii erityisesti tilastolliseen analyysiin. (Pathak 2014, s. 2) R-ohjelmointikieltä pidetään yleisesti parempana datan visualisoinnin ja kuvaajien teon kannalta. Sillä on myös suurempi valikoima erilaisia paketteja ja toiminnallisuuksia data-analyysia varten. Pythonin etuja R:n nähden ovat puolestaan sen monipuoliset käyttömahdollisuudet data-analytiikan lisäksi.

Microsoftin työkalut, kuten Excel ja Power BI ovat hyvin suosittuja monessa yrityksessä. Excel on kansainvälisesti käytetyin laskentataulukko ja se kuuluu Microsoft Office pakettiin. Exceliä voidaan käyttää muun muassa numeerisiin laskelmiin, taulukoiden ja kaavioiden luomiseen, tekstin käsittelyyn sekä tehtävien automatisaatioon. Muita ominaisuuksia on esimerkiksi Power Query, 3D kartta ja ennustemallit. (Walkenbach 2016, s. 3) Power BI -työkalua käytetään etenkin raporttien ja ohjausnäkymien tekoon. Sen avulla voi yhdistää dataa eri lähteistä, analysoida dataa ja luoda siitä visuaalisia, vuorovaikutteisia kaavioita. (Clark 2020, s. 1-2) Verrattuna Exceliin Power BI on tehokkaampi, ja se kykenee käsittelemään suurempaa määrää tietoa lyhyemmässä ajassa. Ero näiden kahden työkalun välillä liittyy myös niiden käyttötarkoitukseen: Power BI on business intelligence työkalu, joka on suunniteltu erityisesti datan visualisointiin. Excelin yleisimmät käyttökohteet ovat puolestaan datan analysointi ja organisointi sekä matemaattiset laskut. (Morrison 2020)

Muita kaupallisia analytiikan työkaluja ovat muun muassa Tableau, Qlik Sense, SAS ja MATLAB. Tableau on business intelligence työkalu, joka on keskittynyt erityisesti datan visualisointiratkaisuihin (Tableau n.d.). SAS, MATLAB sekä Qlik Sense tarjoavat puolestaan laajemmin erilaisia ohjelmistoja. SAS:lla on esimerkiksi markkinointianalytiikkaan, riskienhallintaan, edistyskelliseen analytiikkaan sekä tekoälyyn ja koneoppimiseen erikoistuneita ohjelmistoja (SAS 2021). MATLAB on ohjelmoinnin ja numeerisen laskennan alusta, jota käytetään esimerkiksi koneoppimiseen ja robotiikkaan (MATLAB n.d.a). MATLAB:lla on data-analytiikkaan soveltuva ominaisuus, jolla pystyy esimerkiksi tutkimaan ja mallintamaan dataa, sekä luomaan visuaalisia ja interaktiivisia kaavioita (MATLAB n.d.b). Qlik Sense on dataintegraation ja -analytiikan ohjelmistoja tarjoavan Qlik:n työkalu. Qlik Sense

on suosittu visuaalisen analytiikan ja liiketoimintatiedon hallinnan alusta, mutta se tarjoaa myös muita palveluja keskustelevasta analytiikasta datan tutkimiseen. (Qlik n.d.)

3 LUOTTORISKI

3.1 Luottoriskin määritelmä ja keskeiset käsitteet

Luottoriski tarkoittaa riskiä siitä, että lainanottaja ei täytä taloudellisia velvollisuuksiaan lainan takaisinmaksussa eli laiminlyö sopimuksen mukaisen maksuvelvoitteen. Maksuvelvoitteen laiminlyönti määritellään maksukyvyttömyytenä (default). (Neal 1996) Suomessa luottolaitosten luottoriskiä valvoo Finanssivalvonta. Finanssivalvonta määrittelee luottoriskin seuraavasti: ”Luottoriskillä tarkoitetaan riskiä siitä, että rahoitusinstrumentin sopimusosapuoli ei pysty täyttämään velvoitteitaan ja aiheuttaa siten toiselle osapuolelle taloudellisen tappion. Luotot ovat suurin luottoriskin lähde, mutta luottoriskiä (vastapuoliriski) voi syntyä myös joukkovelkakirjalainoista, lyhytaikaisista saamistodistuksista ja johdannaissopimuksista sekä taseen ulkopuolisista sitoumuksista, kuten käyttämättömistä luottojärjestelyistä ja limiiteistä, takauksista ja rembursseista. Myös maa- ja selvitysriskit luetaan luottoriskeihin.” (Finanssivalvonta 2018) Luottoriski voidaan määritellä kolmen ominaisuuden avulla:

1. Probability of default, PD ~ Kuvaa maksukyvyttömyyden todennäköisyyttä. Yleisesti käytetty kriteeri, jolla maksukyvyttömyyden katsotaan tapahtuneen, on kun velallisen takaisinmaksu on yli 90 päivää myöhässä (Harris 2013, s. 4404). PD arvioidaan yleensä analyttisten luottoluokittelumallien avulla.
2. Exposure at default, EAD ~ Kuvaa riskin suuruutta (takaisin maksettavan pääoman määrää) maksukyvyttömyyden hetkellä.
3. Loss given default, LGD ~ Arvio luottotappion määrästä prosentteissa. (Doumpos et al. 2019, s. 7)

Luottoriskit ovat merkittävä riski rahoitussektorilla, kuten pankeissa, jotka tarjoavat asiakkailleen erilaisia yritys- ja kulutusluottoja, mutta ovat olennainen riski myös rahoitussektorin ulkopuolella. Esimerkiksi yrityksille syntyy päivittäisissä toiminnoissaan luottoriskejä, kun he saavat lainaa toimittajiltaan ja tarjoavat sitä asiakkailleen. (Doumpos et al. 2018, s. 1) Perintäpalvelukonserni Intrum kertoo eurooppalaisessa maksutaparaportissaan suomalaisten yritysten luottotappion olevan keskimäärin prosenttiliikevaihdosta vuonna 2019. Vastaavasti Euroopassa keskimääräinen luottotappio oli 2,3 % yritysten liikevaihdosta. Luottoriski on pankkitoiminnan merkittävin rahallinen riski (Neal 1996).

