

MISKOLCI EGYETEM
GAZDASÁGTUDOMÁNYI KAR

VÁLLALKOZÁSELMÉLET ÉS GYAKORLAT PH.D. PROGRAM
PROGRAMVEZETŐ: DR. NAGY ALADÁR

IMRE BALÁZS

BÁZEL II DEFINÍCIÓKON ALAPULÓ NEMFIZETÉS-ELŐREJELZÉSI
MODELLEK MAGYARORSZÁGI VÁLLALATI MINTÁN (2002-2006)

PH.D. ÉRTEKEZÉS

TUDOMÁNYOS VEZETŐ:

DR. BOZSIK SÁNDOR

MISKOLC, 2008.

ÖSSZEFOGLALÁS

Az értekezés az előrejelzési modellek témakörében folytatott kutatásaim összefoglalása.

Kiindulási alapként a pénzügyi előrejelzési modellek egyik legelterjedtebb típusa, a csődmodellek szolgáltak. A szakirodalom feldolgozása során összesen 126 csődmodellt vizsgáltam, elsősorban az alkalmazott módszertanra fókuszálva. Ezek alapján felvázoltam a csődmodellek módszertani fejlődéstörténetét, behatároltam az egyes fejlődéstörténeti szakaszokat. A fő fejlődési irányvonal mellett azonosítottam két fejlődési mellékágot is.

Kvantitatív vizsgálataim során a csődmodellek mintájára olyan prediktív modell kialakítását tűztem ki célul, amely a csődnél korábban, 90 napos fizetési késedelemnél figyelmeztet a fennálló kockázatra. Ez megfelel a Bazel II. tőkeegyezmény követelményeinek és definícióinak is.

A rendelkezésemre álló adatbázis 2000 magyarországi vállalat adatait tartalmazta a 2002-2006 közötti időszakra vonatkozóan. Az adatbázis felhasználásával több vizsgálatot is elvégeztem, három különböző modellezési technika: a döntési fa, a logisztikus regresszió és a neurális háló felhasználásával.

A kidolgozott modellek előrejelzés pontossága ugyanakkor némileg elmarad a csődmodellekétől, amelynek magyarázata az lehet, hogy a pénzügyileg egészséges és a problémás vállalatok profilja még nem különbözik olyan markánsan egymástól, mint a csődbe ment és a túlélő vállalatoké, ezért az előrejelzés is nehezebb, pontatlanabb.

Az elemzések további eredményeket is hoztak: kiválasztottam a legjobban teljesítő modellezési módszert, illetve meghatároztam az előrejelzésre leghatékonyabban alkalmazható mutatók körét. Igazoltam továbbá, hogy a modellek teljesítménye növelhető az egyes mutatók éves változásait kifejező, ún. dinamikus mutatók felhasználásával.

A kutatás eredményei elsősorban a banki hitelezés területén alkalmazhatók közvetlenül, ugyanakkor a szakirodalom módszertani alapú fejlődéstörténeti feldolgozása hozzájárult a csődmodellek elméletének fejlődéséhez is, ösztönözve a további kutatásokat.

ABSTRACT

The thesis is a summary of my researches in the topic of predictive modelling.

My starting point was the theory of bankruptcy predictions, which are clearly one of the most widespread financial predictive models. I have summarised altogether 126 bankruptcy models, focusing mainly on the applied methodology. Based on these researches I have drawn up the historical trends of the methodological evolution of the bankruptcy models and defined the historical stages of the development. Besides the mainstream evolutionary stages I have identified two dead-ends as well.

In my quantitative researches I aimed to develop a predictive model, which is able to warn earlier than the bankruptcy, at the default of 90 days past due. This is also in line with the requirements and definitions of Basel II Capital Accord.

By using a database of 2000 Hungarian companies between 2002 and 2006, I have carried out several analyses, applying three modelling methods: decision trees, logistic regression and neural networks.

The predictive power of the models are however slightly below the ones of the bankruptcy models, which is coming from the fact that the profiles of the defaulter and non-defaulter companies does not yet differ as sharply as in case of the bankrupted and survival companies.

Further results of the analyses were the identification of the best performing modelling method and the selection of the most predictive financial ratios. I have also justified that the predictive power of the models can be increased by the usage of dynamic indicators, which show the change of the financial ratios from one year to the other.

The results of my researches can be applied directly in credit activities of banks, whereas the synthesis of the methodological evolution of bankruptcy modelling has added value to the theory of bankruptcy modelling, inspiring further researches.

A TUDOMÁNYOS VEZETŐ AJÁNLÁSA

IMRE BALÁZS: *BÁZEL II DEFINÍCIÓKON ALAPULÓ NEMFIZETÉS-ELŐREJELZÉSI MODELLEK
MAGYARORSZÁGI VÁLLALATI MINTÁN (2002-2006)* CÍMŰ PH.D. ÉRTEKEZÉSÉHEZ

Imre Balázs kutatási témaválasztása gyakorlati munkájához kötődik és nagy jelentőséggel bír egy kereskedelmi bank hitelkockázatainak kezelésében. Kutatómunkájára az alaposság, az elmélyült elméleti ismeretek és a gyakorlatias megközelítés, a szakmai nyitottság és rugalmasság jellemző. Az értekezés fő részeit megvitattuk, a szerző a javaslataimat megfontolta és disszertációjának elkészítése során figyelembe vette. Kutatási eredményeit hazai konferenciákon, illetve különböző cikkekben nyilvánosságra hozta.

A műhelyvitára elkészített disszertáció-tervezetet az előopponensek tartalmilag és módszertanilag részletesen megvizsgálták. A vitán megfogalmazott értékes javaslatokat Imre Balázs a végső változat elkészítésekor messzemenőkéig figyelembe vette.

Összhangban a sikeres műhelyvitán elhangzottakkal, magam is megerősítem, hogy a benyújtott értekezés Imre Balázs saját munkája, amely a közgazdász szakmai közösség által használt tartalmi és módszertani sztendereknek mindenben megfelel. Az értekezés jelenlegi formájában nyilvános védelemre bocsátható.

A jelölt tudományos vezetőjeként meggyőződésem, hogy a kutatási eredményei az intézetben folytatott oktató munkában jól hasznosíthatók, egyes részek publikálását javaslom.

Mindezek alapján a doktori eljárás megindítását feltétlenül támogatom.

Bozsik Sándor (Ph.D)
egyetemi docens
tudományos vezető

Miskolc, 2008. május 27.

KÖSZÖNETNYILVÁNÍTÁS

Elemzés adatok nélkül sosem készülhet. Számomra is sokáig tartott, amíg a munkámhoz és kvantitatív vizsgálataimhoz szükséges adatokat találtam. Hosszadalmas keresés és számos visszautasítás után találtam rá az Info-Datex Kft-re és Szirtes Gáborra, aki rendelkezésemre bocsátotta az elemzés alapjául szolgáló értékes adatbázist. Elsősorban neki tartozom köszönettel, hogy egyáltalán elkezdhettem munkámat.

Az elemzésekhez szükséges szoftvert a SAS Institute-nál használhattam, ahol Bóna Gábortól kaptam sok segítséget.

Munkámat nagyban segítette tudományos vezetőm, Dr. Bozsik Sándor, aki szakmai konzultációkkal és személyes tanácsaival járult hozzá az értekezés folyamatos csiszolásához, végső formájának kialakításához.

Számos esetben kaptam nemcsak segítséget, hanem biztatást, támogatást is Dr. Szegedi Krisztinától. Mind neki, mind a Miskolci Egyetem Gazdaságtudományi Karán akkreditált PhD program vezetőinek, oktatóinak, munkatársainak hálás vagyok mindazért a segítségért, amelyet tanulmányaim során kaptam.

Valamennyiüknek köszönettel tartozom.

Imre Balázs

Budapest, 2008. május 28.

TARTALOM

1. BEVEZETÉS, CÉLKITŰZÉS	1
2. A CSŐDMODELLEK TÖRTÉNETI ÁTTEKINTÉSE	3
2.1. A KEZDETEK: PÉNZÜGYI MUTATÓK, MINT FIGYELMEZTETŐ JELEK	4
2.1.1. <i>Ramser, Foster, Fitzpatrick, Winakor, Smith, Merwin</i>	4
2.1.2. <i>Beaver egyváltozós elemzése</i>	5
2.2. AZ ÚTTÖRŐK: ALTMAN ÉS TÁRSAI.....	6
2.2.1. <i>Altman Z-modellje</i>	6
2.2.2. <i>Deakin</i>	9
2.2.3. <i>Blum</i>	10
2.2.4. <i>Altman-Lorris</i>	11
2.2.5. <i>A ZETA-modell</i>	11
2.2.6. <i>További modellek</i>	14
2.3. A REGRESSZIÓS MODELLEK TÉRHÓDÍTÁSA	15
2.3.1. <i>Ohlson</i>	15
2.3.2. <i>Zmijewski</i>	16
2.3.3. <i>Két banki csődmodell: Martin, illetve Pantalone-Platt modelljei</i>	17
2.3.4. <i>Zavgren</i>	18
2.3.5. <i>Keasey-McGuinness</i>	19
2.3.6. <i>Rudolph-Hamdan</i>	19
2.3.7. <i>További modellek</i>	20
2.4. MODELLEK A REKURZÍV FELOSZTÁS MÓDSZERÉVEL	21
2.4.1. <i>Frydman-Altman-Kao</i>	21
2.4.2. <i>McKee-Greenstein</i>	22
2.5. A MESTERSÉGES INTELLIGENCIA MEGJELENÉSE: NEURÁLIS HÁLÓK.....	22
2.5.1. <i>Odom és Sharda</i>	22
2.5.2. <i>Coats és Fant</i>	23
2.5.3. <i>Back, Laitinen, Sere, Wesel</i>	24
2.5.4. <i>Olmeda – Fernandez</i>	25
2.5.5. <i>Kiviluoto</i>	26
2.5.6. <i>Zhang-Hu-Patuwo</i>	26
2.5.7. <i>Laitinen – Kankaanpaa</i>	26
2.5.8. <i>További modellek</i>	27
2.6. MAGYARORSZÁGI CSŐDMODELL-KUTATÁSOK	28
2.6.1. <i>Futó Péter empirikus vizsgálatai</i>	28
2.6.2. <i>Virág Miklós első hazai csődmodellje</i>	30
2.6.3. <i>Virág Miklós– Hajdu Ottó ágazati modellcsaládja</i>	34
2.6.4. <i>Virág-Kristóf neurális háló modellje</i>	36
2.6.5. <i>Virág-Kristóf iparági rátákon alapuló modellje</i>	37
3. A CSŐDMODELLEK FEJLŐDÉSTÖRTÉNETÉNEK EGY LEHETSÉGES RENDSZEREZÉSE, MÓDSZERTANI MEGKÖZELÍTÉssel	41
3.1. A CSŐDMODELLEK ÁLTALÁNOS JELLEMZŐI.....	41
3.2. MÓDSZERTANI FEJLŐDÉSTÖRTÉNET	41
3.2.1. <i>A csődkutatások kialakulásának korszaka (1931-67)</i>	43
3.2.2. <i>A diszkriminancia analízis korszaka (1968-79)</i>	43
3.2.3. <i>A logisztikus regresszió térhódításának korszaka (1980-1989)</i>	43
3.2.4. <i>A mesterséges intelligencia megjelenésének korszaka (1990-től napjainkig)</i>	44

4. MÓDSZERTANI ÁTTEKINTÉS.....	45
4.1. A CSŐDMODELLEKNÉL ALKALMAZOTT STATISZTIKAI MÓDSZEREK	45
4.1.1. <i>Diszkriminancia analízis</i>	45
4.1.2. <i>Regressziós modellek</i>	47
4.1.2.1. Logisztikus (logit) regresszió.....	48
4.1.2.2. Probit regresszió	50
4.1.3. <i>Rekurzív partícionáló algoritmus (döntési fák)</i>	50
4.1.4. <i>Neurális hálók</i>	53
4.1.5. <i>Módszertani összefoglaló</i>	56
4.2. A MÓDSZEREK ÉRTÉKELÉSÉRE HASZNÁLT TECHNIKÁK	58
4.2.1. <i>Klasszifikációs mátrix</i>	58
4.2.2. <i>ROC (Receiver Operating Characteristic)</i>	60
4.2.3. <i>Gini-koefficiens</i>	62
4.2.4. <i>Kolmogorov-Smirnov teszt (K-S)</i>	64
5. A PREDIKTÍV MODELLEK KIDOLGOZÁSA.....	67
5.1. CSŐD VAGY DEFAULT? DEFINÍCIÓS KÉRDÉSEK	67
5.1.1. <i>A csődmodellek leggyakoribb kritikái, problémái</i>	67
5.1.2. <i>Egy lehetséges megoldás: Bázel II. definíciókon alapuló default-modellek</i>	68
5.2. A MODELLEZÉSNEEL HASZNÁLT ADATBÁZIS FELÉPÍTÉSE.....	70
5.2.1. <i>A modellezés alapját képező minta</i>	70
5.2.2. <i>A minta reprezentativitása</i>	71
5.2.3. <i>Adattisztítás, adatkorrekció</i>	72
5.2.4. <i>A modellépítésnél vizsgált mutatók köre</i>	74
5.2.4.1. <i>Pénzügyi-számviteli mutatók (objektív, hard faktorok)</i>	74
5.2.4.2. <i>Kvalitatív adatok (szubjektív, soft faktorok)</i>	79
5.3. A MUTATÓK ELŐZETES ÁTTEKINTÉSE	79
5.4. A MODELLFEJLESZTÉSI MUNKA LÉPÉSEI.....	84
5.4.1. <i>A modellező szoftver</i>	84
5.4.2. <i>A modellezés folyamata</i>	84
5.4.2.1. <i>Adatbeolvasás (Input Data Node)</i>	86
5.4.2.2. <i>Mintafelosztás (Data Partition Node)</i>	86
5.4.2.3. <i>Döntési fa (Decision Tree Node)</i>	87
5.4.2.4. <i>Logit modell (Regression Node)</i>	89
5.4.2.5. <i>Neurális háló (Neural Network Node)</i>	91
5.4.2.6. <i>Modellek kiértékelése (Model Comparison Node)</i>	93
5.5. A KIDOLGOZOTT MODELLEK, EREDMÉNYEK.....	94
5.5.1. <i>A döntési fa</i>	94
5.5.2. <i>A logit modell</i>	95
5.5.3. <i>A neurális háló</i>	96
5.5.4. <i>A modellekben szereplő döntési faktorok</i>	98
5.5.5. <i>Az egyes modellek teljesítményének összehasonlítása, értékelése</i>	102
6. TOVÁBBI TESZTEK, ELEMZÉSEK.....	107
6.1. IPARÁGI RÁTÁKON ALAPULÓ MUTATÓK FELHASZNÁLÁSA A MODELLEZÉSBEN	107
6.2. A MODELLEK TESZTELÉSE DINAMIKUS MUTATÓK BEVEZETÉSÉVEL.....	109
6.2.1. <i>A dinamikus mutatók hatása a modellekre</i>	110
6.2.2. <i>A dinamikus modellek értékelése</i>	116
7. A KUTATÁSOK FOLYTATÁSÁNAK LEHETSÉGES IRÁNYAI	121
8. ÖSSZEFOGLALÁS	124
9. FELHASZNÁLT IRODALOM	127
10. MELLÉKLETEK.....	142

