



Aalto-yliopisto
Perustieteiden
korkeakoulu

Teollisuusyritysten konkurssien ennustaminen logistisella regressiolla (valmiin työn esittely)

Juho Lahti

12.04.2024

Ohjaaja: DI *Leevi Olander*

Valvoja: Prof. *Ahti Salo*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Tausta

- Yrityskonkurssi on menettely, jossa maksukyvyttömän yrityksen ulosmitattava omaisuus myydään ja jaetaan velkojien kesken
- Konkurssit aiheuttavat suuria tappioita eri sidosryhmille
- Konkurssien ennustamista on laajalti tutkittu
- Ennustuksen tavoitteena on luoda ”hälytysjärjestelmä”

Tavoitteet

- Rakennetaan tilinpäätöstietoihin tukeutuva logistinen regressiomalli yritysten konkurssiriskien arvioimiseksi
- Selvitetään, miten konkurssitodennäköisyys riippuu konkurssiriskiin vaikuttavista tekijöistä
- Selvitetään aikaisempien ennustemallien soveltuvuutta teollisuusyritysten konkurssien ennustamiseen

Menetelmät

- Logistinen regressio
 - Yleistetty lineaarinen malli, soveltuu binääriseen luokitteluongelmaan
 - Selittävinä muuttujina tilinpäätöksestä johdetut tunnusluvut, vastemuuttujana konkurssitapahtuma (0 = ei konkurssia, 1 = konkurssi)
- Muuttujien esivalinta
 - Pohjautuu aikaisemmille tutkimuksille, muuttujia kategorisoitu (kannattavuus, maksuvalmius, vakavaraisuus...)

Menetelmät

- $g(E(y_i)) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$ 1)

- $\ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$ 2)

$$\Leftrightarrow \pi(\mathbf{x}) = \frac{1}{1 + e^{-\beta_0 - \sum_{j=1}^p \beta_j x_{ij}}} \quad 3)$$

- 1) Yleistetty lineaarinen malli
- 2) Logit-muunnos
- 3) Ratkaistu todennäköisyys $\pi(\mathbf{x})$

Menetelmät

- Mallin arviointi luokittelutaulukon avulla

	Ennustettu 0	Ennustettu 1
Todellinen 0	TN	FP
Todellinen 1	FN	TP

- Luokittelutaulukon avulla saadaan laskettua mallin suorituskykyä kuvaavia

metriikoita, esim: tarkkuus (accuracy) = $\frac{TP+TN}{TP+TN+FP+FN}$,

sensitiivisyys (recall) = $\frac{TP}{TP+FN}$ sekä positiivinen ennustearvo (precision) = $\frac{TP}{TP+FP}$

- Mallin arvioinnissa hyödynnetään lisäksi erottelukykykäyrää (ROC), joka on graafinen esitys mallin suorituskyvystä eri kynnyksisarvoilla

Aineisto

- Aineistona teolliset pk-yritykset EU:ssa ja niiden tilinpäätöstiedot
- Konkurssitapahtumat vuosilta 2020-2021, tilinpäätöstiedot kattavat kolme tilikautta ennen konkurssitapahtumaa
- Esivalinnan perusteella valittiin muuttujia 17 eri kategorioista (kannattavuus, vakavaraisuus, maksuvalmius ja kasvu)
- Ennusteita arvioitiin laatimalla ne kaksi vuotta konkurssitapahtumaa edeltäneen aineiston pohjalta

Aineisto

	Konkurssiyrietykset	Aktiiviset maksukykyiset yritykset
Yhteensä	104	156 759
Lopullinen aineisto	62	41 813