3.2 Luottoriskin tekijät

Luottoriskin määrittelevien tekijöiden tunnistaminen on tärkeää taloudellisen vakauden kannalta. Selkeämpi käsitys luottoriskiin vaikuttavista tekijöistä voi auttaa yritysten maksukyvykkyyden ennustamisessa. (Bonfim 2009, s. 281) Luottoriskiin vaikuttavat tekijät jaetaan systemaattisiin eli yleisiin, sekä omaperäisiin eli lainaajakohtaisiin tekijöihin. (Memmel et al. 2015, s. 232)

Bonfim (2009, s. 298) suoritti tutkimuksen yli 30 000:n yrityksen taloustietojen pohjalta, joka osoitti usean yrityskohtaisen tekijän vaikuttavan maksukyvyttömyyden todennäköisyyteen. Esimerkiksi yrityksen taloudellisella rakenteella, kannattavuudella, likviditeetillä sekä lähiaikaisella myyntituloksella todettiin olevan merkittävä vaikutus maksukyvykkyyteen. Yrityksen aikaisemmat maksuvaikeudet ovat myös olennaista ottaa huomioon riskin arvioinnissa: tutkimuksessa yrityksillä, joilla oli ilmennyt maksuvaikeuksia lähihistoriassa oli myös huomattavasti suurempi todennäköisyys uuteen maksukyvyttömyyteen. Maksuhistorian on todettu olevan merkittävin yksittäinen tekijä maksukyvyttömyyden todennäköisyyden ennustamisessa (Lawrence et al. 1992, s. 299). Henkilökohtaisten ominaisuuksien, kuten lainanottajan iän, sukupuolen sekä nykyisen työsuhteen pituudella on todettu olevan merkitys pankkilainan luottoriskiin: todennäköisimmin maksuvaikeuksia on todettu olevan miehillä, nuorilla sekä lyhyessä työsuhteessa olevilla, kuin naisilla, vanhemmilla tai pidemmässä työsuhteessa olevilla. Myös asuinalueen, kotitalouden tulojen sekä asumismuodon on todettu vaikuttavan maksukyvyttömyyden todennäköisyyteen. (Azam et al. 2012)

Memmel et al. (2015, s. 244) tekemässä analyysissä havaitaan systemaattisten tekijöiden, kuten valtakunnallisen luottotappioasteen, toimialan kokoonpanon, maturiteettirakenteen sekä alueellisten tekijöiden vaikuttavan huomattavasti luottotappioriskiin. Makroekonomiset tekijät ovat myös merkittävä luottoriskiin vaikuttava systemaattinen tekijä. Esimerkiksi tuotantokuilu, tuottokäyrä sekä kotitalouksien odotukset taloudesta ovat tärkeitä indikaattoreita luottoriskin kehityksessä (Carling et al. 2007, s. 865). Bonfimin (2009, s. 298) tutkimus osoittaa myös makroekonomisten tekijöiden suuren merkityksen yritysten maksukyvykkyyteen. Tutkimuksessa todettiin, että voimakkaan talouskasvun jaksoilla yrityksillä saattaa olla taipumusta liialliseen riskinottoon, josta seuraa toisinaan voimakasta luottojen kasvua. Tällaisina aikoina syntynyt epätasapaino ilmenee kuitenkin vasta talouskasvun hidastuttua.

Castro (2013, s. 681) päätyi samaan tulokseen makroekonomisten tekijöiden merkittävästä vaikutuksesta luottoriskin suuruuteen hänen teettämässään analyysissä EU alueella. Hän huomauttaa BKT:n kasvun hidastumisen, sekä osakekurssi-indeksien ja asuntojen hintojen laskun vaikuttuvan luottoriskin kasvuun. Myös työttömyysasteen sekä korkojen ja luottojen kasvulla todettiin olevan luottoriskiä lisäävä vaikutus.

3.3 Luottoluokitus

Luottoriskin merkittävyyden vuoksi sen hallinnasta on tullut avainkysymys luottoa tarjoaville yrityksille ja rahoituslaitoksille. Luottoriskin hallintaan käytetään hyvin erilaisia tekniikoita ja lähestymistapoja käyttötarkoituksesta riippuen. (van Thiel & van Raaij 2019, s. 270) Luottoluokittelu on yksi yleisimmistä menetelmistä luottoriskin arviointiin. (Harris 2013, s. 4404)

Luottoluokitus ja -pisteytys menetelmät (credit rating and scoring) viittaavat prosesseihin, joilla määritellään hakijan luottoriskin taso. Menetelmiä käytetään luottoriskin arvioinnin välineenä, jossa pääajatuksena on luokitella lainanhakijat 'hyviin' ja 'huonoihin' ryhmiin maksukyvyttömyyden todennäköisyyden perusteella. Menetelmissä otetaan huomioon kaikki käytettävissä oleva informaatio, joka liittyy lainanhakijan kykyyn tai halukkuuteen maksaa laina takaisin sovittujen ehtojen mukaisesti. Luottoluokitus ja -pisteytys menetelmiä käytetään päätöksenteon tukena luottopäätöstä tehdessä sekä luoton hinnoittelun apuna. Korkean riskitason lainaajat joutuvat maksamaan suuremmat lainan korkomaksut kuin vähäriskisiksi todetut yritykset tai henkilöasiakkaat. Perinteisesti luokitteluun on käytetty tilastollisia malleja, kuten syrjivää analyysia tai lineaarista regressiota, mutta lähivuosina uudet menetelmät ovat yleistyneet, joita käsitellään seuraavassa luvussa. Luokitus- ja pisteytysmalleja voidaan kehittää sisäisesti tai ostaa ulkoisilta toimittajilta, kuten luottoluokituslaitoksilta. Nykyään kolme johtavaa luottoluokituslaitosta, Moody's, Standard and Poor's (S&P) ja Fitch hallitsevat globaaleja markkinoita. Ne luokittelevat yritys- ja valtionvelkoja käyttämällä usean luokan järjestelmää, jotka vaihtelevat matalan luottoriskin tapauksista maksukyvyttömyyden tapauksiin. Luokitukset voidaan jakaa kahteen suurempaan osioon: matalan riskin investointiluokkiin ja korkean riskin spekulatiivisiin luokkiin. (Doumpos et al. 2018, s. 23-38; Harris 2013, s. 4404-4405; Getter 2006, s. 41) Tarkemmat kuvaukset luokitteluista on kuvattu taulukossa 3.