ÁBRAJEGYZÉK

1. ábra: Virág-Kristóf döntési fa modellje.....	39
2. ábra: A csődmodellek fejlődéstörténete módszertani megközelítéssel	42
3. ábra: Egy tipikus döntési fa	51
4. ábra: Neurális háló két köztes réteggel.....	54
5. ábra: Receiver Operating Characteristic (ROC) grafikon	60
6. ábra: Néhány lehetséges ROC-görbe.....	61
7. ábra: A CAP-görbe és a Gini-koefficiens.....	63
8. ábra: A Kolmogorov-Smirnov (K-S) teszt	65
9. ábra: Vállalatok pénzügyi „egészségi állapotának” romlása	69
10. ábra: A fejlesztési minta és a nemzetgazdaság megoszlása árbevétel alapján	72
11. ábra: A fejlesztési minta és a nemzetgazdaság megoszlása gazdasági ágazatok alapján	72
12. ábra: Az előzetesen kiválasztott mutatók átlagai és mediánjai I.	82
13. ábra: Az előzetesen kiválasztott mutatók átlagai és mediánjai II.	82
14. ábra: A modellfejlesztés folyamatábrája	85
15. ábra: Az Enterprise Miner Data Partition Node paraméterezése.....	87
16. ábra: A Decision Tree Node paraméterezése	89
17. ábra: A Regression Node paraméterezése	91
18. ábra: A Neural Network Node paraméterezése	93
19. ábra: A Model Comparison Node paraméterezése	94
20. ábra: A kidolgozott modellek ROC-görbéi	104
21. ábra: A dinamikus default-modellek ROC-görbéi.....	118

TÁBLÁZATOK JEGYZÉKE

1. táblázat: Altman Z-modelljének változói és csoportokon belüli átlagaik	7
2. táblázat: Altman Z-modelljében szereplő változók magyarázó ereje.....	8
3. táblázat: Altman Z-modelljének besorolási pontossága	9
4. táblázat: A ZETA-modell hibaarányai	12
5. táblázat: Altman modellje és a ZETA-modell pontossága	13
6. táblázat: Ohlson modelljének besorolási pontossága	16
7. táblázat: Az első magyar csődmódel kialakításánál használt pénzügyi mutatók	31
8. táblázat: Virág-Kristóf neurális hálóinak összefoglaló táblázata	36
9. táblázat: Virág-Kristóf iparági rátákon alapuló csődmódeljeinek besorolási pontosságai.....	40
10. táblázat: Az egyes modellezési módszerek összehasonlító táblázata	57
11. táblázat: Klasszifikációs mátrix csődmódelre értelmezve	58
12. táblázat: A csődmódellek kritikái és a javasolt modell megfelelő megoldásai	70
13. táblázat: A vizsgált likviditási mutatók	75
14. táblázat: A vizsgált tőkeszerkezeti mutatók	75
15. táblázat: A vizsgált adósságszolgálati mutatók	76
16. táblázat: A vizsgált működőtőke mutatók	76
17. táblázat: A vizsgált eszközforgási mutatók	77
18. táblázat: A vizsgált termelékenység mutatók	77
19. táblázat: A vizsgált jövedelmezőség mutatók	79
20. táblázat: A leíró statisztika alapján legjobb szétválasztó erejű mutatók.....	81
21. táblázat: A kidolgozott default-módellek által felhasznált mutatók	99
22. táblázat: A kidolgozott módellek besorolási pontosságai és hibaarányai.....	102
23. táblázat: A kidolgozott default-módellek teljesítménymutatói	105
24. táblázat: Az iparági ráták tesztelése során kapott módellek besorolási pontosságai és hibaarányai	108
25. táblázat: Az iparági ráták második tesztelése során kapott módellek besorolási pontosságai és hibaarányai	109
26. táblázat: A dinamikus default-módellek által felhasznált mutatók.....	115
27. táblázat: A dinamikus mutatók tesztelése során kapott módellek besorolási pontosságai és hibaarányai	117
28. táblázat: A dinamikus default-módellek teljesítménymutatói	119

MELLÉKLETEK

1. melléklet: A csődmodell-kutatások szerzőinek és módszereinek összefoglaló táblázata..... 142
2. melléklet: Az egyes pénzügyi mutatók főbb leíró statisztikai fizetési késedelembe esett és kontroll vállalatokra 145

1. Bevezetés, célkitűzés

Az értekezés az előrejelzési modellek témakörében folytatott kutatásaim összefoglalása. A témaválasztás oka nemcsak személyes érdeklődésből fakad, hanem szakmám jellegéből is: immár nyolcadik éve foglalkozom banki kockázatkezeléssel, azon belül hitelkockázatokkal.

Ezen a szakterületen elkerülhetetlen, hogy a hitelezési döntések során előrejelzéseket alkalmazzunk: minden egyes hitelkérelem esetén mérlegelni kell, hogy vajon vissza fogja-e fizetni az ügyfél a meghitelezett összeget és a bank bevételét jelentő kamatot, vagy inkább fennáll annak a kockázata, hogy a nemfizetés miatt az ügylet veszteséget hoz majd a bank számára.

A szakirodalomban a pénzügyi előrejelzési modellek egyik legelterjedtebb típusai a csődmodellek, elsősorban tehát ezeket tekintettem kiindulási alapnak. Azt tanulmányoztam, hogy milyen statisztikai módszertan alkalmazása, milyen adatok illetve definíciók felhasználása teszi lehetővé e prediktív modellek kialakítását.

A vállalati hitelezéssel foglalkozó bank számára rendkívül hasznosak a csődmodellek, alkalmazzák ezeket a gyakorlatban is. Ugyanakkor a bank pénzügyi veszteségei nemcsak egy csődbe jutott vállalat esetében lehetnek jelentősek, hanem csődöt egyébként nem jelentő, ugyanakkor komolyabb fizetési késedelembe eső cég is problémát jelent.

Ehhez járult még hozzá a Bázel II. szabályozás életbe lépése, amely - többek között - a hitelkockázatok előrejelzését és számszerűsítését követeli meg, a default bekövetkezését a 90 napos fizetési késedelemnél határozva meg.

A csődmodellek mintájára tehát olyan modell kialakítását tűztem ki célul, amely a csődnél korábban, 90 napos fizetési késedelemnél figyelmeztet a fennálló kockázatra.

Hosszas keresés után sikerült olyan adatbázishoz jutnom, amelynek felhasználásával országos mintán tudtam elvégezni a tesztek, összesen 2000 vállalat öt éves időszakot felölelő, 2002-2006 közötti adatainak felhasználásával. Annak érdekében, hogy a kidolgozott modellek minél szélesebb körben legyenek alkalmasak az előrejelzésre, a

minta két szempont alapján képezi le a hazai nemzetgazdaságot: a vállalatok árbevétele és iparági megoszlásuk alapján. A default esemény bekövetkezése szintén országos adatbázisból, a BISZ Rt. Központi Hitelinformációs Rendszeréből származik.

Értekezésem első részében a csődmodellek fejlődéstörténetét kísérem meg átfogóan bemutatni, összesen 126 modell áttekintésével, elsősorban az alkalmazott módszertan szerinti megközelítéssel.

Az értekezés második részében a csődmodellekhez hasonló előrejelzési modellt próbálok kialakítani, amely a Bázeli II. definícióknak megfelelően a vállalatok nemfizetését, 90 napos késedelembe esését jelzi előre.

A modellezés során a csődmodelleknél alkalmazott statisztikai módszereket próbálom alkalmazni, és megkísérlem összehasonlítani az egyes modellezési technikák előrejelző képességét, besorolási pontosságát, kiválasztani a legjobban teljesítő modellezési módszert, illetve meghatározni az előrejelzésre leghatékonyabban alkalmazható mutatók körét.

Az adatbázis segítségével két további hipotézist is megvizsgállok:

- Növelhető-e a modellek besorolási pontossága, ha iparági középértékekhez viszonyított pénzügyi mutatókat vonunk be a vizsgálandó változók körébe?
- Növelhető-e a modellek pontossága dinamikus mutatók felhasználásával?

2. A csődmodellek történeti áttekintése

A csődelőrejelzés, mint kutatási terület igen fiatal. Kialakulása az 1930-as évekre nyúlik vissza, de a mai értelemben vett csődmodellek csak a 60-as évek végére jelentek meg. Az első modellek óta eltelt közel 40 év alatt az alkalmazott módszertan jelentős fejlődésen ment keresztül.

Az alábbiakban a csődmodellek történetét fogom átfogóan bemutatni, az alkalmazott módszertan szerinti megközelítéssel. Az egyes módszereknél a szakirodalomban legtöbbször idézett, illetve legjelentősebbnek tartott modelleket részletesen ismertetem, de a többi fellelt modellt is besorolom a megfelelő módszertani részhez.

A fontosabb modelleket az alábbi szempontok alapján ismertetem:

- a modellek alapját képező adatminta: az elemzésekben részt vevő vállalatok jellegzetességei, a minta összetétele,
- az alkalmazott módszertan,
- a modell által a csődveszély előrejelzése szempontjából relevánsnak ítélt pénzügyi mutatók,
- a modell „erőssége”, azaz besorolási pontossága.

A szakirodalom feldolgozása során összesen 126 csődmodell-kutatást vizsgáltam, amelyekre épülve a fejezet végén megkísérlem rendszerezni, összefoglalni a csődmodellek irodalmát és történetét, illetve meghatározni és behatárolni az egyes fejlődéstörténeti szakaszokat.

Előljáróban az alábbi hipotézis fogalmazható meg:

H1. A csődmodellek fejlődéstörténete viszonylag egyenes úton haladt. Módszertani szempontból az első, egyváltozós elemzéseken alapuló modelleket hamarosan többváltozós elemzésekkel készült modellek követték. Napjainkra a mesterséges intelligencia, illetve a szakértő által kontrollált módszerek terjedtek el, és ezek tekinthetők a legfejlettebb technikáknak.

2.1. A kezdetek: pénzügyi mutatók, mint figyelmeztető jelek

2.1.1. *Ramser, Foster, Fitzpatrick, Winakor, Smith, Merwin*

A kezdeti kutatások még nem tekinthetők valódi modellépítésnek. Az elemzések mindössze olyan mutatók kialakítására, illetve kiválasztására koncentráltak, amelyek átlagos értékei szignifikánsan különböznek a csődbe jutott és a működő vállalatok csoportjai esetén.¹

Az első előfutárok Ramster és Foster, valamint Fitzpatrick voltak, 1931-ben illetve 1932-ben publikált tanulmányaikkal. Winakor és Smith 1935-ben, Merwin pedig 1942-ben jelentette meg kutatási eredményeit.²

A legtöbb szerző által kiválasztott csődelőrejelző faktor, amelynek tekintetében jelentős különbség mutatkozott a csődbe ment és a túlélő vállalatok tekintetében, a likviditási mutató volt. A tanulmányokban megjelenő egyéb mutatók³:

- jegyzett tőke / nettó árbevétel,
- jegyzett tőke / befektetett eszközök,
- összes adósság / jegyzett tőke,
- forgótőke / összes eszköz.

A későbbi kutatások szempontjából e kezdeti korszak legjelentősebb felfedezésének azt tekinthetjük, hogy a kutatók a pénzügyi mutatók rendszerét előrejelző eszközként, figyelmeztető jelekként kezelték, és ezek segítségével próbálták értékelni a vállalatok teljesítményét.

¹ Virág (2004)

² A kutatási terület hamar népszerű lett. Cikkek egész sorozata jelent meg, a szerzők egymás munkáira reflektáltak, kiegészítették egymást, vitatkoztak. Lásd pl. Mears (1966), Horrigan (1966), Neter (1966), Wilcox (1971), Wilcox (1973), Libby (1975).

³ Back-Laitinen-Sere-Wezel (1996)

2.1.2. Beaver egyváltozós elemzése

A szakirodalom a csődelőrejelzés első előfutárának W.H. Beavert tekinti. Beaver tanulmányában⁴ egyetlen olyan mutatót próbált definiálni, amelynek felhasználásával szét lehet választani a csődbe jutott és a túlélő vállalatokat.

Az elemzéshez 79 csődbe ment illetve 79 túlélő vállalat adatait használta fel. A vállalatok éves beszámolóiból 30 mutatót képzett, amelyeket egyesével hasonlított össze a két csoportra vonatkozóan. Arra a következtetésre jutott, hogy a csoportok adatai leginkább az alábbi mutatók terén különböztek egymástól:

- cash-flow / összes eszköz,
- cash-flow / adósság,
- nettó árbevétel / adósság.