- Monen yrityksen osalta tilinpäätöstiedoissa puutteita tai vajavaisuutta

Esivalinnan perusteella valitut tunnusluvut

Nro	Tunnusluku	Kategoria
1	Sijoitetun pääoman tuotto prosentti (ROI)	Kannattavuus
2	Oman pääoman tuotto prosentti (ROE)	Kannattavuus
3	Kokonaispääoman tuotto prosentti (ROA)	Kannattavuus
4	Liiketulos (EBIT)	Kannattavuus
5	Käyttökate (EBITDA)	Kannattavuus
6	Net gearing	Kannattavuus
7	Omavaraisuusaste prosentti (Oma pääoma / taseen loppusumma)	Vakavaraisuus
8	Vieraan pääoman takaisinmaksukyky (Rahoitustulos / [vastattavaa yhteensä – oma pääoma])	Vakavaraisuus
9	Lyhytaikainen vieras pääoma / taseen loppusumma	Vakavaraisuus
10	Oma pääoma / vieras pääoma	Vakavaraisuus
11	Quick ratio (Rahoitusomaisuus / lyhytaikainen vieras pääoma)	Maksuvalmius
12	Current ratio	Maksuvalmius
13	Rahoitustulos prosentti	Maksuvalmius
14	Käyttöpääoma / taseen loppusumma	Maksuvalmius
15	Myyntisaamisten kiertoaika	Maksuvalmius
16	Liikevaihdon kasvuprosentti	Kasvu
17	Jalostusarvon kasvuprosentti	Kasvu

Kuvailevat tunnusluvut kaksi vuotta ennen konkurssia

Tunnusluku	Aineiston ominaisuus	Konkurssit	Aktiiviset
Oman pääoman tuotto-% (ROE)	Keskiarvo	-4,03	11,03
	Mediaani	3,38	8,88
	Keskihajonta	66,68	24,08
Kokonaispääoman tuotto-% (ROA)	Keskiarvo	-0,22	4,12
	Mediaani	1,18	2,75
	Keskihajonta	6,17	6,60
Liiketulos (EBIT)	Keskiarvo	140,75	413,01
	Mediaani	47,81	168,06
	Keskihajonta	632,20	1033,85
Käyttökate (EBITDA)	Keskiarvo	381,77	692,84
	Mediaani	125,21	337,25
	Keskihajonta	969,51	1206,87
Net gearing	Keskiarvo	170,56	142,65
	Mediaani	120,18	87,32
	Keskihajonta	151,20	157,12
Omavaraisuus-%	Keskiarvo	0,31	0,37
	Mediaani	0,26	0,35
	Keskihajonta	0,18	0,20
Vieraan pääoman takaisinmaksukyky	Keskiarvo	-0,02	-0,01
	Mediaani	-0,01	-0,01
	Keskihajonta	0,03	0,05
Lyhytaikainen vieras pääoma / taseen loppusumma	Keskiarvo	0,15	0,13
	Mediaani	0,12	0,09
	Keskihajonta	0,13	0,11
Oma pääoma / vieras pääoma	Keskiarvo	0,66	0,91
	Mediaani	0,35	0,54
	Keskihajonta	0,99	1,32
Quick ratio	Keskiarvo	-0,03	-0,02
	Mediaani	-0,02	-0,01
	Keskihajonta	0,04	0,08
Current ratio	Keskiarvo	1,42	1,77
	Mediaani	1,16	1,43
	Keskihajonta	0,92	1,42
Rahoitustulos-%	Keskiarvo	-1,50	-0,78
	Mediaani	-0,99	-0,53
	Keskihajonta	1,92	8,97
Käyttöpääoma / taseen loppusumma	Keskiarvo	0,32	0,28
	Mediaani	0,32	0,27
	Keskihajonta	0,23	0,20
Liikevaihdon kasvuprosentti	Keskiarvo	19,57	11,72
	Mediaani	4,79	5,40
	Keskihajonta	64,57	388,48
Jalostusarvon kasvuprosentti	Keskiarvo	3,77	21,41
	Mediaani	1,28	5,85
	Keskihajonta	27,14	2368,46