Taulukko 3. Luottoluokitukset (Doumpos et al. 2018, s. 38)

Kategoria	Moody's	S&P/Fitch	Kuvaus
Investointi- luokitus	Aaa	AAA	Korkea suositus. Pieni maksukyvyttömyyden todennäköisyys. Vakaa sijoitus.
	Aa	AA	Korkealaatuinen sijoitus. Hieman korkeamman tason pitkäaikainen riski.
	A	A	Korkea - keskiverto sijoitus. Haavoittuvainen taloudellisille muutoksille.
	Baa	BBB	Keskivertolaatuinen. Melko turvallinen nykyhetkellä, riski pitkällä aikavälillä.
Spekulatiivinen luokitus	Ba	BB	Keskivertolaatuinen. Ei hyvin turvattu sijoitus.
	B	B	Melko turvallinen nykyhetkellä. Maksukyvyttömyys todennäköinen tulevaisuudessa.
	Caa	CCC	Heikkolaatuinen. Korkea todennäköisyys maksukyvyttömyyteen.
	Ca	CC	Spekulatiivinen sijoitus. Lähellä maksukyvyttömyyttä.
		C	Heikkolaatuinen. Korkea todennäköisyys maksukyvyttömyyteen. Maksaa nykyhetkellä.
	C	D	Maksukyvyttömyys.

4 DATA-ANALYTIIKAN KÄYTTÖ LUOTTORISKIN HALLINNASSA

Vanhimmat lähestymistavat luottoriskin hallintaan perustuivat laadulliseen eli subjektiiviseen arviointiin. Siinä asiantuntija arvioi lainanhakijan ominaisuuksia, kapasiteettia, pääomaa, vakuuksia ja olosuhteita. Henkilöasiakkaista arvioitiin muun muassa hakijan persoonallisuus, tulot, menot ja muut mahdolliset velat sekä lainan ominaisuudet, kuten korkotasot. Yritysten arvioinnissa puolestaan huomioon otettiin julkiset taloudelliset tiedot, kuten tuloslaskelma, tase ja kassavirta sekä informaatio markkinoista ja yleisestä taloudellisesta tilasta. Pienemmissä lainoissa lainapäätöksen teki usein yksi asiantuntija, kun taas suuremmissa lainahakemuksissa päätöksessä otettiin usein huomioon sekä asiantuntijoiden että luottokomitean arviointi. Tällaisessa lähestymistavassa ongelmana on objektiivisen algoritmin puuttuminen sekä korkeat asiantuntijoiden kustannukset. (Doumpos et al. 2018, s. 10-11; Adamko et al. 2014, s. 148)

Datan runsaan lisääntymisen myötä luottoriskin arviointi on siirtynyt asiantuntijan arvioinneista dataohjautuviin lähestymistapoihin, jotka mahdollistavat tehokkaamman ja tarkemman luottoriskin arvioinnin sekä asiakkaiden objektiivisen luottoluokittelun. Dataohjautuvat lähestymistavat tarkoittavat malleja, jotka ovat rakennettu historiallisen datan perusteella hyväksytyistä, hylätyistä ja ajallaan maksetuista lainoista sekä tapauksista, joissa on ilmennyt lainan maksukyvyttömyys. Mallien rakentamiseen käytetään dataa lainanhakijan ominaisuuksista, ulkoisesta ympäristöstä sekä lainan tilasta (erottamalla ajallaan maksetut ja maksamattomat lainat). Näin mallit tunnistavat kuvioita datasta ja muodostavat suhteita maksukyvyttömyyden todennäköisyyden ja syötetietojen/riskitekijöiden välillä. (Doumpos et al. 2018, s. 12) Perinteisesti mallien syötetietona käytetään sosiaalidemograafista dataa, dataa hakijan taloudellisesta tilanteesta, työllisyys- ja koulutustietoja sekä käyttäytymiseen liittyvää dataa. Osa valtioista, kuten Yhdysvallat, on kieltänyt joidenkin sosiaalidemograafisten tietojen käytön luottoluokittelussa, kuten tiedon hakijan sukupuolesta, iästä, etnisestä taustasta ja uskonnosta. Joissain tutkimuksissa syötetietona käytetään myös makroekonomista dataa, jolloin hakemuksen tekohetken markkinaolosuhteet otetaan luottopäätöksessä huomioon. (Toback & Martens 2019, s. 1228) Myös datan louhinnan tekniikat, kuten päätöspuut, logistinen regressio, erotteluanalyysi ja luokituspuut ovat yleisiä datalähtöisiä tekniikoita luottoriskin hallinnassa. Datan louhinta pyrkii tunnistamaan datasta suhteita, kuvioita ja korrelaatioita käyttäen hyväksi mallinnusteknologioita tai tilastollisia ja matemaattisia

tekniikoita. (Moradi & Mokhatab Rafiei 2019, s. 3-4; Tobback & Martens 2019, s. 1228; Mandala et al. 2012, s. 408-409)