Mindegyik viszonyszámra meghatározott egy kritikus értéket (cut-off-ot), amely fölött a vállalatot túlélőnek, alatta pedig csődveszélyesnek ítélte.

A legjobb eredményt a cash-flow és az összes eszköz arányának mutatójával érte el, amely 90 százalékos megbízhatósággal mutatta meg a fizetéseképtelenséget egy évvel a csőd bekövetkezése előtt. Modellje hosszú távon is hatékonynak bizonyult: a cash-flow és adósság arányát figyelembe véve még öt évvel a csőd előtt is mindössze 22% volt a rossz besorolások aránya.

Bár Beaver modellje kezdetlegesnek tűnik a mai fejlett módszertanokhoz képest, jelentősége vitathatatlan. Ő volt ugyanis az első, aki a viszonyszám-analízist a csődelőrejelző módszerek sorába emelte, ezzel alapvetően meghatározva a későbbi csődmodell-kutatások irányvonalát.

A modell alkalmazása ugyanakkor számos problémát vet fel, amelyek mind az egyváltozós jellegéből fakadnak. Például nem veszi figyelembe, hogy a csődöt több mutató „együttállása” is jelezheti, és nem számol a mutatók közötti korrelációval sem.

⁴ Beaver (1966)

Könnyen előfordulhat továbbá, hogy az egyik mutató szerint csődvészélyesnek kellene tekinteni a vizsgált vállalatot, míg egy másik mutató szerint nem. Ezzel a helyzettel Beaver módszere nem tud mit kezdeni, egyetlen univerzális mutató pedig, amellyel minden vállalatról eldönthető lenne a továbbélés esélye, aligha adható.⁵

2.2. Az úttörők: Altman és társai

A csődmodell-kutatásokban az áttörést és a többváltozós módszerek térhódítását a diszkriminancia analízis megjelenése hozta meg.

2.2.1. Altman Z-modellje

A csődmodellek történetének első igazi úttörője Edward I. Altman volt. 1968-as, diszkriminancia analízissel készült modellje már valódi többváltozós elemzés volt, és jelentős eredményt ért el.

Összesen 66 közepes méretű iparvállalatot elemzett, amelyből 33 csődbe ment az 1946-1965 közötti időszakban, 33 pedig még 1966-ban is működött. A mintában szereplő vállalatokra 22 lehetséges pénzügyi mutatószámot vizsgált meg a likviditási, megtérülési, tőkeáttételi, eszköz-megfelelőségi és eszköz-kihasználtsági mutatócsoportokból. A mutatókat a szakirodalomban való népszerűségük és potenciális jelentőségük szerint választotta ki.

Végeredményként Altman egy öt változót tartalmazó lineáris függvényt állított fel:

$$Z = 0,012 X_1 + 0,014 X_2 + 0,033 X_3 + 0,006 X_4 + 0,999 X_5$$

Az egyenletben szereplő változók a következők:

- X_1 : forgótőke / összes eszköz,
- X_2 : visszatartott nyereség / összes eszköz,
- X_3 : adózás és kamatfizetés előtti nyereség (EBIT) / összes eszköz,

⁵ Virág (2004)

- X_4 : részvények piaci értéke / adósság,
- X_5 : nettó árbevétel / összes eszköz.

A változók a hozzájuk kiszámított súlyok segítségével adják azt a csődfüggvényt, amely lehetővé tette a cégek megbízható besorolását a csődveszélyeztetettség szempontjából.

Az értékelés úgy történik, hogy a mutatószámok értéke a súlyokkal megszorozva és összeadva ad egy Z értéket. Ezt a Z értéket összehasonlítjuk a cut-off ponttal, és ez alapján lehetőség nyílik az adott cég valamely csoporthoz történő besorolására:

- amennyiben a kiszámított Z érték 1,81 alatti, a céget egyértelműen a csődveszélyes kategóriába sorolhatjuk;
- a legalább 2,99-es Z értéket elérő vállalatokat egyértelműen túlélőnek tekinthetjük;
- a két érték között található tartományban a téves besorolások aránya magas, így ebben a „szürke zónában” nagyobb körültekintéssel kell eljárni és a modell eredményét fenntartásokkal kell kezelni.

További vizsgálatok alapján az egységes cut-off értékét Altman 2,675-nél húzta meg. Ez tehát az a kritikus Z érték, amelynél a modell szétválasztja a csődbe ment és a túlélő vállalatokat.

Tekintve, hogy mind az öt változó negatív kapcsolatban áll a csőddel, minél nagyobb egy vállalat Z értéke, annál kisebb a csőd valószínűsége. Látható volt továbbá az is, hogy a változóknak a túlélő csoporton belüli átlagai mindig meghaladják a csődbe jutott csoport ugyanezen mutatóinak értékeit.

Változó	Csődbe jutott vállalatok átlaga	Túlélő vállalatok átlaga
X_1	-6,1	41,4
X_2	-62,6	35,5
X_3	-31,8	15,4
X_4	40,1	247,7
X_5	1,5	1,9

1. táblázat: Altman Z-modelljének változói és csoportokon belüli átlagai

Mivel a modell függvényében az X_4 mutató tartalmazza a saját részvények piaci értékét, a megalkotott csődfüggvény csak a tőzsdén jegyzett részvényekkel rendelkező vállalatokra alkalmazható.

A modell egy érdekes pontja az árbevétel és eszközök aránya (X_5) kérdése: Altman megjegyzi, hogy ez egy szeparált szignifikancia-vizsgálat esetén be sem került volna a modellbe.

Az egyes változóknak a csődfüggvény szétválasztó erejéhez való hozzájárulását a statisztikai F-próbával mérte, egyenként és együttesen is. Az F-próba alapján a változók magyarázó ereje a következőképp alakult:

Változó	F-arány
X_1	32,6
X_2	58,86
X_3	26,56
X_4	33,26
X_5	2,84

2. táblázat: Altman Z-modelljében szereplő változók magyarázó ereje

Az X_1 , X_2 , X_3 és X_4 változók szignifikánsnak mutatkoztak 1%-os szignifikancia-szinten is, kivéve az ötödik változót. A változó ennek ellenére benne maradt a végső modellben, amit a többi változóval való erős korrelációja indokolt.

A modell besorolási pontossága magas, 95%-os a csődöt megelőző évben. Ez azonban a csőd időpontjától távolodva fokozatosan csökken:

Évek a csőd bekövetkezése előtt	Vizsgált vállalatok száma	Helyes besorolás	Hibás besorolás	Besorolási pontosság
1 év	33	31	2	95%
2 év	32	23	9	72%
3 év	29	14	15	48%
4 év	28	8	20	29%
5 év	25	9	16	36%

3. táblázat: Altman Z-modelljének besorolási pontossága

2.2.2. Deakin

Az Altman által bevezetett diszkriminancia analízis segítségével dolgozta ki csődmodelljét Deakin is, 1972-ben. Összesen 64 cég adatait használta fel, amelyek fele-fele arányban voltak fizetésképtelenek illetve túlélők. A csőd bekövetkezését megelőző öt év adatait vizsgálva alakított ki csődelőrejelző modellt, amely az alábbi mutatószámokra épült:

- készpénz / rövid lejáratú kötelezettségek,
- cash flow / összes tartozás,
- készpénz / nettó értékesítés,
- készpénz / összes eszköz,
- likviditási ráta,
- forgóeszközök / nettó értékesítés,
- forgóeszközök / összes eszköz,
- nettó árbevétel / összes eszköz,
- likvid eszközök / rövid lejáratú kötelezettségek,
- likvid eszközök / nettó értékesítés,
- likvid eszközök / összes eszköz,

- összes tartozás / összes eszköz,
- forgótőke / nettó értékesítés,
- forgótőke / összes eszköz.

A modell előrejelzési pontosságát egy 11 csődbe jutott, illetve 23 túlélő vállalatból álló hold-out mintán is tesztelte. A modell igen jó eredményeket mutatott, besorolási pontossága 97%-os volt.

2.2.3. Blum

1974-ben Blum is a diszkriminancia analízis felhasználásával dolgozta ki előrejelző modelljét, az alábbi mutatókra támaszkodva:

- piaci megtérülési ráta,
- likviditási gyorsráta,
- cash flow / összes tartozás,
- nettó vagyon méltányos piaci értéke,
- nettó likvid eszközök / készletek,
- nettó vagyon könyv szerinti értéke / összes tartozás,
- bevételek szórása,
- nettó likvid eszközök szórása / készletek,
- bevételek trendje,
- nettó likvid eszközök trendje / készletek,
- a bevétel trendjének törése,
- nettó likvid eszközök trendjének törése / készletek.

Modellje igen pontos volt: több mint 90%-os pontossággal volt képes előre jelezni a vállalatok csődbe jutását.

2.2.4. Altman-Lorris

Altman és Lorris 1976-ban publikált közös modelljének alapja is a diszkriminancia analízis volt.

A modell 5 pénzügyi mutató felhasználásával készült:

- nettó árbevétel / összes eszköz,
- (összes kötelezettség + alárendelt hitelek) / saját tőke,
- összes eszköz / nettó szavatoló tőke,
- (záró tőke - tőkenövelések) / kezdő tőke,
- a fenti mutatók kombinációjaként előállított változó.

Modelljük teljesítménye meglehetősen magas volt, összességében 90%-os besorolási pontosságot ért el.

2.2.5. A ZETA-modell

Körülbelül egy évtizeddel az áttörést jelentő modelljének publikálása után, 1977-ben alkotta meg Altman az eredeti modell továbbfejlesztett változatát Haldemannal és Narayanannal. A továbbfejlesztést több ok miatt is szükségessé vált:

- Egyrészt megváltozott a gazdasági környezet: egyre több nagyvállalati csődöt is regisztráltak, így szükségessé vált az eredetileg középvállalatokra épülő modell kiterjesztése a nagyvállalatokra is.
- Hasonló okból ki kellett terjeszteni a vizsgált vállalatok körét az iparvállalatok mellett a kereskedelmi vállalatokra is.
- A fentiekén túlmenően igazodni kellett a számviteli előírásokban időközben bekövetkezett változásokhoz.

Az új modellt 111 vállalatból álló mintán dolgozták ki: 53 csődbe ment és 58 túlélő vállalatra. A mintában fele-fele arányban szerepeltek kereskedelmi illetve iparvállalatok. A szerzők kiterjesztették a vizsgált mutatók körét is: az eredeti kutatás 22

mutatójával szemben 27 mutatót képeztek a vállalatok éves beszámolóí alapján, és ezek segítségével állították fel modelljüket.

A ZETA-modell összesen hét pénzügyi mutatót tartalmazott, amelyeket a modellben alkalmazott súlyaik nélkül publikáltak. A mutatók:

- X_1 : eszközjövodelmezőség: EBIT / eszközök,
- X_2 : a jövodelmezőség stabilitása: a 10 éves trendtől való eltérés,
- X_3 : adósságszolgálat: EBIT / kamatfizetés logaritmusa,
- X_4 : kumulatív jövodelmezőség: visszatartott nyereség / eszközök,
- X_5 : likviditás: forgóeszközök / rövid lejáratú kötelezettségek,
- X_6 : kapitalizáció: jegyzett tőke / saját tőke,
- X_7 : méret: összes eszköz logaritmusa.

A modellben a legerősebb három mutató sorrendben a kumulatív jövodelmezőség (X_4), a jövodelmezőség stabilitása (X_2) és a kapitalizáció (X_6), míg a legkisebb súlyt az eszközjövodelmezőség (X_1) mutató kapta.

Az alábbi táblázat mutatja a modell hibaarányait a csőd bekövetkezése előtti évekre:

Időhorizont	Elsőfajú hiba	Másodfajú hiba	Összes hiba
1 év	3,80%	10,30%	7,20%
2 év	15,10%	6,90%	10,90%
3 év	25,50%	8,60%	16,70%
4 év	31,90%	10,50%	20,70%
5 év	30,20%	17,90%	23,40%

4. táblázat: A ZETA-modell hibaarányai

Ha ezeket az adatokat összevetjük Altman eredeti modelljével, a következőt láthatjuk:

Évek száma	Altman modellje	ZETA-modell
1 év	95%	92,80%
2 év	72%	89,10%
3 év	48%	83,30%
4 év	29%	79,30%
5 év	36%	76,60%

5. táblázat: Altman modellje és a ZETA-modell pontossága

Jól látható, hogy rövid távú előrejelzés szempontjából a ZETA-modell körülbelül ugyanolyan hatékonysággal dolgozik, mint 1968-as elődje: mindkét modell 93-95%-ban ad helyes minősítést a csődöt megelőző évben. A ZETA-modell igazi jelentőségét azonban a hosszú távú előrejelző képessége adta. Még három évvel a bekövetkezés előtt is 80% körüli pontossággal volt képes előre jelezni a csődöt, miközben ilyen időtávra az eredeti modell megbízhatósága már az 50%-os szint alá süllyed. Sőt, ez a teljesítmény még a csődöt megelőző ötödik évre is csak 77%-ra csökken.

A ZETA-modellnél a cut-off értékét 0-nál húzták meg, vagyis a negatív értékkel rendelkező vállalatokat csődveszélyesnek, a pozitív értékűeket pedig túlélőnek minősíti a modell. Hasonlóan az eredeti modellhez, itt is meghatározható egy „szürke tartomány”, ahol a modell értékelése bizonytalan: -1,45 és +0,87 közötti értékek esetén kell körültekintően kezelni a modell által jelzett besorolást.

A modell megalkotói új elemet is bevettek a kutatásba: az a priori valószínűségeket és a költségeket. Ezekkel az eredeti modell nem számolt. A vizsgálatok célja az volt, hogy a döntéshozók saját tapasztalataik és döntési környezetük alapján rugalmasan állíthassák be a cut-off értékét.