Malli

- Esivalittujen tunnuslukujen pohjalta rakennetaan kaksi mallia
- Ensimmäinen malli hyödyntää kaikkia esivalittuja muuttujia (15 kpl), toinen malli hyödyntää viittä muuttujaa
- Aktiivisista yrityksistä valitaan satunnaisotannalla 310 yritystä mallien rakentamista varten, jolloin konkurssiyritysten (62) osuus aineistosta on 20 %
- Aineisto jaetaan mallin opetusdataan (75 %) ja testidataan (25 %), jolla mallin ennustekykyä arvioidaan
- Muuttujat skaalataan scikit-learn-kirjaston avulla, jotta regressiokertoimet ovat vertailukelpoisia keskenään

Tulokset

- Itseisarvoltaan suuret muuttujat vaikuttavat eniten konkurssiriskiin ja vastaavasti itseisarvoltaan pienet muuttujat vaikuttavat vähän konkurssiriskiin
- Muuttujan positiivinen kerroin tarkoittaa, että ao. muuttuja kasvattaa konkurssiriskiä kun taas negatiivinen vähentää sitä

Tulokset

- Regressiokertoimet kaikki esivalitut muuttujat sisältävässä mallissa

Tunnusluku	Regressiokerroin
ROE	0,2599
ROA	-0,7882
EBIT	0,3403
EBITDA	-0,1925
Net gearing	0,1317
Omavaraisuusaste prosentti	-0,0163
Current ratio	-0,6610
Vieraan pääoman takaisinmaksukyky	-0,3598
Lyhytaikainen vieras pääoma / taseen loppusumma	-0,2844
Oma pääoma / vieras pääoma	0,4636
Quick ratio	-0,1107
Rahoitustulos prosentti	-0,1325
Käyttöpääoma / taseen loppusumma	0,4825
Liikevaihdon kasvuprosentti	0,4079
Jalostusarvon kasvuprosentti	0,0111

Tulokset

- Regressiokertoimet viisimuuttujaisessa mallissa

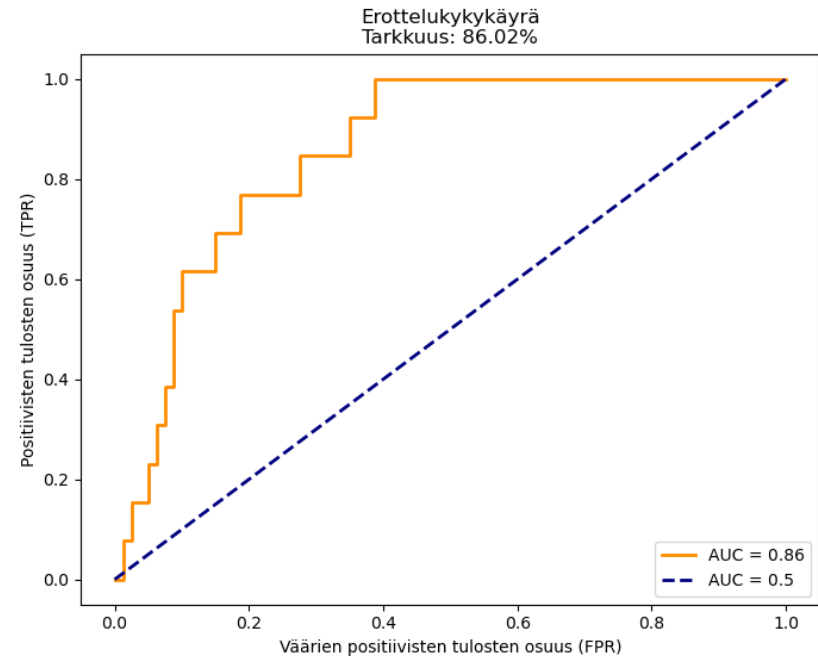
Tunnusluku	Regressiokerroin
EBITDA	-0,6691
Omavaraisuusaste	-0,1545
Vieraan pääoman takaisinmaksukyky	-0,5959
Lyhytaikainen vieras pääoma / taseen loppusumma	0,1235
Rahoitustulosprosentti	-0,7992