Tekoälyn ja big datan kehittyminen on mahdollistanut monipuolisemman datan käytön sekä tarkemmat algoritmit luottoriskin hallintaan. Niin kuin edellä mainittiin, perinteisesti luottoluokitukseen on käytetty asiakkaan historiallisia taloudellisia tietoja sekä sosiaalidemograafista ja makroekonomista dataa. Big datan kehittymisen myötä uusia datanlähteitä on alettu ottamaan huomioon luottoriskin arvioinnissa. (Königstorfer & Thalmann 2020, s. 5; Onay & Öztürk 2018, s. 382) Esimerkiksi käyttäytymiseen liittyvän datan on todettu parantavan mallien tehokkuutta. Henkilön käyttäytymiseen liittyvää dataa voidaan kerätä muun muassa luottokortti tiedoista ja rahan siirroista. Esimerkiksi jos lainahakija tekee säännöllisesti kauppaa maksuhäiriöisen henkilön kanssa tai käyttää jatkuvasti rahaa kasinoihin, voi tällainen käyttäytyminen kertoa mahdollisesta maksukyvyttömyydestä. (Tobback & Martens 2019, s. 1227-1229) Käyttäytymiseen liittyvää dataa on myös paljon saatavilla sosiaalisessa mediassa, joka voi edesauttaa maksukyvykkyyden ennustamisessa. Zhang et al. (2016) muun muassa huomauttaa, että henkilön sosiaalinen käyttäytyminen heijastaa hänen luonnettaan, ja tätä tietoa voidaan käyttää hyväksi luottoriskin ennustamisessa. Osa luottoyhtiöistä käyttää jo sosiaalisesta mediasta, kuten Facebookista, Twitteristä ja LinkedInista saatavaa tietoa hyväksi luottoriskin hallinnassa. Esimerkiksi henkilön sosiaalinen identiteetti, nettimaine ja ammatilliset kontaktit ovat tekijöitä, joita voi hyödyntää luottoriskin ennustamisessa. (Zhang et al. 2016, s. 169) Sosiaalisessa mediassa on myös laajasti saatavilla yritysten uutisia, jotka tarjoavat tietoa yrityksen luottokelpoisuudesta. Esimerkiksi informaatio yrityksen suhteista, asemasta markkinoilta, strategiasta tai johdon ja hallituksen pätevyydestä voi indikoida yrityksen taloudellisesta vakaudesta. (Doumpos et al. 2019, s. 32) Myös puhelin käyttäytymisellä, kuten tekstiviesti- ja soittokäyttäytymisellä on todettu olevan yhteys luottoriskin todennäköisyyteen. Tekemässään tutkimuksessa Oskarsdottir et al. (2019, s. 37) huomasivat pelkästään soittokäyttäytymisdatan lisäämisen malleihin parantavan ennustettavuutta niin huomattavasti, että puhelindataa voidaan käyttää jopa ainoana datanlähteenä luottopäätöstä tehdessä. Vahvuutena näillä uusilla datalähteillä on myös tiedon saatavuus: monista kotitalouksista, etenkin kehitysmaissa, ei ole virallista taloudellista tietoa saatavilla, mikä hankaloittaa lainanhakuprosessia. Suuri osa väestöstä omistaa kuitenkin

matkapuhelimen, jolloin yhä useammat voivat tulla luottoluokitelluksi. (van Thiel & van Raaij 2019, s. 273)

Monipuolisemman datan käytön lisäksi tekoäly mahdollistaa myös uusia, tarkempia algoritmeja riskin arviointiin. Sigrist ja Hirnschall ovat esimerkiksi kehittäneet niin sanotun Grabit algoritmin, jossa perinteisten tietolähteiden tukena käytetään käyttäytymiseen liittyvää dataa, kuten sosiaalisen median dataa, sisäänkirjautumistietoja sekä click dataa. (Königstorfer & Thalmann 2020, s. 5) Grabit malli, joka lukeutuu regression malleihin, voi parantaa huomattavasti maksukyvyttömyyden ennustusten tarkkuutta (Sigrist & Hirnschall 2019, s. 190). Koneoppimisen tekniikat mahdollistavat myös big datan käytön luottoriskin hallinnassa. Koneoppiminen on tekoälyn osa-alue, joka voi tunnistaa kuvioita, oppia datasta ja tehdä siitä ennusteita. (Zhu et al. 2017, s. 42) Sen avulla laadullisten tekijöiden, kuten maksuhalukkuuden, huomioon ottaminen päätösprosessissa on helpottunut. Koneoppimisalgoritmit mahdollistavat myös nopeamman ja halvemmän luottopäätösprosessin. (Tadapaneni 2020, s. 2793) Neuroverkot ovat myös yleisesti tutkittu lähestymistapa luottoriskin hallinnassa. Neuroverkot ovat mallinnustyökalu, joka kehitettiin simuloimalla ihmisen aivoja. Neuroverkoissa ihmisen oppimiskyky on siirretty tietokoneympäristöön, joka mahdollistaa sen oppimisen esimerkkien pohjalta. Niitä on käytetty laajasti ennustukseen jo 1990-luvulta lähtien ja niiden tarkkuuden on havaittu olevan perinteisiä malleja parempi. Neuroverkkojen käytössä on kuitenkin havaittu olevan joitain haasteita, kuten niiden vaatima pitkä koulutusprosessi. (Akkoç 2012, s. 170)

Vaikka näillä uusilla menetelmillä on monia hyötyjä perinteisiin malleihin verrattuna, liittyy niihin myös erilaisia haasteita. Esimerkiksi koneoppimisen käyttö lainapäätöksessä aiheuttaa huolta niin sanotusta 'musta laatikko' ongelmasta, jossa asiakas ei saa perusteluja lainan hylkäys- tai hyväksymispäätöksestä. Myös ei-perinteisten datalähteiden käyttö luottopäätöksessä voi aiheuttaa erilaisia ongelmia. Kuluttajajärjestöt huomauttavat niiden voivan aiheuttaa ennakkoluuloja esimerkiksi lainanhakijan sukupuoleen tai etniseen taustaan liittyen. Myös tietosuojan ja yksityisyyteen liittyy ongelmia ei-perinteisten datalähteiden, kuten puhelindatan käytössä. Tämä tuo haasteita lainanantajille, joiden täytyy harkita näiden uusien datalähteiden käyttöä eettisestä näkökulmasta, jotta he voivat säilyttää asiakkaidensa ja työntekijöidensä luottamuksen. (Tadapaneni 2020, s. 2793; Königstorfer & Thalmann 2020, s. 10) Valtioiden on määriteltävä erilaisia lakeja ja määräyksiä vähentääkseen näitä riskejä. Esimerkiksi vuonna 2021 Euroopan komissio esitti uudet sääntelyt tekoälyn käytön

rajoittamiseksi, jossa luottoluokitus tekoälyn avulla todettiin korkean riskin käyttötarkoitukseksi. Käytännössä tämä tarkoittaa, että tekoälyn tekniikoiden täytyisi läpäistä tiukat velvoitteet, jotta niitä voidaan käyttää luottopäätöksessä. Luottopäätöksessä tulisi esimerkiksi olla mukana asianmukaiset ihmisen valvontatoimenpiteet riskien minimoimiseksi, datan täytyisi olla riittävän korkealaatuista syrjivien tulosten vähentämiseksi ja tekniikan pitäisi täyttää korkeat kestävyden, turvallisuuden ja tarkkuuden vaatimukset. (European Commission 2021) Monissa maissa on myös käytössä lainsäädäntö, joka vaatii rahoituslaitoksia perustelemaan syyt lainan hylkäyspäätöksen takana (Toback & Martens 2019, s. 1228). Myös itse toteutus saattaa aiheuttaa organisaatioille erilaisia ongelmia. Haasteellista voi olla esimerkiksi oikean datan ja algoritmin valitseminen sekä osaavan henkilökunnan löytäminen. (Königstorfer & Thalmann 2020, s. 10) Näiden haasteiden vuoksi luottopisteitys perustuu käytännössä vielä pääasiassa yksinkertaisiin luokitteluihin, kuten logistiseen regressioon, erotteluanalyysiin sekä luokituspuihin (Toback & Martens 2019, s. 1228).