A cut-off meghatározása az alábbi képlet alapján lehetséges:

$$\ln \frac{q_1 C_1}{q_2 C_2}$$

ahol:

q_1 és q_2 a csődbe jutott illetve a túlélő csoportba sorolás a priori valószínűségeit,
 C_1 és C_2 az első illetve másodfajú hibák költségeit jelentik.

A költségek figyelembe vételét az indokolja, hogy gazdaságilag más-más hatása van a modell által vétett hibáknak. Míg egy tévesen elfogadott kérelem esetén egy később csődbe menő vállalatnak nyújtott hitelen tényleges vesztesége képződik a banknak, addig egy téves döntés alapján elutasított kérelemnél a döntésnek „csak” alternatív költsége, elmaradt haszna van.

A fenti két tényező modellbe iktatása ugyan gazdaságilag indokolható és elvileg növeli a modell alkalmazhatóságának rugalmasságát, ugyanakkor gyakorlatban nem tudták igazolni. A szerzők a különböző q és C értékek vizsgálatával arra a végkövetkeztetésre jutottak, hogy az eredeti modell mindezek ellenére jobb eredményeket mutat.

2.2.6. További modellek

Altman korszakalkotó modellje kutatások egész sorát indította el: az 1970-es években sokan tesztelték a diszkriminancia analízis felhasználását a csőd előrejelzésére. Érdemes megemlíteni az alábbi tanulmányokat: Edmister (1972), Lane (1972), Bilderbeek (1973), Awh – Waters (1974), Elam (1975), Sinkey (1975), Taffler - Tisshaw (1977), van Frederikslust (1978), Ketz (1978), Norton - Smith (1979), Bilderbeek (1979), Altman – Levallee (1980), Dambolena - Khoury (1980), Sharma - Mahajan (1980).

A kísérletek még a 80-as évekre is áthúzódtak, igaz, ekkor már jóval kevesebb modell alkalmazta a diszkriminancia analízist: pl. Castagna – Matolcsy (1981), Ooghe - Verbaere (1982), El-Hennawy – Morris (1983), Taffler (1983), Altman – Izan (1984), Micha (1984), Lincoln (1984), Takahashi - Kurokawa (1984), Levitan – Knoblett (1985), Betts - Belhoul (1987), Gombola – Haskins – Ketz - Williams (1987), Dambolena – Shulman (1988).

De előfordulnak diszkriminancia analízison alapuló modellek még jóval Altman után is: pl. Lussier (1995) vagy McGurr - Devaney (1998).

2.3. A regressziós modellek térhódítása

2.3.1. Ohlson

A logisztikus regressziót először Ohlson emelte a csődelőrejelző modellek kelléktárába 1980-as publikációjában. Ezzel körülbelül egy évtizedre meghatározta a kutatások irányvonalát.

Ohlson modellje az 1970-76 közötti időszak adataira épült. A vizsgált minta összesen 2163 amerikai iparvállalatból állt, amelyekből 105 csődbe ment, 2058 pedig túlélő volt.

Összesen 9 mutatót vizsgált, amelyeket a szakirodalom feldolgozása alapján állított össze, mint leggyakrabban előforduló mutatókat.

A logisztikus regressziós modell végeredménye a következő volt:

$$Z = -1,32 - 0,407 X_1 + 6,03 X_2 - 1,43 X_3 + 0,076 X_4 - 2,37 X_5 - 1,83 X_6 + 0,285 X_7 - 1,72 X_8 - 0,521 X_9$$

A modellben szereplő mutatók:

- X_1 : log(összes eszköz / GNP árszínvonal index),
- X_2 : kötelezettségek / összes eszköz,
- X_3 : forgótőke / összes eszköz,
- X_4 : rövid lejáratú kötelezettségek / forgóeszközök,
- X_5 : dummy változó. Értéke 1, ha a teljes tartozás meghaladja az összes eszközt, egyébként 0;
- X_6 : nettó árbevétel / összes eszköz,
- X_7 : üzleti eredmény / kötelezettségek,

- X_8 : dummy változó. Értéke 1, ha a nettó árbevétel negatív volt az elmúlt két évben, egyébként 0;
- X_9 : nettó árbevétel változása: $(NI_t - NI_{t-1}) / (|NI_t| + |NI_{t-1}|)$, ahol NI a nettó árbevétel.

A cut-off értéke Ohlsonnál 0,038. Vagyis a 0,038-as érték feletti P értékkel rendelkező vállalatokat a modell csődvészélyesnek ítéli meg, ez alatt az érték alatt a túlélő kategóriába sorolja.

A modell pontossága kb. 83%-os volt. Besorolási hibáit az alábbi táblázat mutatja.

	Elsőfajú hiba	Másodfajú hiba	Helyes besorolás
<i>Vállalatok száma</i>	13	358	1792
<i>Vállalatok aránya (%)</i>	12,4%	17,4%	82,9%

6. táblázat: Ohlson modelljének besorolási pontossága

2.3.2. Zmijewski

A probit modellt először Zmijewski alkalmazta csődelőrejelzésre 1984-es tanulmányában.

Összesen 840 iparvállalatot vizsgált meg, 40 csődbe jutott céget és 800 túlélőt. A cégek adatait az 1970 és 1978 közötti időszakra dolgozta fel.

Modelljének egyenlete a következő:

$$P = -4,3 - 4,5 X_1 + 5,7 X_2 - 0,004 X_3$$

A modellben szereplő mutatók:

- X_1 : nettó árbevétel / összes eszköz,
- X_2 : összes tartozás / összes eszköz,
- X_3 : rövid lejáratú kötelezettségek / forgóeszközök.

Zmijewski a modell számára ismeretlen, 25%-os ún. hold-out mintán is tesztelte a modell megbízhatóságát. Ez azt jelenti, hogy a modellépítés során a minta egy részét kihagyta az elemzésből, majd a tanulási mintán kifejlesztett modellt lefuttatta a kimaradó részmintára is, így mérve a modell teljesítményét. Ezen a mintán 98%-os besorolási pontosságot mutatott ki, amely magas fokú stabilitást jelez.

Zmijewski további tesztek is elvégzett. Azt vizsgálta, hogy mennyire torzítja a csődelőrejelzés becsléseit a minta összetétele, pontosabban a csődbe ment és a túlélő vállalatok mintabeli aránya. A szakirodalomban addig elterjedt 1:1 arány helyett különböző összetételű mintákat készített. Összesen hat új mintát vizsgált, amelyekben az eredeti 1:1 arányról egészen 1:20-ig növelte a mintába kerülő túlélő vállalatok arányát. Ez azonban a modellek besorolási pontosságában nem hozott szignifikáns különbségeket.

2.3.3. Két banki csődmodell: Martin, illetve Pantalone-Platt modelljei

Martin 1977-ben publikált tanulmánya a kereskedelmi bankok csődelőrejelzését vizsgálta. Összesen 5700 bank adatait tartalmazó mintán dolgozott, amelyből 58 ment csődbe az 1970-76 közötti időszakban.

Modelljének alapja a logit regresszió volt, és a következő mutatókat találta erős magyarázó erejűnek:

- nettó árbevétel / összes eszköz,
- bruttó leírások / nettó üzleti eredmény,
- kereskedelmi és ipari hitelek / összes hitel,
- bruttó tőke / kockázati eszközök.

Tíz évvel később, 1987-ben Pantalone és Platt is kidolgozott egy bankokra vonatkozó csődmodell. Elemzésük 226 túlélő és 113 fizetésképtelenné vált bankból álló mintán készült, a logisztikus regresszió felhasználásával.

Modelljük az alábbi mutatókat tartotta szignifikánsnak a csőd előrejelzése szempontjából:

- nettó árbevétel / összes eszköz,
- saját tőke / összes eszköz,
- összes hitel / összes eszköz,
- kereskedelmi és ipari hitelek / összes hitel,
- lakásépítések %-os változása.

Amint látható, az újabb modell magyarázó változói nagyban hasonlítanak az egy évtizeddel korábban készült modelléhez. A korábbi modell 4 mutatójából kettő az új modellben is ugyanúgy megjelenik. A későbbi modell érdekességét ugyanakkor az utolsó magyarázó változó adja: Pantalone és Platt ugyanis a pénzügyi beszámolók adataiból képzett viszonyszámok mellett beépített egy helyi gazdasági feltételekkel összefüggő mutatót is.

A Pantalone és Platt modell besorolási pontossága is meggyőző: 98%-ot ért el a csődveszély előrejelzésében, és 92%-ot a túlélőkre vonatkozóan. A modell összesített besorolási pontossága 96% volt.

2.3.4. Zavgren

Zavgren 1985-ös csődmodellje az Amerikai Egyesült Államok tőzsdén jegyzett vállalataira készült, az 1972-1978 közötti időszakra vonatkozóan. Összesen 90 vállalat adataiból dolgozott, amelyből 45 ment csődbe a vizsgált időszak alatt, és ugyanennyi működött tovább.

A logisztikus regresszió módszerével kialakított csődmodell az alábbi mutatókat tartotta szignifikánsnak a csődelőrejelzés szempontjából:

- készletek / értékesítés,
- lejárt követelések / készletek,
- (készpénz + rövid távú befektetések) / összes eszköz,

- likviditási gyorsráta,
- adózás utáni sajáttőke-arányos nyereség,
- (kötelezettségek – passzív időbeli elhatárolások) / saját tőke,
- értékesítés / összes eszköz.

A modell besorolási pontossága egy évvel a csőd bekövetkezése előtt 82%-os volt.

2.3.5. Keasey-McGuinness

Keasey és McGuinness csődmmodelljük⁶ kialakításához az Egyesült Királyságban működő vállalatok adataiból vettek mintát. A minta fele-fele arányban tartalmazott fizetéseképtelenné vált illetve túlélő vállalatot, összesen 116 cég adatai szerepeltek benne.

Az elemzéseket a csődöt megelőző 1-5 évre készítették el, modellezési módszerük a logisztikus regresszió volt. A tanulmányban vizsgált időszak a 1976-1984 közötti évekre terjedt ki.

A csődöt egy évvel előrejelző modell az alábbi mutatókat használta:

- saját tőke / kötelezettségek,
- beszerzés / szállítók,
- adózás előtti eredmény / értékesítés.

A csődöt megelőző évben modelljük előrejelzési pontossága 86%-os volt.

2.3.6. Rudolph-Hamdan

Rudolph és Hamdan modelljének⁷ különlegességét is a vizsgálat alá vont gazdálkodási egységek köre adja: a szerzőpáros ugyanis a takarékszövetkezeti csődöket vizsgálta. 1988-ban publikált elemzésük az 1984-85 közötti időszak takarékszövetkezeti csődjeit helyezte górcső alá, eszköze a logit módszer volt.

⁶ Keasey-McGuinness (1990)

⁷ Rudolph – Hamdan (1988)

A modellépítés során összesen 18 pénzügyi mutatót vizsgáltak meg, a likviditási, jövedelmezőségi, tőkemegfelelési és betétszerkezeti, eszközminőségi mutatócsoportokból. Logit modelljük a csőd előrejelzése szempontjából a tőkemegfelelési, betétszerkezeti és hitelkockázati mutatókat találta erős magyarázó erejű változóknak.

2.3.7. További modellek

A regressziós csődmodellek hamar elterjedtek a szakirodalomban, és ugyanolyan széles körű irodalom Ohlson nyomán ugyanolyan széles körű irodalom keletkezett, mint annak idején Altman publikációja után.

A legtöbb szerző a logisztikus regressziót alkalmazta. Ezek között érdemes kiemelni az alábbiakat: Dambolena - Shulman (1980), Barth – Brumbaugh – Sauerhaft – Wang (1985), Hopwood - McKeown - Mutchler (1989), Gilbert - Menon - Schwartz (1990), Platt - Platt (1990), Flagg - Giroux (1991), Ooghe - Joos - De Vos (1991), Bahnson - Bartley (1992), Berg – Hexeberg (1994), Ward (1994), Sheppard - Fraser (1994), Platt - Platt - Pedersen (1994), Johnsen - Melicher (1994), Hill - Perry (1996), Richardson - Kane - Lobinger (1998), Hunter - Isachenkova (2001), Hámori (2001), Darayseh - Waples - Tsoukalas (2003), Neophytou - Molinero (2004), Andersen (2008).

A legtöbb tanulmányban azonban a logisztikus regresszió nem csupán önmagában szerepel, hanem a publikáció lényeges eleme, hogy az eredményeket összevetik az ugyanarra az adatbázisra épülő diszkriminancia analízisen alapuló modellel. A legtöbb szerző a regresszió mellett teszi le a voksát: pl. Collins - Green (1982), Hamer (1983), Gentry - Newbold - Whitford (1985), Casey - Bartczak (1985), Lo (1986), Peel - Peel (1987), Aziz – Emanuel – Lawson (1988), Lau (1987), Keasey - Watson (1987), Gloubos - Grammatikos (1988), Barniv - Raveh (1989), Virág - Hajdu (1993), Lussier (1995), Lennox (1999), Ginoglou - Agorastos - Hatzigagios (2002).

A probit regresszió módszerét már jóval kevesebben alkalmazták: pl. Skogsvik (1990), Sjøvoll (1999), Bongini – Ferri – Hahm (2000), Bernhardsen (2001).

2.4. Modellek a rekurzív felosztás módszerével

2.4.1. *Frydman–Altman–Kao*

A 80-as évek közepén jelent meg a rekurzív felosztás (vagy más néven döntési fa, illetve rekurzív partícionáló algoritmus) a csődmodell-kutatásokban. Először 1985-ben, Altman, Frydman és Kao alkalmazta.