Tulokset

Kaikkia muuttujia hyödyntävän mallin ennustekyky

	Ennustettu 0	Ennustettu 1
Todellinen 0	78	2
Todellinen 1	11	2

- Tarkkuus = 0,86
- Sensitiivisyys = 0,15
- Positiivinen ennustearvo = 0,50

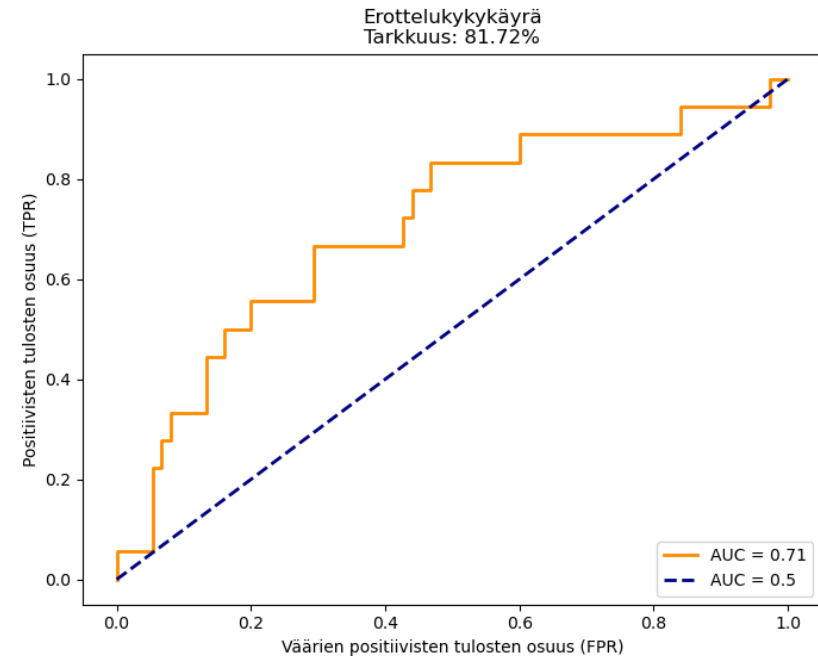


Tulokset

Viisimuuttujaisen mallin ennustekyky

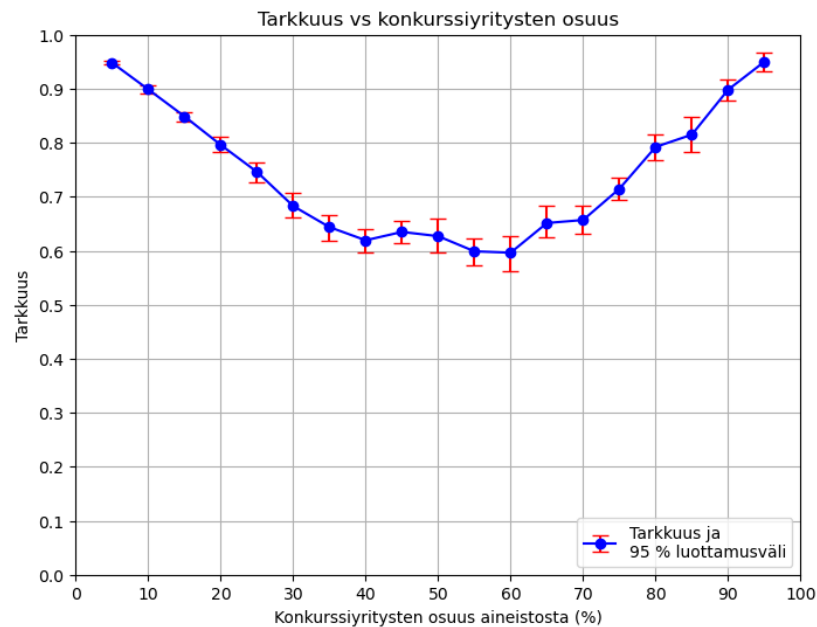
	Ennustettu 0	Ennustettu 1
Todellinen 0	75	0
Todellinen 1	17	1