5 ANALYYSI LUOTTORISKIN ENNUSTAMISESTA

Tässä osiossa suoritetaan teorian mukainen luottoriskianalyysi 50:lle julkisesti noteeratulle yhdysvaltalaiselle teollisuusyritykselle. Tarkoituksena on esittää resurssien puitteissa yksinkertaistetusti ja sovelletusti, miten luottopääätösprosessi etenee. Niin kuin teoriassa on esitetty, yleinen dataohjautuva lähestymistapa luottoriskin ennustamiseen on kerätä eri tietolähteistä saatavaa dataa mahdollisista tekijöistä, jotka vaikuttavat lainanhakijan kykyyn tai halukkuuteen maksaa laina takaisin sovittujen ehtojen mukaisesti. Tietoja verrataan lainan takaisinmaksuun, jonka perusteella voidaan rakentaa malli, joka ennustaa tulevien lainahakijoiden maksukyvyttömyyden todennäköisyyden. Tässä työssä suoritettavassa analyysissä ei kehitetä omaa luokitteluun käytettävää mallia, vaan käytetään julkisesti saatavilla olevaa Altman Z-pisteytystä. Altman Z-pisteytysmalli on Edward Altmanin vuonna 1968 kehittämä malli konkurssin ennustamiseen. Mallissa käytetään hyväksi yritysten taloudellisia tunnuslukuja, ja se on todettu luotettavaksi tekniikaksi ennustamaan yrityksen konkurssi kahden vuoden päähän. Mallia on sovellettu monipuolisesti eri käyttötarkoituksiin, joista yksi yleisimmistä on luottoriskin ennustaminen. Malli on kehitelty yhdysvaltaisten teollisuusyritysten datan pohjalta, jonka vuoksi tässä työssä valittiin vastaavia yrityksiä mahdollisimman luotettavien tulosten varmentamiseksi. (Altman & Hotchkiss 2006, s. 235; Altman et al. 2017, s. 132-135; Altman 2018, s. 1-2; Altman 1968, s. 609) Altman Z-luku (Z) lasketaan seuraavan kaavan mukaisesti (Altman et al. 2017, s. 135):

$$Z = 0.012*A + 0.014*B + 0.033*C + 0.006*D + 0.999*E$$

A: Nettokäyttöpääoma / Koko pääoma

B: Kertyneet voittovarot / Koko pääoma

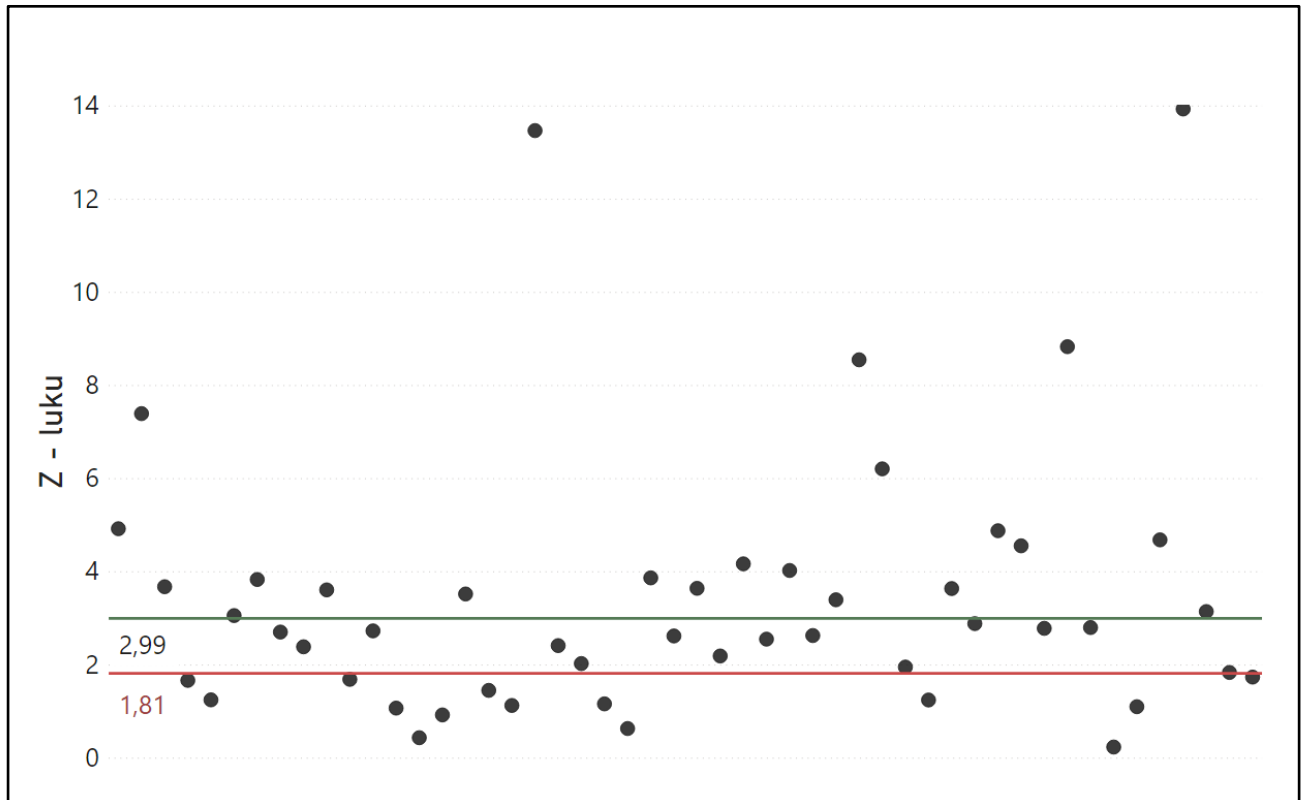
C: Tulos ennen korkoja ja veroja / Koko pääoma