Az általuk vizsgált mutatók a következők voltak:

- nettó árbevétel / összes eszköz,
- forgóeszközök / rövid lejáratú kötelezettségek,
- $\log(\text{forgóeszközök})$,
- saját tőke piaci értéke / teljes tőkésítettség,
- forgóeszközök / összes eszköz,
- cash-flow / összes tartozás,
- likvid eszközök / rövid lejáratú kötelezettségek,
- EBIT / összes eszköz,
- $\log(\text{kamatfedezet}+15)$,
- készpénz / összes eszköz,
- összes tartozás / összes eszköz,
- likvid eszközök / összes eszköz.

A modell besorolási pontossága összességében 94%-os volt. Ez igen jónak mondható, azonban ha külön-külön nézzük az egyes csoportokba jutás előrejelzésének pontosságát, igen nagy különbséget látunk: míg a túlélők csoportjába kerülés esetén a besorolási pontosság 99%-os, addig ugyanez a csődveszélyes csoport esetén csak 84%.

2.4.2. McKee–Greenstein

A McKee–Greenstein szerzőpáros 2000-ben kísérletezett a rekurzív felosztás módszerével. Vizsgálataikat amerikai vállalatokból álló mintán végezték a csődöt megelőző évre vonatkoztatva.

Tanulmányukban a következő mutatók használták fel:

- nettó árbevétel / összes eszköz,
- forgóeszközök / összes eszköz,
- forgóeszközök / rövid lejáratú kötelezettségek,
- készpénz / összes eszköz,
- (forgóeszközök) / forgalom,
- hosszú lejáratú tartozások / összes eszköz.

Eredményeik azt mutatták, hogy a döntési fák másodfajú hibák tekintetében pontosabbak mind a diszkriminancia analízisnél, mind a neurális hálónál, elsőfajú hibák tekintetében viszont alulmaradnak.

Annak ellenére, hogy a rekurzív felosztás módszere egyszerűen kezelhető, eredménye pedig könnyen interpretálható és algoritmizálható, mégsem terjedt el a szakirodalomban.

2.5. A mesterséges intelligencia megjelenése: neurális hálók

2.5.1. Odom és Sharda

A neurális háló a 90-es évek elején jelent meg a csődelőrejelzés területén. A technika bevezetése Odom és Sharda nevéhez fűződik, akik 1990-ben publikálták tanulmányukat.

A szerzőpáros Altman 1968-as modelljének mutatóira alkalmazta a háromrétegű backpropagation-háló módszerét, figyelemreméltó eredménnyel. A tanulmány 74

vállalat 1975 és 1982 közötti adatait vizsgálta. A minta körülbelül fele-fele arányban tartalmazott csődbe jutott (38) és tovább működő (36) vállalatot.

A modell eredményeit a diszkriminancia analízisével vetették össze. A neurális háló nemcsak felülmúlta a diszkriminancia elemzést, hanem tökéletes besorolást sikerült vele megvalósítani, vagyis sem elsőfajú, sem másodfajú hiba nem fordult elő a tanulási mintán.

A kifejlesztett modellt a korábban már ismertetett hold-out mintán is tesztelték. 55 további, a modell számára ismeretlen vállalat mintáján is lefuttaták mindkét modellt. A hold-out minta – a tanulási mintához hasonlóan – szintén fele-fele arányban tartalmazott csődbe ment és túlélő vállalatot.

Az eredmények itt is a neurális hálót igazolták: a tesztelés során a neurális háló a 27 csődbe jutott vállalatból 82%-os pontossággal adott csődveszélyes minősítést, szemben a diszkriminancia analízis kb. 60%-os pontosságával.

A szerzők megállapították azt is, hogy a neurális háló nemcsak a tanulási és a hold-out mintán nyújtott pontosság terén előzi meg elődjét, hanem kismintás tesztelések során is jobb eredményeket mutat.

2.5.2. Coats és Fant

Tanulmányukban⁸ Coats és Fant is a klasszikus 1968-as Altman-modell változóit vizsgálta újra, az 1970 és 1989 közötti időszakra. Az elemzést elvégezték mind a diszkriminancia analízis, mind a neurális háló technikáival is, és összevetették a kettő eredményeit.

A besorolási pontosságok terén a neurális hálók egyértelműen jobban teljesítettek, mint a diszkriminancia analízis. Az egy éven belüli csőd előrejelzésénél pontosságuk 95%-os volt, szemben a diszkriminancia analízis által elért 88%-kal.

Az eltérés elsődleges oka az volt, hogy a diszkriminancia analízis különösen magas besorolási hibával volt alkalmazható a csődveszélyes csoport esetén.

⁸ Coats - Fant (1993)

2.5.3. *Back, Laitinen, Sere, Wesel*

Back és szerzőtársai, Laitinen, Sere és Wesel 1996-ban ismertette több módszert összehasonlító elemzésének eredményeit. A többrétegű neurális háló mellett tesztelték a diszkriminancia analízis és a logisztikus regresszió alapú modelleket is.

A szerzők által vizsgált mutatók az alábbiak voltak:

- készpénz / rövid lejáratú kötelezettségek,
- cash-flow / rövid lejáratú kötelezettségek,
- cash-flow / összes eszköz,
- cash-flow / összes tartozás,
- készpénz / nettó értékesítés,
- készpénz / összes eszköz,
- forgóeszközök / rövid lejáratú kötelezettségek,
- forgóeszközök / nettó értékesítés,
- forgóeszközök / összes eszköz,
- rövid lejáratú kötelezettségek / saját tőke,
- saját tőke / befektetett eszközök,
- saját tőke / nettó értékesítés,
- készletek / nettó értékesítés,
- hosszú lejáratú tartozások / saját tőke,
- saját tőke piaci értéke / tartozások könyv szerinti értéke,
- összes tartozás / saját tőke,
- nettó árbevétel / összes eszköz,
- nettó likvid eszközök / készletek,
- nettó értékesítés / összes eszköz,

- üzemi eredmény / összes eszköz,
- EBIT / összes fizetett kamat,
- likvid eszközök / rövid lejáratú kötelezettségek,
- likvid eszközök / nettó értékesítés,
- likvid eszközök / összes eszköz,
- megtérülési ráta / törzsrészesvények,
- fel nem osztott nyereség / összes eszköz,
- Return on Stock (ROS),
- összes tartozás / összes eszköz,
- forgótőke / nettó értékesítés,
- forgótőke / saját tőke,
- forgótőke / összes eszköz.

A szerzők megállapították, hogy a neurális háló segítségével felállított modell teljesítménye jelentősen felülmúlja a diszkriminancia analízis és a logisztikus regresszió eredményeit.

2.5.4. Olmeda – Fernandez

Olmeda és Fernandez 1997-ben publikálta kutatási eredményeit. Modelljük különlegességét az adja, hogy nem vállalati adatmintán dolgoztak, hanem spanyol bankok adatait dolgozták fel különböző időtávokra.

A szerzők módszertani összehasonlítást végeztek a neurális hálók, a diszkriminancia analízis, a logisztikus regresszió és kétféle döntési fa alkalmazásával kialakított csődmodellekre.

Eredményeik azt mutatták, hogy a neurális hálóval kb. 83%-os pontosságú modellt sikerült kifejleszteni, míg a többi módszerrel ennél kisebb, 62% és 79% közötti pontosságot értek el.

2.5.5. Kiviluoto

1998-ban Kiviluoto publikált olyan csődmódellet, amelyet Fisher-metrikára épülő önszerveződő térképek segítségével állított elő, több időhorizontra vonatkozóan. A módellet finnországi vállalatokból álló mintára alapozta, az alábbi mutatók felhasználásával:

- üzemi árás,
- rendkívüli tételek,
- NIBD (Net Income before Depreciation),
- előző év rendkívüli tételei,
- saját tőke arány.

Kiviluoto a módszer alkalmazásával 80% fölötti besorolási pontosságot ért el.

2.5.6. Zhang–Hu–Patuwo

Zhang, Hu és Patuwo 1999-ben publikálta tanulmányát, amelyben a logisztikus regressziót és a neurális hálókat hasonlították össze termelő vállalatokból álló adatbázison. Kutatásaikat ötrétegű hálóval, többszörös kölcsönhatás-vizsgálattal végezték el.

A neurális hálók input-neuronjaihoz az Altman-modell pénzügyi mutatóit használták fel, kiegészítve a forgóeszközök/rövid lejáratú kötelezettségek likviditási mutatóval.

A neurális háló eredményei jobbak voltak a logisztikus regressziónál: míg az előbbi 88%-os besorolási pontosságot mutatott, az utóbbi csak 79%-ot.

2.5.7. Laitinen – Kankaanpaa

Laitinen és Kankaanpaa 1999-ben készített tanulmányt. Összesen 76 finn vállalatot vizsgáltak meg, amelyekből 38 lett fizetésektelen, és ugyanennyi túlélő. Vizsgálataikat az 1986 – 1989 közötti időszakra végezték. A cégek többsége feldolgozóipari tevékenységet végzett, illetve a kis- és középvállalati méretbe tartozott.

A szerzők az összes modellt ugyanazon három mutató alapján építették fel, kizárva ezzel azt, hogy a modellek a módszertanon kívül más tekintetben is különbözzenek.

A felhasznált mutatók:

- összes hitel / összes eszköz
- likvid források / esedékes hitelek
- EBITDA / átlagos eszköz állomány

A tanulmányban hat módszert hasonlítottak össze: a diszkriminancia analízis, a logisztikus regresszió, döntési fa és a neurális háló mellett foglalkoznak még a túlélés elemzéssel és egy szakértői modellel is.

A szerzők megállapítják, hogy a tanuló minták eredményei alapján a neurális hálók és döntési fák mutatkoznak a legpontosabbnak, megelőzve az összes többi modellt. A neurális háló eredményei kiemelkedően magas, 97%-os pontosságot mutattak.

A hold-out mintán mért teljesítmények tekintetében azonban már nem volt ilyen markáns különbség a modellek besorolási pontosságai között. Az eredmények 86% és 89% között alakultak minden modell esetében.

2.5.8. További modellek

A neurális háló alapján alkotott modelleket is általában a szakirodalomban fellelhető korábbi módszerekkel vetették össze a szerzők.

Legtöbbször a neurális hálóval és az addig legsikeresebb regressziós modellekkel kísérleteztek, pl. Salchenberger – Cinar – Lash (1992), Fletcher - Goss (1993), Udo (1993), Leshno – Spector (1996), Piramuthu – Raghavan – Shaw (1998), Tan (1999), Yang – Temple (2000), Neophytou – Mar-Molinero (2001), Charitou – Neophytou - Charalambous (2004).

A diszkriminancia analízis, a regressziós modellek és a neurális hálók összehasonlító elemzését publikálták: pl. Tam – Kiang (1992), Alici (1995), Boritz - Kennedy - Albuquerque (1995), Pompe – Feelders (1997), Bardos – Zhu (1997), Lin - McClean (2001), Yang (2001).

Előfordulnak olyan kutatások is, amelyek a neurális hálók eredményeit csak a diszkriminancia analízissel készített modellek eredményeivel vetik össze: pl. Cadden (1991), Coats – Fant (1991), Coats – Fant (1993), Kerling–Poddig (1994), Wilson - Sharda (1994), Altman – Marco – Varetto (1994), Lacher - Coats - Sharma - Fant (1995), Gritta - Wang - Davalos - Chow (2000), Pompe – Bilderbeek (2000), Zapranis - Ginoglou (2000).

Viszonylag kevés tanulmány dolgozik kizárólag neurális hálókkal: pl. Chung – Tam (1993), Atiya (2001).

A neurális hálók modelljei általában jobb eredményeket mutatnak a két korábbi módszerekkel készített modelleknél. Meg kell jegyezni ugyanakkor, hogy a bonyolultabb szerkezet miatt ezeket a modelleket kevésbé részletesen publikálják szerzőik, megnehezítve ezáltal az eredmények érdemi követhetőségét és a rendszerező feldolgozást.

2.6. Magyarországi csődmodell-kutatások

Hazánkban a nemzetközi szakirodalomhoz képest csak viszonylag későn, a 90-es évek elején indultak meg a csődelőrejelzéssel kapcsolatos kutatások. A piacgazdasági környezet csak a rendszerváltás után alakult ki, és ezzel összefüggésben készültek el a törvényi szabályozások is. A csődeljárásról, felszámolási eljárásról és végelszámolásról szóló 1991. évi XLIX. törvény hatályba lépésével létrejöttek a csőd szabályozásának jogi keretei, lehetővé téve a gazdasági csőd vizsgálatára irányuló kutatásokat is.

2.6.1. Futó Péter empirikus vizsgálata

Futó Péter 1989-ben az Ipargazdasági Intézetben készített tanulmányt a magyarországi vállalatok fizetéseképtelensége és éves beszámolóikból képzett mutatószámai közötti összefüggések feltérképezésére. Kutatásait nem publikálta, annak eredményeit Virág-Hajdu (1998) ismerteti.

Futó Péter mintájában 24 iparvállalat (19 gépipari, négy könnyűipari és egy kohászati vállalat) adatai szerepeltek. A vizsgált vállalatok közül kettőről már a vizsgálat időpontjában ismert volt, hogy 1988 első hat hónapjában nem tudta fizetni tartozásait.

Kontrollként a szerző 150 könnyűipari és 150 gépipari vállalat 1986-87-es mérlegadataiból számított átlagos mutatószámokat használt. Módszere a szórásелемzés volt, továbbá a „diszkriminancia-analízis szemléletesre leegyszerűsített formája”.⁹

Munkája tulajdonképpen ugyanarra a megközelítésre épült, amelyet a nemzetközi szakirodalom a 30-as években tárgyalt. Vagyis a szerző azt vizsgálta, melyek azok a viszonyszámok, amelyekben a legnagyobb különbség mutatkozik a később csődbe menő, illetve a túlélő vállalatok csoportjai között. Azokat a mutatókat, amelyeknél a két átlag között nagy eltérés mutatkozott, szignifikánsnak tekintette a fizetéseképtelenség előre jelezhetősége szempontjából.