- Tarkkuus = 0,82
- Sensitiivisyys = 0,06
- Positiivinen ennustearvo = 1,0

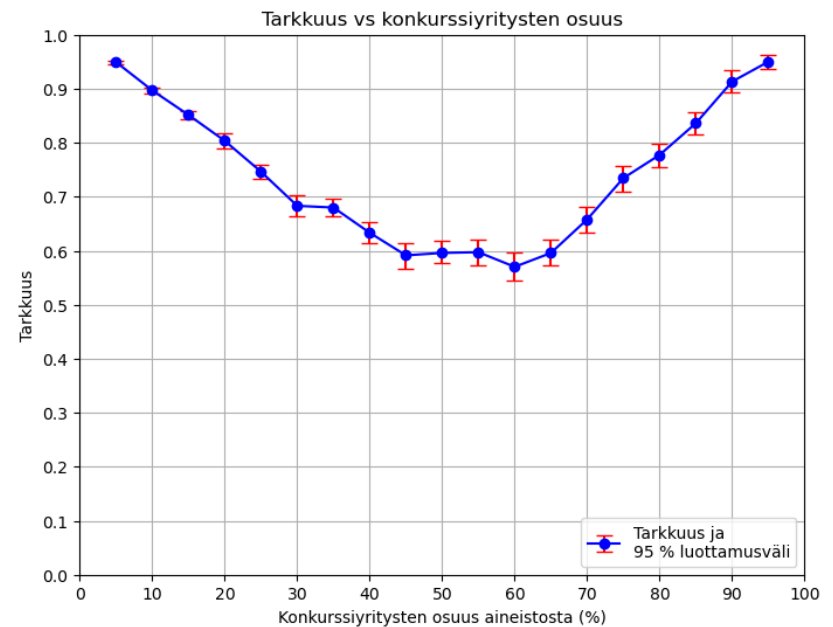


Tulokset

Mallien tarkkuudet, kun konkurssiyritysten osuutta aineistosta muutetaan



Kaikkia muuttujia hyödyntävä malli



Viisimuuttujainen malli

Yhteenveto

- Mallien perusteella konkurssiriskiä alentavat eniten kokonaispääoman tuottoprosentin (ROA), maksuvalmiutta kuvaavan current ration, rahoitustulosprosentin sekä käyttökatteen eli EBITDA:n kasvu riippuen mallissa käytetyistä muuttujista
- Vastaavasti konkurssiriskiä eniten kasvattavia tekijöitä ovat oman pääoman kasvu vieraaseen pääomaan nähden sekä käyttöpääoman kasvu taseen loppusummaan nähden
- Mallien ennustuskyky oli kuitenkin heikohko
- Mallien ennustekykyä voitaisiin parantaa etsimällä optimaalinen kynnysarvo luokittelulle tai valitsemalla muuttujat eri tavalla

Viitteet

- Aineisto: Moody's Orbis –tietokanta
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, 3(10).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
- Hackeling, G. (2017). Mastering Machine Learning with scikit-learn. Packt Publishing Ltd.
- Hilbe, J. M. (2009). Logistic Regression Models. Chapman and hall/CRC.
- Hosmer Jr, D. W., Lemeshow, S. & Sturdivant R. X. (2013). Applied Logistic Regression. John Wiley & Sons.
- Laitinen, E. K. (1990). Konkurssin ennustaminen. Vaasan Yritysinformaatio Oy.
- Lappalainen, O. (2021). EU:n rakennusalan yritysten konkurssien ennustaminen. Pro gradu -tutkielma. Itä-Suomen yliopisto.
- Lehtinen, L. (2016). Konkurssin ennustaminen tilinpäätöstiedoilla. Pro gradu -tutkielma. Aalto-yliopisto.
- McCullagh, P. (2019). Generalized Linear Models. Routledge.