D: Oman pääoman markkina-arvo / Vieraan pääoman kirjanpitoarvo

E: Myynti / Koko pääoma

Tässä työssä käytettävien yritysten taloudelliset tunnusluvut on haettu Yahoo Finance:n verkkosivuilta ja kullekin yritykselle on laskettu yllä olevan kaavan mukaisesti oma Z-luku. Z-luvun perusteella yritykset voidaan jakaa kolmeen eri luokkaan taloudellisen hyvinvoinnin perusteella. Yritykset, jotka saavat Z-luvuksi yli 2.99 ovat taloudellisesti vahvassa asemassa. Z-luvuksi alle 1.81 saaville yrityksille puolestaan konkurssi ja maksukyvyttömyys on todennäköinen. Yritykset, jotka saavat Z-luvun tältä väliltä luokitellaan niin sanottuun harmaaseen alueeseen, jossa arvojen suhteen syntyy virheluokittelua. (Altman & Hotchkiss 2006, s. 244; Altman 1968, s. 606) Kuvassa 2 esitetään tämän työn yritysten Z-lukujen arvot. Kuvaajat on toteutettu Power BI -ohjelmiston avulla.

Suurin osa yrityksistä (22 yritystä) saivat Z-luvuksi yli 2,99 ja ovat taloudellisesti vahvassa asemassa eikä konkurssi tai maksukyvyttömyys ole todennäköinen. 14 yrityksen Z-luku on alle 1,81. Nämä yritykset ovat mallin mukaan taloudellisessa ahdingossa ja maksukyvyttömyyden uhka on suuri. Toiset 14 yritystä kuuluvat harmaaseen alueeseen, jolloin niitä ei luokitella kumpaakaan kategoriaan.



Kuva 2. Z-lukujen hajonta

Vaikka Altmanin malli, joka on pääasiassa tarkoitettu konkurssin ennustamiseen, kuvaa hyvin yritysten taloudellista hyvinvointia ja luottokelpoisuutta, se ei kerro luottoriskin kannalta olennaista yrityksen luottoluokkaa. Standard & Poor's on koontanut Z-luvuista kullekin joukkovelkojen luokalle mediaani luvun eri ajankohtina. Pisteet on esitetty taulukossa 4. Tässä työssä yritysten luottoluokitteluun käytetään tätä taulukkoa ja kullekin yritykselle annetaan lähimpänä olevan mediaani Z-luvun mukainen luokitus. Pisteet katsotaan vuoden 2017 sarakkeesta, jotka kertovat viimeisimmät mediaaniluvut kussakin luokassa.

Taulukko 4. Yhdysvaltaisten teollisuusyritysten mediaani Z-luku S&P joukkolainojen luottoluokille 1992-2017 (Altman 2018, s. 4)

Rating	2017	2013	2004-2010	1996-2001	1992-1995
AAA/A	4.20	4.13	4.18	6.20	4.80
A	3.85	4.00	3.71	4.22	3.87
BBB	3.10	3.01	3.26	3.74	2.75
BB	2.45	2.69	2.48	2.81	2.25
B	1.65	1.66	1.74	1.80	1.87
CCC/CC	0.73	0.23	0.46	0.33	0.40
D	-0.10	0.01	-0.04	-0.20	0.05

Tässä työssä käytettävien yritysten luokitukset näkyvät taulukossa 5. Yrityksistä 25 kuuluu investointiluokkaan (vähintään BBB-luokituksen saaneet) ja näiden yritysten maksukyvyttömyyden todennäköisyys on suhteellisen pieni. Toiset 25 yritystä kuuluvat puolestaan spekulatiiviseen luokitukseen, jossa riski maksukyvyttömyydelle on suurempi. Lainanantajat tekevät luokituksen perusteella päätöksen lainan myönnöstä. Usein mitä matalamman luokituksen yritys saa sitä korkeammat korkomaksut se joutuu lainasta maksamaan.

Taulukko 5. Analysoitujen yritysten luokitukset

Rating	Yritysten lukumäärä	Prosenttiosuus
AAA/A	11	22 %
A	8	16 %
BBB	6	12 %
BB	8	16 %
B	9	18 %
CCC/CC	7	14 %
D	1	2 %

6 JOHTOPÄÄTÖKSET

Tässä kandidaatintyössä käsiteltiin data-analytiikan käyttöä luottoriskin hallinnassa. Työ toteutettiin kirjallisuuskatsauksena ja viimeisessä kappaleessa suoritettiin esimerkinmukainen analyysi luottoriskin ennustamisesta. Päättökysymys, johon pyrittiin vastaamaan, oli:

1. Miten data-analytiikkaa voidaan hyödyntää luottoriskin hallinnassa?

Seuraavien alatutkimuskysymysten avulla pyrittiin esittämään tarkempaa tietoa työn eri osaluista:

2. Mitä data-analytiikka on ja miten data-analytiikka prosessi etenee?

3. Mitä luottoriskit ovat?

Luottoriski tarkoittaa riskiä siitä, että lainanottaja ei maksa lainaa sopimuksen mukaisesti takaisin. Luottoriskit ovat merkittävä riski yrityksille ja rahoituslaitoksille, jotka tarjoavat asiakkailleen erilaisia yritys- ja kulutusluottoja. Luottoriskejä pyritään hallitsemaan muun muassa luottoluokitus ja -pisteytysmenetelmillä, jossa pääajatuksena on luokitella lainanhakijat 'hyviin' ja 'huonoihin' ryhmiin maksukyvyttömyyden todennäköisyyden perusteella. Työssä pyrittiin hahmottamaan luottopisteytysprosessia toteuttamalla sovellettu esimerkki luottopisteytyksestä Altman Z-luvun avulla. Altman Z-pisteytysmalli on Edward Altmanin kehittämä malli konkurssin ennustamiseen, jonka yksi yleisimmistä sovelluskohteista on luottoriskin ennustaminen. Analyysissä otettiin huomioon kaavan mukaisesti yritysten taloudellisia tunnuslukuja, joiden pohjalta kullekin yritykselle laskettiin oma Z-luku. Lukua verrattiin yhdysvaltalaisen teollisuusyritysten mediaani Z-lukuun S&P joukkolainojen luottoluokille, ja jokainen yritys luokiteltiin sitä vastaavaan luokkaan. Analyysi oli sovellettu esimerkki luottopäätösprosessin etenemisestä. Malli on kehitelty yhdysvaltalaisen julkisesti noteerattujen teollisuusyritysten datan pohjalta.