A szerző az alábbi mutatók esetén tapasztalta a legnagyobb eltéréseket csődbe jutott és túlélő vállalatok esetén:

- likvid passzíva / vevőállomány,
- vevőállomány / árbevétel,
- Bank felé adósság / likvid passzívák,
- hiteltörlesztés / Bank felé adósság,
- szállítóállomány változása,
- hiteltörlesztés változása,
- dollárexport változása,
- rubelexport változása,
- költségvetési befizetések és támogatások alakulása.

Futó Péter ezzel a megközelítéssel nem tudott ugyan választ adni arra, hogy az egyes mutatószámok milyen súllyal és jelentőséggel bírnak a megkülönböztetés vonatkozásában, de maga a tanulmány figyelemfelkeltő hatásában jelentős: rámutatott arra, hogy hazai viszonyok között is megvizsgálható és megvizsgálandó, melyek azok a viszonyszámok, amelyek leginkább alkalmasak a közelgő csőd előrejelzésére.¹⁰

⁹ Virág-Hajdu (1998)

¹⁰ Virág-Hajdu (1998)

2.6.2. Virág Miklós első hazai csődmodellje

A hazai viszonyokra kialakított csődmodell-kutatások Virág Miklós nevéhez köthetők: ő készítette el Hajdu Ottóval együttműködve az első magyarországi csődmodellt.

A munka alapját egy 154 vállalatot tartalmazó pénzügyminisztériumi adatbázis képezte. A vállalatok a feldolgozóiparban tevékenykedtek, a régi KSH ágazati osztályozási rendszer szerint a 13-17. és 19. ágazatokhoz tartoztak. Mindegyik létszáma meghaladta a 300 főt.

A minta fele-fele arányban tartalmazott fizetéseképtelenné vált és túlélő (ahogy a szerzők nevezték: kontroll) vállalatokat, csoportonként 77 céget. Csődbe ment vállalatoknak azokat tekintették, amelyek ellen 1992. augusztusában a Céglözlönyben közzétett, folyamatban lévő csődeljárás folyt, vagyis a csődtörvény hatályba lépése után gyakorlatilag azonnal csődeljárást kértek maguk ellen.

A szerzők a vállalatok 1990. és 1991. évi beszámolóinak adatai felhasználásával összesen 15 mutatót vizsgáltak a csődhelyzet kialakulásával összefüggésben. A mutatókat és számítási módjukat az alábbi táblázat tartalmazza:

A mutató megnevezése	A mutató számítási módja
Likviditási gyorsráta	$(\text{Forgóeszközök} - \text{Készletek}) / \text{Rövid lejáratú kötelezettségek}$
Likviditási ráta	$\text{Forgóeszközök} / \text{Rövid lejáratú kötelezettségek}$
Pénzeszközök aránya (%)	$(\text{Pénzeszközök} / \text{Forgóeszközök}) \times 100$
Cash-flow és összes tartozás aránya	$\text{Cash-flow} / \text{Összes tartozás}$
Forgóeszközök aránya (%)	$(\text{Forgóeszközök} / \text{Mérlegfőösszeg}) \times 100$
Tőkeellátottsági mutató (%)	$(\text{Befektetett eszközök} + \text{Készletek}) / \text{Saját vagyon} \times 100$
Eszközök forgási sebessége	$\text{Nettó árbevétel} / \text{Mérlegfőösszeg}$
Készletek forgási sebessége	$\text{Nettó árbevétel} / \text{Készletek}$
Vevők forgási sebessége (nap)	$(\text{Vevők} \times 360) / \text{Nettó árbevétel}$
Eladósodottság mértéke (%)	$(\text{Kötelezettségek} / \text{Mérlegfőösszeg}) \times 100$
Saját vagyon aránya (%)	$(\text{Saját tőke} / \text{Mérlegfőösszeg}) \times 100$
Bonitás	$\text{Kötelezettségek} / \text{Saját tőke}$
Befektetett eszközök hosszú lejáratú hitelekkel fedezett aránya (%)	$(\text{Hosszú lejáratú hitelek} / \text{Befektetett eszközök}) \times 100$
Forgóeszközök rövid lejáratú hitelekkel fedezett aránya	$(\text{Rövid lejáratú hitelek} / \text{Forgóeszközök}) \times 100$
Árbevétel arányos nyereség (%)	$(\text{Adózott eredmény} / \text{Nettó árbevétel}) \times 100$
Sajátvagyon-arányos nyereség (%)	$(\text{Adózott eredmény} / \text{Saját tőke}) \times 100$

7. táblázat: Az első magyar csődmodell kialakításánál használt pénzügyi mutatók¹¹

Modelljeiket a szakirodalomban a 90-es években uralkodó diszkriminancia analízis és logisztikus regresszió alkalmazásával alakították ki. Az előbbi modellt a csőd bekövetkezése előtti egy illetve két évre, a regressziós modellt pedig a csődöt megelőző év adataira vonatkozóan.

A végső modellekbe a fentiek közül csak azok a mutatók kerülnek be, amelyek az előrejelzés szempontjából szignifikánsak.

¹¹ Virág-Kristóf (2006)

Az 1991. év adataira épülő, diszkriminancia analízis felhasználásával készült modell eredményeként a következő függvénpár adódott:

$$\text{Csőd} = -7,73405 + 1,72122 X_1 - 0,22514 X_2 - 2,29162 X_3 + 0,21935 X_4$$

$$\text{Kontroll} = -10,35017 + 3,07788 X_1 + 1,40883 X_2 + 1,37222 X_3 + 0,25301 X_4$$

Mivel a besorolást csak két csoportra kell elvégezni, ebben a speciális esetben a besorolási függvényeket egymásból kivonva összevonhatók egyetlen egyenletbe:

$$z = 1,3566 X_1 + 1,63397 X_2 + 3,66384 X_3 + 0,03366 X_4$$

A modellben szereplő mutatók:

- X_1 : likviditási gyorsráta,
- X_2 : cash-flow / összes tartozás (dinamikus likviditás),
- X_3 : forgóeszközök / összes eszköz,
- X_4 : cash-flow / összes eszköz (dinamikus jövedelmezőség).

A modellnél a cut-off értéke 2,61612. Ennél kisebb z-érték esetén a modell a vizsgált vállalatot csődveszélyesnek, fölötte pedig túlélőnek minősíti.

A függvény besorolási pontossága a csődveszélyes csoportba kerülésre 82%-os, a kontroll csoportnál 74%-os volt, összességében nézve pedig 77,9%-os pontossággal teljesített.

A csődöt két évvel megelőző, 1990-es adatokon alapuló modell eredményeként a következő függvénpár adódott:

$$\text{Csőd} = -2,45586 + 5,13716 X_1 + 0,11897 X_2 - 12,932 X_3 + 0,07857 X_4$$

$$\text{Kontroll} = -2,85781 + 7,98291 X_1 + 0,1777 X_2 - 4,30594 X_3 + 0,06303 X_4$$

Vagy egymásból kivonva a kettőt:

$$z = 2,84575 X_1 + 0,05880 X_2 + 8,62619 X_3 - 0,01544 X_4$$

A modellben megjelenő mutatók csak részben egyeznek meg az előbbi modellével:

- X_1 : cash-flow / összes tartozás (dinamikus likviditási ráta),
- X_2 : likvid pénzeszközök aránya,
- X_3 : cash-flow / összes eszköz (dinamikus jövedelmezőség),
- X_4 : forgóeszközök rövid- és középlejártú hitelekkel fedezett aránya.

A cut-off értéke 0,40195. A modell a csődveszélyt 76,6%-os, míg a túlélést 70%-os megbízhatósággal jelezte előre. A teljes mintára vonatkozó besorolási pontossága 73,38% volt.

Az előzetes várakozásoknak megfelelően látható, hogy két évvel a csőd bekövetkezése előtt az előrejelzés megbízhatósága alacsonyabb. Ennek oka, hogy ilyenkor a csődbe jutott és túlélő cégek profilja még nem különül el olyan élesen, mint egy évvel a csődöt megelőzően.

A logisztikus regressziót a szerzők csak az 1991. évi adatokra alkalmazták. A logisztikus függvény:

$$\Pr(\text{csőd}) = \frac{e^{3,432-10,32X_1+0,01439X_2-4,438X_3-0,02992X_4+8,17X_5}}{1+e^{3,432-10,32X_1+0,01439X_2-4,438X_3-0,02992X_4+8,17X_5}}$$

A modell mutatói itt is csak részben egyeznek meg a diszkriminancia analízis eredményeivel. A logisztikus függvény mutatói:

- X_1 : likviditási gyorsráta,
- X_2 : árbevétel- arányos nyereség,
- X_3 : cash-flow / összes tartozás,
- X_4 : forgóeszközök aránya,
- X_5 : vevőállomány / szállítóállomány.

A cut-off érték ebben az esetben 0,525, tehát az ennél nagyobb függvényérték esetén tekintjük a vállalatot csődveszélyesnek, ez alatt pedig túlélőnek.

A csődvészélyesség előrejelzésénél a modell 84,4%-os pontosságot ért el, a túlélőknél 79,2%-ot. Összességében a megbízhatósága 81,8%. Ha ezt összevetjük a diszkriminancia analízissel készült modell eredményével, látható, hogy a logisztikus regresszió jobban teljesített, előrejelzése megbízhatóbb volt.

2.6.3. Virág Miklós– Hajdu Ottó ágazati modellcsaládja

A Virág Miklós – Hajdu Ottó szerzőpáros 1996-ban újabb csődmodellt mutatott be, Early Warning System elnevezéssel. Egészen pontosan nem egy modell, hanem egy csődmodell-család megalkotásáról van szó, amelyek a teljes nemzetgazdasági mintára vonatkozó modell mellett számos, egyes ágazatokra speciálisan felállított modelleket tartalmaznak. Az egyes ágazatokat a TEÁOR besorolás szerinti kódok alapján azonosították, majd ez alapján végezték el a csődmodell-számításokat.

A feldolgozott adatbázis mérete nemzetközi viszonylatban is jelentős, körülbelül tízezres nagyságrendű. Az alkalmazott módszertan a diszkriminancia analízis volt.

A szerzők összesen 41 modellt készítettek:

- egy átfogó modellt, amelynél a fejlesztési adatbázis a teljes nemzetgazdaságot lefedte,
- további 10 modell készült az egyes nemzetgazdasági ágakra,
- ágazati szintre lebontva pedig összesen 30 modellt dolgoztak ki:
 - 01: Mezőgazdaság, vadgazdálkodás és kapcsolódó szolgáltatások,
 - 05: Halászat és kapcsolódó szolgáltatások,
 - 17: Textíliák gyártása,
 - 18: Ruházati termékek gyártása, szőrmekikészítés és -festés,
 - 20: Fafeldolgozás,
 - 22: Kiadói és nyomdaipari tevékenység, hang- és képfelvételek,
 - 25: Gumi- és műanyagtermékek gyártása,
 - 28: Fémfeldolgozási termékek gyártása,

- 29: Gépek, gépi berendezések gyártása és javítása,
- 31: Villamosipari gépek és készülékek gyártása és javítása,
- 32: Híradástechnikai termékek gyártása és javítása,
- 33: Műszergyártás és -javítás,
- 34: Közúti járműgyártás,
- 35: Egyéb járművek gyártása és javítása,
- 36: Bútorgyártás, egyéb feldolgozóipari termékek gyártása,
- 41: Vízttermelés, -kezelés és -elosztás,
- 45: Építőipar,
- 50: Közúti jármű- és üzemanyag-kereskedelem,
- 51: Nagykereskedelem (közúti jármű- és üzemanyag-kereskedelem nélkül),
- 52: Kiskereskedelem (közúti jármű- és üzemanyag-kereskedelem nélkül),
- 60: Szárazföldi és csővezetékes szállítás,
- 63: A szállítás kiegészítő tevékenységei,
- 70: Ingatlanügyletek,
- 71: Ingóvagyon kölcsönzése,
- 72: Számítástechnikai és ehhez kapcsolódó tevékenységek,
- 74: Gazdasági tevékenységet segítő szolgáltatások.

Az egyes ágazatok specifikumait figyelembe vevő csődmódel-család pontossága felülmúlta a korábbi modellekét: a nemzetgazdaság egészére vonatkozó modell pontossága 98% fölötti, és az egyes ágazati modelleké is meghaladja a 95%-ot.

A modell részleteit, azok besorolási egyenleteit és a pontos klasszifikációs eredményeit a szerzők nem publikálták, mivel azok üzleti titkot képeznek. Annyit azonban nyilvánosságra hoztak, hogy a modelleket több magyar bank is megvásárolta, és

alkalmazza a hitelkihelyezés kockázatának megítélésére. Az eddigi tapasztalatok a gyakorlati alkalmazások során is kedvezőek voltak.

2.6.4. Virág-Kristóf neurális háló modellje

Körülbelül tíz év elteltével az első magyar csődmodellt a Virág-Kristóf szerzőpáros ismét frissítette, két szempont szerint újraszámolva azokat. Egyrészt módszertani fejlesztésként a neurális hálók segítségével is megkíséreltek új modellt felállítani, másrészt pedig olyan modelleket is teszteltek, amelyek nem az egyes vállalatok mutatóra épülnek, hanem a vállalati mutatók iparági átlagokhoz való viszonyán alapulnak.

Mindkét új tanulmány ugyanarra a pénzügyminisztériumi adatbázisra épült, amelyet az 1993-as kutatások során használtak. Az eltérés mindössze annyi, hogy nem 154, hanem 156 pár fizetéseképtelenné vált illetve túlélő feldolgozóipari vállalat adataival dolgoztak.

A neurális háló tanulási algoritmus az ún. „backpropagation” eljárás volt. A számításokat több beállítással is elvégezték, minden változatnál más-más számú közbenső réteget alkalmazva.