Data-analytiikka tarkoittaa tietoaaineistoista tehtävää analyysia erilaisia tietokonejärjestelmiä hyväksikäyttäen. Datasta pyritään tunnistamaan kaavoja ja yhtäläisyyksiä, jotka voivat edesauttaa organisaatioita ja niiden päätöksentekoa. Data-analytiikan prosessi on monivaiheinen, ja sillä pyritään muuttamaan raakadata tietoon ja toimintoihin. Prosessiin

kuuluu muun muassa datan kerääminen ja käsittely, tutkiminen ja arviointi ja viimeisessä vaiheessa saadun informaation muuntaminen toiminnaksi.

Työssä tunnistettiin data-analytiikalla olevan jo suuri merkitys luottoriskin hallinnassa. Sitä käytetään uusien mallien rakentamiseen historiallisen datan pohjalta ajallaan maksetuista lainoista ja lainoista, jossa on ilmennyt maksukyvyttömyys. Näin mallit tunnistavat kuvioita datasta ja muodostavat suhteita maksukyvyttömyyden todennäköisyyden ja riskitekijöiden välille. Tekoälyn ja big datan kehittymisen myötä uusia tekniikoita sekä uusia datanlähteitä on alettu ottamaan huomioon luottoriskin ennustamisessa. Esimerkiksi sosiaalisesta mediasta tai puhelinkäyttäytymisestä kerättävä data voi antaa merkittävää tietoa hakijan lainan takaisinmaksukyvyistä. Tällaisen tiedon tehokas käyttäminen luottopäätöksessä vaatii kuitenkin kehittyntä teknologiaa, joka voi olla sen käytön esteenä monille yrityksille. Se on osasy sille, miksi luottopisteytys perustuu käytännössä vielä yksinkertaisiin luokitteluihin, kuten logistiseen regressioon, erotteluanalyysihin sekä luokituspuihin. Tekoälyn laajempi kehittyminen näyttää siten olevan tärkein tekijä, jonka avulla luottoriskin arvioinnista on mahdollista tulla tehokkaampaa, halvempaa ja automatisoidumpaa kuin yksikään aikaisemmin käytetty menetelmä. Suurimpina haasteina ovat kuitenkin valtioiden erilaiset lait ja rajoitukset, jotka säännöstelevät sen käyttöä luoton myönnössä.

LÄHTEET

- Adamko, P., Kliestik, T. & Birtus, M. 2014. History of credit risk models. Economics and Social Science. 2nd International Conference on Economics and Social Science (ICISS 2014). *Advances in Education Research*. Vol. 61, s. 3-336.
- Adelman, S. & Moss, L. 2000. Data Warehouse Project Management. Addison-Wesley, NJ.
- Akkoç, S. 2012. An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European journal of operational research*. Vol. 222, nro. 1, s. 168–178.
- Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of finance (New York)*. Vol. 23, nro. 4, s. 589–609.
- Altman, E. I. & Hotchkiss, E. 2006. Corporate financial distress and bankruptcy predict and avoid bankruptcy, analyze and invest in distressed debt. 3rd ed. Hoboken, N.J: Wiley.
- Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E.K. & Suvas, A. 2017. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman’s Z-Score Model. *Journal of international financial management & accounting*. Vol. 28, nro. 2, s. 131–171.
- Altman, E. I. 2018. Applications of Distress Prediction Models: What Have We Learned After 50 Years from the Z-Score Models? *International journal of financial studies*. Vol. 6, nro. 3, s. 70–.
- Azam, R., Danish, M. & Akbar, S. S. 2012. The significance of socioeconomic factors on personal loan decision (a study of consumer banking local private banks of Pakistan). *MPRA Paper 4322*, University Library Munich, Germany.
- Bonfim, D. 2009. Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics. *Journal of Banking & Finance*. Vol. 33, nro. 2, s. 281-299.
- Buyya, R., Calheiros, R. N. & Dastjerdi, A.V. 2016. Big Data: Principles and Paradigms. Cambridge, MA: Elsevier/Morgan Kaufmann.
- Carling, K., Jacobson T., Linde, J. & Roszbach, K. 2007. Corporate credit risk modeling and the macroeconomy. *Journal of Banking & Finance*. Vol. 31, nro. 3, s. 845-868.

- Castro, V. 2013. Macroeconomic determinants of the credit risk in the banking system: The case of the GIPSI. *Economic modelling*. Vol. 31, s. 672-683.
- Chen, H., Chiang, R. H. L. & Storey, V. C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*. Vol. 36, nro. 4, s. 1165-1188.
- Clark, D. 2020. Beginning Microsoft Power BI: A Practical Guide to Self-Service Data Analytics. 3rd edition. Berkeley, CA: Apress L. P. 403 s.
- Davenport, T. H. & Harris, J. 2007. Analysoi ja voita – kilpailun uusi tiede. Helsinki: Talentum. 253 s.
- Doumpos, M., Lemonakis, C., Niklis, D. & Zopounidis, C. 2019. Analytical Techniques in the Assessment of Credit Risk: An Overview of Methodologies and Applications. Cham: Springer International Publishing AG. 111 s.
- Finanssivalvonta. 2018. Määräykset ja ohjeet 4/2018: Luottoriskien hallinta rahoitussektorin valvottavissa. [WWW-dokumentti]. [viitattu 9.3.2021]. Saatavissa: [2018_04.m2.pdf \(finanssivalvonta.fi\)](#)
- Gandomi, A. & Haider, M. 2015. Beyond the hype: Big data concepts, methods and analytics. *International journal of information management*. Vol. 35, nro. 2, s. 137-144.
- Getter, D. E. 2006. Consumer Credit Risk and Pricing. *The Journal of consumer affairs*. Vol. 40, nro. 1, s. 41–63.
- Goodwyn, G. 2019. Business Analytics. *The Journal of American Society of Military comptroller*. Vol. 64, nro. 1, s. 33–34.
- Harris, T. 2013. Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions. *Expert Systems with Applications*. Vol. 40, s. 4404-4413.
- Intrum. 2019. European Payment Report 2019: 21st Annual Edition.
- Kowalczyk, M. 2017. The Support of Decision Processes with Business Intelligence and Analytics: Insights on the Roles of Ambidexterity, Information Processing and Advice. Wiesbaden: Springer Fachmedien Wiesbaden GmbH. 146 s.

Kwon, O. Lee N. & Shin, B. 2014. Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*. Vol. 34, nro. 2, s. 387–394.

Königstorfer, F. & Thalmann, S. 2020. Applications of Artificial Intelligence in commercial banks – A research agenda for behavioral finance. *Journal of behavioral and experimental finance*. Vol. 27, s. 1-15.

Lawrence, E.C., Smith, L.D. & Rhoades, M. 1992. An analysis of default risk in mobile home credit. *Journal of Banking & Finance*. Vol. 16, nro. 2, s. 299-312.

Liberatore, M. J. & Luo, W. 2010. The Analytics Movement: Implications for Operations Research. *Interfaces*. Vol. 40, nro. 4, s. 313-324.

Mach-Król, M., M. Olzak, C. & Pelech-Pilichowski, T. 2015. Advances in ICT for Business, Industry and Public Sector. Cham: Springer International Publishing. 189 s.

Mandala, G., Nawangpalupi, C. & Praktikto, F. 2012. Assessing Credit Risk: An Application of Data Mining in a Rural Bank. *Procedia Economics and Finance*. Vol. 4, s. 406-412.

Martinsuo, M. & Kärri, T. 2017. Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa. 1. painos. Helsinki: Kunnossapitoyhdistys Promaint ry. 238 s.

MATLAB. n.d.a. Math. Graphics. Programming. [WWW-dokumentti]. [viitattu 2.3.2021]. Saatavissa: [MATLAB - MathWorks - MATLAB & Simulink](#)

MATLAB. n.d.b. MATLAB for Data Analysis. [WWW-dokumentti]. [viitattu 2.3.2021]. Saatavissa: [Data Analysis – MATLAB & Simulink - MATLAB & Simulink \(mathworks.com\)](#)

Memmel, C., Gündüz, Y. & Raupach, P. 2015. The common drivers of default risk. *Journal of Financial Stability*. Vol. 16, s. 232-247.

Moradi, S. & Mokhatab Rafiei, F. 2019. A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial innovation (Heidelberg)*. Vol. 5, nro. 1, s. 1–27.

Morrison, A. 2020. Power BI Vs. Excel. [WWW-dokumentti]. [viitattu 1.3.2020]. Saatavissa: [Power BI Vs. Excel | Zarantech](#)

Neal, R. S. 1996. Credit derivatives: New financial instruments for controlling credit risk. *Economic Review*. Vol. 81, nro. 2.

Nelli, F. 2018. Python Data Analytics: With Pandas, NumPy and Matplotlib. 2nd edition. Berkley, CA: Apress L. P. 569 s.

Onay, C. & Öztürk, E. 2018. A review of credit scoring research in the age of Big Data. *Journal of financial regulation and compliance*. Vol. 26, nro. 3, s. 382-405.

Oskarsdottir, M., Bravo, C., Sarraute, C., Vanthienen, J. & Baesens, B. 2019. The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing Journal*. Vol. 74, s. 26-39.

Pathak, M. A. 2014. Beginning Data Science with R. Cham: Springer International Publishing. 157 s.

Qlik. n.d. Turn raw data into remarkable outcomes. [WWW-dokumentti]. [viitattu 2.3.2021].
Saatavissa: [The Platform for the Next Generation of BI \(qlik.com\)](https://www.qlik.com)

Rausch, P., Sheta, A. F. & Ayes, A. 2013. Business Intelligence and Performance Management: Theory, Systems and Industrial Applications. Springer. 269 s.

Romano, F., Phillips, D. & Hattem, R. 2016. Python: Journey from Novice to Expert. Birmingham: Packt Publishing Ltd. 1287 s.

Runkler, T.A. 2016. Data Analytics Models and Algorithms for Intelligent Data Analysis. 2nd ed. 2016. Wiesbaden: Springer Fachmedien Wiesbaden.

SAS. 2021. SAS Products, Technology & Solutions. [WWW-dokumentti]. [viitattu 2.3.2021].
Saatavissa: [Products & Solutions | SAS](https://www.sas.com)

Saxena, R. & Srinivasan, A. 2013. Business Analytics: A Practitioner's Guide. New York: NY: Springer New York. 162 s.

Schuster, A. J. 2017. Understanding Information: From the Big Bang to Big Data. Cham: Springer International Publishing. 237 s.

Seddon, P. B., Constantinidis, D., Toomas, T. & Harjot, D. 2017. How does business analytics contribute to business value? *Information systems journal (Oxford, England)*. Vol. 27, nro. 3, s. 237–269.

Shmueli, G., Bruce, P. C. & Patel, N. R. 2016. Data mining for business analytics: concepts, techniques and applications with XLMiner. 3rd edition. Wiley-Blackwell. 514 s.

Sigrist, F. & Hirnischall, C. 2019. Grabit: Gradient tree-boosted Tobit models for default prediction. *Journal of banking and finance*. Vol. 102, s. 177-192.

Tableau. n.d. Turn insights into action with the Tableau product suite. [WWW-dokumentti]. [viitattu 2.3.2021]. Saatavissa: [Our Products \(tableau.com\)](https://www.tableau.com)

Tadapaneni, N. 2020. Artificial Intelligence in Finance and Investments. *International Journal of Innovative Research in Science, Engineering and Technology*. Vol. 9, nro. 5, s. 2792-2795.

Tobback, E. & Martens, D. 2019. Retail credit scoring using fine-grained payment data. *Journal of the Royal Statistical Society. Series A, Statistics in society*. Vol. 182, nro. 4, s. 1227-1246.

Tukey, J. W. 1962. The Future of Data Analysis. *The Annals of Mathematical Statistics*. Vol. 33, nro. 1, s. 1-67.

van Thiel, D. & van Raaij, W. F. 2019. Artificial intelligence credit risk prediction: An empirical study of analytical artificial intelligence tools for credit risk prediction in a digital era. *Journal of risk management in financial institutions*. Vol. 12, nro. 3, s. 268–286.

Walkenbach, J. 2016. Microsoft Excel 2016 bible. 1st edition. New York: Wiley. 1061 s.

Zhang, Y., Jia, H., Diao, Y., Hai, M. & Li, H. 2016. Research on Credit Scoring by fusing social media information in Online Peer-to-Peer Lending. *Procedia Computer Science*. Vol. 91, s. 168-174.

Zhu, Y., Xie, C., Wang, G. & Yan, X. 2017. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. *Neural computing and applications*. Vol. 28, s. 41-50.