A különböző neurális hálók eredményeit foglalja össze az alábbi táblázat:

	Neuronok száma a két köztes rétegben (darab)					
	4-4	5-4	6-4	4-5	5-5	6-5
<i>Rontott fizetőképesség (darab)</i>	9	9	12	9	11	9
<i>Rontott fizetőképesség (százalék)</i>	11,5	11,5	15,4	11,5	14,1	11,5
<i>Rontott fizetéseképtelen (darab)</i>	17	18	13	23	12	12
<i>Rontott fizetéseképtelen (százalék)</i>	21,8	23,1	16,7	29,5	15,4	15,4
<i>Összes rontott (darab)</i>	26	27	25	32	23	21
<i>Összes rontott (százalék)</i>	16,7	17,3	16	20,5	14,7	13,5
<i>Besorolási pontosság (százalék)</i>	83,3	82,7	84	79,5	85,3	86,5

8. táblázat: Virág-Kristóf neurális hálóinak összefoglaló táblázata¹²

¹² Virág-Kristóf (2005)

A neurális hálókval való kísérlet eredménye egyértelműen alátámasztotta a nemzetközi szakirodalom tapasztalatait: a neurális hálók pontossága meghaladja mind a diszkriminancia analízist, mind a logisztikus regressziót. A modellek közül a legmegbízhatóbbnak a 17-6-5-1 struktúrájú neurális hálót találták, amelynek besorolási pontossága 86,5% volt, megelőzve ezzel a 77,9%-os diszkriminancia analízist és a 81,8%-os regressziós modellt is.

Az elsőfajú hibákat tekintve a diszkriminancia analízisnél 2,8 százalékponttal, a logisztikus regresszióval 5,4 százalékponttal teljesít jobban a neurális háló, másodfajú hiba tekintetében pedig 14,5 illetve 4,1 százalékponttal mutat jobb eredményt.

2.6.5. Virág-Kristóf iparági rátákon alapuló modellje

A Virág-Kristóf szerzőpáros másik tanulmánya, amely az első hazai csődmódel továbbfejlesztésére irányult, az iparági ráták felhasználására épült.

Az alkalmazott módszer lényege, hogy nem közvetlenül a vállalat pénzügyi mutatói felhasználásával épít előrejelzési modellt, hanem a mutatókat a vállalat tevékenysége alapján meghatározott nemzetgazdasági ágazat átlagos mutatójához viszonyítja, és ez képezi a modell bemenő adatait. Ez a következőképpen számítható:

$$(\text{iparági relatív ráta})_{k,j,t} = \frac{(\text{vállalati mutatószám})_{k,j,t}}{(\text{iparági átlagos ráta})_{j,t}} \times 100$$

ahol

- k: a vállalat,
- j: az iparág,
- t: a mutatószám fajtája.

A nevezőben szereplő százszoros célja, hogy a százalékos viszonyszámokat és az egynél nagyobb skaláris értékeket egymáshoz igazítsa. Az iparági átlag ezek alapján mindig 0,01-es értéket vesz fel, ehhez viszonyítva határozzuk meg az iparágához tartozó egyes vállalatok mutatóinak arányát.

A szerzők által felhasznált adatbázis megegyezett a fentiekben ismertetett mintával, azzal a kitételrel, hogy a 156 vállalatot kétfelé osztották: 75% tanulási mintára, és 25% tesztelési mintára. A tanulási mintán fejlesztett modelleket tesztelték a modell számára ismeretlen adatokon, ezáltal adtak becslést azok előrejelzési pontosságára.

Az adatbázis segítségével a szakirodalomban legelterjedtebb négy módszer (diszkriminancia analízis, logisztikus regresszió, rekurzív partícionáló algoritmus és neurális háló) alkalmazásával állítottak fel modelleket.

A diszkriminancia-függvény:

$$Z = -14101 X_1 + 33,74902 X_2 - 66,36 X_3 + 78,16907 X_4 + 146,1178 X_5 + 52,5579 X_6$$

A függvényben szereplő mutatók:

- X_1 : iparági átlaggal korrigált likviditási gyorsráta,
- X_2 : iparági átlaggal korrigált forgóeszközök aránya,
- X_3 : iparági átlaggal korrigált eszközök forgási sebessége,
- X_4 : iparági átlaggal korrigált készletek forgási sebessége,
- X_5 : iparági átlaggal korrigált saját vagyon aránya,
- X_6 : iparági átlaggal korrigált saját vagyonarányos nyereség.

A cut-off értéke 1,06252. Ennél nagyobb Z-érték esetén a modell a vállalatokat csődveszélyesnek ítéli, az alatt pedig túlélőnek minősíti.

A kapott logisztikus regressziós egyenlet:

$$\Pr(\text{túlélő}) = \frac{e^{0,04233+621,92432X_1-170,80129X_2-99,43514X_3-245,79408X_4}}{1 + e^{0,04233+621,92432X_1-170,80129X_2-99,43514X_3-245,79408X_4}}$$

Az előrejelzés szempontjából szignifikáns mutatók:

- X_1 : iparági átlagokkal korrigált likviditási gyorsráta,
- X_2 : iparági átlagokkal korrigált készletek forgási sebessége,

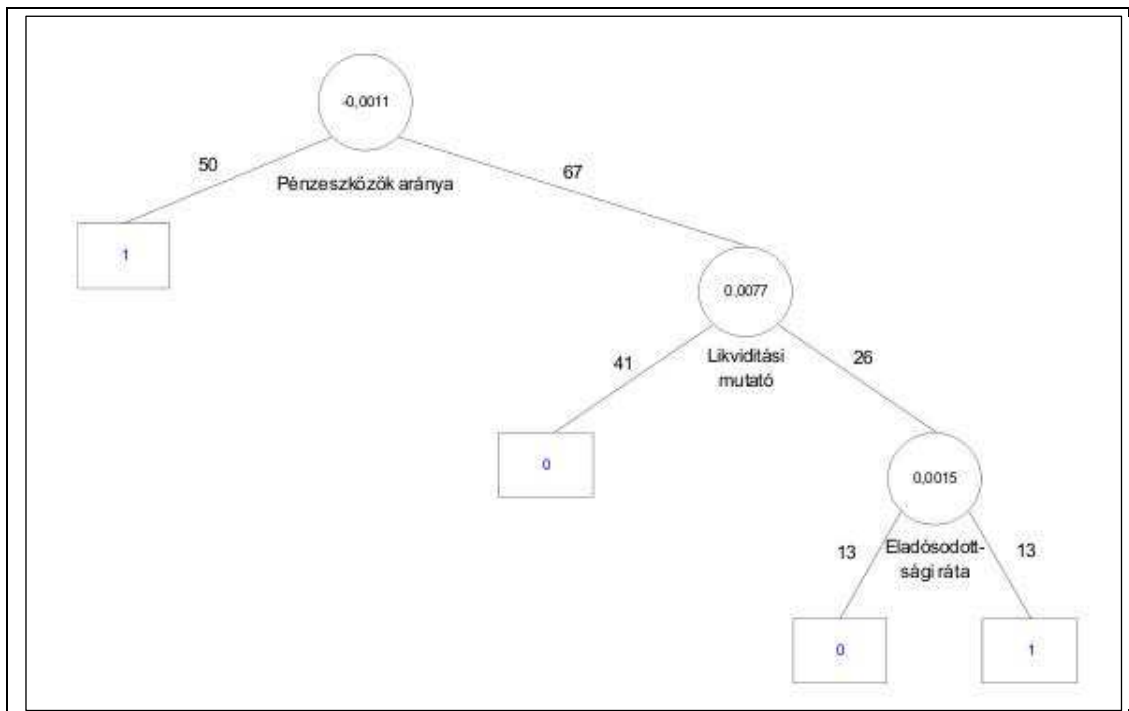
- X_3 : iparági átlagokkal korrigált vevők forgási sebessége,
- X_4 : iparági átlagokkal korrigált saját vagyon aránya.

A modell cut-off értéke 0,48. Ez alatti érték esetén a vállalat túlélőnek, fölötté pedig csődvészélyesnek minősítjük.

A rekurzív felosztási algoritmus 3 mutató felhasználásával osztotta fel a mintát:

- X_1 : iparági átlagokkal korrigált pénzeszközök aránya,
- X_2 : iparági átlagokkal korrigált likviditási mutató,
- X_3 : iparági átlagokkal korrigált eladósodottsági ráta.

A túlélőket 1-gyel, a csődbe jutott vállalatokat 0-val jelölve az alábbi döntési fa rajzolható fel:



1. ábra: Virág-Kristóf döntési fa modellje¹³

¹³ Virág-Kristóf (2006)

Ahogy az a fentiekben látszik, az első vágási szempont az iparági átlaggal korrigált pénzeszközök aránya. Amennyiben a mutató értéke a vizsgált vállalat esetén kisebb, mint $-0,0011$, a modell túlélőnek minősíti. A mintában összesen 50 ilyen vállalat volt. A fennmaradó 67 céget a modell tovább vizsgálja, mégpedig a likviditási mutató szempontjából. Ha ez $0,0077$ -nél kisebb, a modell csődvészélyesnek tekinti a vállalatot, a minta esetén összesen 41-et. A fennmaradó 26 vállalatot, amelynek likviditási mutatója nagyobb, mint $0,0077$, tovább osztja az eladósodottsági ráta alapján: ha $0,0015$ alatt van a mutató értéke, akkor csődvészélyesnek, e fölött pedig fizetőképességnek sorolja be a modell.

A korábbi, neurális hálókval folytatott kutatási eredményeik felhasználásával a szerzők 16-6-4-1 struktúrájú neurális háló felhasználásával, 400 tanulási ciklus lefuttatásával készítettek csődmodellt. A modell besorolási pontossága (85,76%) a legmagasabb volt az összes modell között, a tesztelési mintán elért pontossága terén pedig még inkább a többi módszer fölé emelkedett.

Az egyes modellek besorolási pontosságát az alábbi táblázat foglalja össze:

Előrejelzési módszer	Besorolási pontosság	
	Tanulási minta	Tesztelő minta
<i>Diszkriminancia analízis</i>	82,91	79,49
<i>Logisztikus regresszió</i>	85,47	74,36
<i>Rekurzív partícionáló algoritmus</i>	82,91	79,49
<i>Neurális háló</i>	85,76	87,28

9. táblázat: Virág-Kristóf iparági rátákon alapuló csődmodelljeinek besorolási pontosságai¹⁴

¹⁴ Virág-Kristóf (2006)

3. A csődmodellek fejlődéstörténetének egy lehetséges rendszerezése, módszertani megközelítéssel

3.1. A csődmodellek általános jellemzői

A modellek áttekintése után megfogalmazhatunk bizonyos megállapításokat, amelyek általánosságban jellemzik a csődmodelleket.

A modellek közös jellemzője, hogy a vállalatok éves beszámolóiból, pénzügyi-számviteli kimutatásokból képzett mutatók elemzésével próbálják előre jelezni a néhány éven belül bekövetkező csődöt, illetve ennek elmaradását, és a vállalat további sikeres működését. Néhány esetben azt is tapasztalhatjuk, hogy ezeket a mutatókat kiegészítetik egyéb tényezők vizsgálatával, például gazdasági környezetet jellemző mutatókat, egyéb tényezőket építenek be a modellbe.

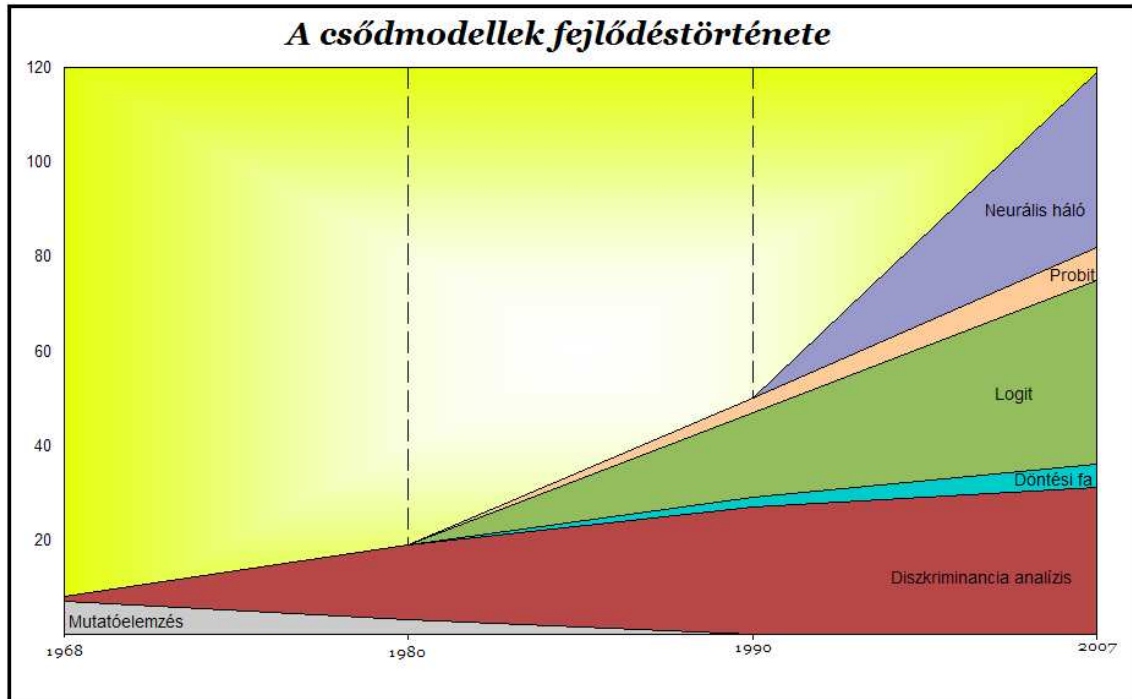
A modellépítés során olyan többváltozós statisztikai elemzéseket végeznek el, amelyek feltárják a mutatók összessége és a csőd bekövetkezése közötti kapcsolatokat. A csőd előrejelzésének szempontjából releváns mutatók lesznek részei a modellnek.

3.2. Módszertani fejlődéstörténet

A szakirodalom feldolgozása során összesen 126 modellt vizsgáltam, amelyeket az alkalmazott módszertan szempontjából csoportosítottam. A kutatások időrendiségét és módszertanát figyelembe véve készítettem áttekintő táblázatot, amelyet az 1. mellékletben helyeztem el.¹⁵

¹⁵ A fejezet eredményei az alábbi tanulmányban kerültek publikálásra: Imre (2007a).

A módszertani fejlődést egy ábrával szemléltetem, amely az idő függvényében mutatja az egyes módszerek felhasználásával készült tanulmányok számát:



2. ábra: A csődmodellek fejlődéstörténete módszertani megközelítéssel

Az ábra alapján behatárolhatók a csődmodellek fejlődési szakaszai. A rendszerezésükre egy 1+3-as felosztást javaslok, amelyet az alábbi tézisben fogalmaztam meg.

T1. A szakirodalom feldolgozása alapján a kutatások időrendiségét és módszertanát figyelembe véve a csődmodellek fejlődéstörténetének 4 szakasza különböztethető meg:

0. A csődkutatások kialakulásának korszaka (1931-67)

1. A diszkriminancia analízis korszaka (1968-79)

2. A logisztikus regresszió térhódításának korszaka (1980-1989)

3. A mesterséges intelligencia megjelenésének korszaka (1990-től napjainkig)

Két fejlődési mellékág azonosítható: a probit és a rekurzív felosztás módszerei.

Az egyes fejlődési szakaszokat és azok jellemzőit a következőkben foglalom össze.

3.2.1. *A csődkutatások kialakulásának korszaka (1931-67)*

Ez a korszak nem igazán eredményezett mai értelemben vett csődmodellt. Inkább a csőd, mint gazdasági esemény előrejelzésének vizsgálata köthető ehhez az időszakhoz, amolyan 0. szakaszként foghatjuk fel.

A korszak a harmincas évektől kezdődik, és 1967-ig tart. A kutatások a pénzügyi mutatók egyváltozós vizsgálatára terjednek ki: a kutatók azt elemzik, hogy a csődbe jutott illetve a túlélő vállalatok csoportjai milyen mutatók esetén mutatnak egymástól jelentősen eltérő átlagos értékeket. Ezeket a mutatókat tekintik előrejelzésre alkalmas eszköznek.

Ide sorolhatjuk Ramser, Foster, Fitzpatrick, Winakor, Smith, Merwin és Beaver tanulmányait.

3.2.2. *A diszkriminancia analízis korszaka (1968-79)*

A korszakot az első csődmodell kidolgozója, Altman neve és a diszkriminancia analízis elterjedése jellemzi. Bár még előfordulnak egyváltozós elemzések, a domináns módszertan egyértelműen a diszkriminancia analízis.

A korszak legjelentősebb kutatói: Altman és szerzőtársai, Deakin, Blum, Edmister, Taffler – Tisshaw, van Frederikslust, Bilderbeek, Dambolena, Khoury, Ooghe, Verbaere, Taffler, Micha, Betts, Belhoul, Gombola, Haskins, Ketz, Williams, Gloubos, Grammatikos, Lussier, stb.

3.2.3. *A logisztikus regresszió térhódításának korszaka (1980-1989)*

A logisztikus regresszió beemelése a csődmodellek módszertanába Ohlson nevéhez fűződik. A módszer gyorsan elterjedt, és mind a mai napig az egyik leggyakrabban alkalmazott modellezési módszernek, „best practice”-nek tekinthető.

A logit modellek felhasználásával alkottak modelleket többek között: Ohlson, Pantalone, Zavgren, Keasey, McGuinness, Rudolph, Hamdan, Collins, Green, Gentry,

Newbold, Whitford, Aziz, Emanuel, Lawson, Gloubos, Grammatikos, Platt és Platt, Ooghe, Joos, De Vos, Hunter, Isachenkova, stb.

Ehhez a korszakhoz tartozik még két fejlődési mellékág: a probit és a rekurzív felosztás. Ezek a módszerek nem terjedtek el annyira a szakirodalomban, bár időről-időre kísérleteztek velük, mégsem tudták megtörni a logit modellek dominanciáját.

3.2.4. A mesterséges intelligencia megjelenésének korszaka (1990-től napjainkig)

A mesterséges intelligencia a neurális hálók technikájával van jelen a csődmódel-irodalomban. 1990-es bevezetése Odom és Sharda nevéhez kötődik, és napjainkban is ez a legjelentősebb témája a csődmódellek módszertani kutatásainak. A kutatások többsége azt a következtetést vonja le, hogy a neurális hálók eredményei és előrejelző képessége jobb, mint a korábbi módszerekkel készült modelleké.

A neurális hálókra alapozták modelljeiket pl. Odom és Sharda, Coats, Fant, Back és szerzőtársai, Olmeda, Fernandez, Kiviluoto, Laitinen, Kankaanpää, Tam, Kiang, Salchenberger és szerzőtársai, Kerling, Poddig, Alici, Leshno, Spector, Tan, Yang, stb.

Az is megfigyelhető, hogy a diszkriminancia analízis és a logit még most is jelentős számban megtalálható a tanulmányok módszerei között. A diszkriminancia analízis szerepe azonban fokozatosan átalakult: általában az újabb modellek teljesítményeinek összehasonlítása miatt alkalmazzák a kutatók, amolyan vonatkoztatási pontként, jól bevált viszonyítási alapként tekintenek a módszerre. A logit ezzel szemben ma is gyakorlati értékkel bír és több területen (pl. bankoknál) töretlenül alkalmazzák ma is.

4. Módszertani áttekintés

4.1. A csődmodelleknél alkalmazott statisztikai módszerek

4.1.1. *Diszkriminancia analízis*

A diszkriminancia analízis egy olyan statisztikai eszköz, amely a megfigyeléseket azok egyedi jellegzetességei alapján besorolja az előre definiált csoportok valamelyikébe. A csődmodellekre lefordítva ez azt jelenti, hogy a vállalatokat kategorizálja aszerint, hogy a csődbe jutott vagy a túlélő vállalatokkal mutatnak nagyobb hasonlóságot.

A módszer azon a gondolatmeneten alapul, hogy a két csoport egy adott tulajdonságban különbözik egymástól, és ez alapján lehet őket elkülöníteni. Ennek feltétele, hogy a két csoporton belül ez a tulajdonság azonos eloszlású legyen, és csak a csoporton belüli várható értékek különbözzenek egymástól. Ekkor kell lennie egy olyan küszöbértéknek, amely fölött az egyik, alatta pedig a másik csoportba tartozás valószínűsége nagyobb.

A csoportba sorolást a diszkriminancia analízis nem egy tulajdonság alapján végzi, hanem a Bayes-féle klasszifikációval, több folytonos változó alapján vizsgálja. Ekkor az egyes csoportokat már nem csak egy változó várható értéke, hanem a csoportra vonatkozó centroid alapján hasonlítjuk össze, amely tulajdonképpen az egyes változók csoporton belüli átlagaiból álló vektor.

Egy adott megfigyelési elemnek a centroidtól való távolsága a Mahalanobis-féle távolsággal mérhető. A megfigyelési elemet abba a csoportba soroljuk, amelyik csoport centroidjától mért ún. Mahalanobis-távolsága kisebb.

A módszer alkalmazásakor tehát minden csoportra külön klasszifikációs függvény készül, amelyek általános formája a következő:

$$Z = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \dots + a_n X_n$$

ahol:

- a_0 = konstans,

- $a_1, a_2, a_3, \dots a_n$ = diszkrimináló koefficiensek,
- $X_1, X_2, X_3, \dots X_n$ = független változók.

A függvények alkalmazása úgy történik, hogy a minta minden egyes megfigyelésére kiszámítjuk a Z-értékeket mind a két csoport függvényével. Azaz minden vállalat adatai alapján a diszkrimináló változók értékeit behelyettesítjük a csődbe ment és a túlélő csoportok diszkriminancia függvényébe. Amelyik függvény esetén magasabb Z-érték adódik, a vizsgált vállalat ahhoz a csoporthoz hasonlít jobban, így oda soroljuk be a vállalatot.

Amennyiben két csoportra vonatkozóan végezzük el a diszkriminancia analízist, a két klasszifikációs függvényt kivonva egymásból egyetlen klasszifikációs függvényt kapunk. Ehhez tartozik egy kritikus érték (cut-off), amely az eredeti két függvény konstans tagjainak különbsége alapján adódik. Ez az érték választja el a két csoportot, illetve határozza meg azt, hogy a megfigyelési elemet melyik csoportba soroljuk.

A módszer előnyei közül kiemelhetők az alábbiak:

- elsősorban a modell többváltozós jellege: a minősítésnél a megfigyelési tulajdonságok teljes körét figyelembe veszi, túlhaladva ezzel az egyváltozós megközelítéseken.
- viszonylag egyszerűen kezelhető,
- egy minta alapján létrehozott klasszifikációs függvény segítségével olyan elemeket is be tud sorolni a két csoport valamelyikébe, amelyek az eredeti mintában nem szerepeltek. Ezt használjuk ki a csődelőrejelzésnél is, amikor a modell számára ismeretlen vállalatról akarjuk eldönteni, hogy a csődveszélyes, vagy inkább a túlélő vállalatokhoz hasonlít jobban.

Hátrányaként említhetjük azonban, hogy több korlátozó feltétellel bír:

- A modell először is feltételezi a független változók normális eloszlását. A gyakorlati életben már ez a feltétel is nehezen teljesíthető¹⁶. Ezen felül feltételezi azt is, hogy a változók bármely lineáris kombinációja is normális eloszlású. Ez az erős feltétel teszi

¹⁶ lásd Halas (2004)

sokszor megkérdőjelezhetővé a módszer alkalmazásának helyességét, és indokolhatja más eljárások használatát.

- A módszer további feltételezése, hogy a csoportok variancia- és kovarianciamátrixai azonosak.
- A függő változókhoz tartozó együtthatók nem értelmezhetők az egyes ismérvek súlyaként.
- Ismerni kell továbbá a csoportokba kerülés előzetes valószínűségét és a hibás besorolás költségét.

4.1.2. Regressziós modellek

A regressziós modellek célja ugyanaz, mint a diszkriminancia analízisé, azaz előre meghatározott csoportokba sorolni a minta megfigyelési elemeit. Bár létezik több csoportra vonatkozó regressziós modell is (ún. polichotom regresszió), a gyakorlatban a két csoporttal dolgozó, ún. dichotom regressziót alkalmazzák leggyakrabban. A csődelőrejelzés szempontjából is ezzel dolgozunk, mivel a cél a vizsgált vállalatoknak a csődveszélyes vagy a túlélő kategóriák egyikébe való besorolása.

A csődmodelleknél kétféle regressziós modell használatos: a logit és a probit analízis. Mindkét modell alapja az odds, amely az előzőleg definiált két csoportba kerülés valószínűségeinek egymáshoz viszonyított arányát jelenti. Csődmodellek esetében gyakorlatilag ez a túlélés és a csőd bekövetkezési esélyeinek hányadosát takarja.

Képlettel kifejezve:

$$odds_x = \frac{P_{\text{túlélő}|x}}{P_{\text{csőd}|x}}$$

ahol x az adott megfigyelésre vonatkozó független változók ($x_1, x_2 \dots x_p$) vektora. Csődmodellekre értelmezve pedig a vizsgált vállalatok mutatószámainak vektora.

Dichotom regresszió esetén csak két csoportot vizsgálunk, ezért az odds kifejezhető csak az egyik esemény bekövetkezési valószínűségével is, hiszen a két esemény együttes valószínűsége 1. Vagyis:

$$odds_x = \frac{P_{\text{túlélő}|x}}{1 - P_{\text{túlélő}|x}}$$

ahol x az adott megfigyelésre vonatkozó független változók ($x_1, x_2 \dots x_p$) vektora.

A logit és a probit módszer ezen a ponton válik el egymástól. Alapvetően abban különböznek, hogy milyen feltételezéssel élnek a P valószínűség eloszlásáról.

4.1.2.1. Logisztikus (logit) regresszió

A logit módszer azzal a feltételezéssel él, hogy az odds arány logaritmusai lineárisan függ a magyarázó változóktól. Az odds logaritmusát nevezik logitnak, innen ered a módszer elnevezése is.

Képlettel ezt a következőképp fejezhetjük ki:

$$\text{logit} = \ln(\text{odds}) = \beta^T x$$

ahol β az x -re vonatkozó paraméterek vektora.

A logisztikus regresszió a független változók ($x_1, x_2, \dots x_n$) eloszlására vonatkozóan viszont semmilyen további feltétellel nem él. Ez adja legfőbb előnyét.

Mivel a P valószínűsége a logisztikus eloszlást feltételezzük, ezért:

$$P_{\text{túlélő}|x} = \frac{e^{\beta^T x}}{1 + e^{\beta^T x}} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}$$

A modell β_i paramétereinek becslése a maximum likelihood elv alapján történik. A likelihood függvény az adott minta bekövetkezésének feltételes valószínűsége a megfelelő x értékek esetén, vagyis a minta modell alapján becsült valószínűsége. Ez úgy számítható, hogy minden mintaelemre kiszámoljuk a saját eseményének valószínűségét, majd ezek szorzatát képezzük. A β paramétereket pedig úgy becsüljük, hogy ezt az értéket maximalizáljuk az adott mintára.

A modell segítségével az ismert x változók alapján minden mintaelemre meghatározható a túlélés valószínűsége, amely kiterjeszthető a mintán kívüli, a modell számára ismeretlen megfigyelési elemek minősítésére is.

A klasszifikációhoz szükség van még egy úgynevezett cut-off pont meghatározására. A cut-offnál nagyobb P érték esetén a vállalatot túlélőnek, alatta pedig csődveszélyesnek valószínűsítjük. A cut-off meghatározásához a klasszifikációs mátrixot használják, amely a későbbiekben kerül ismertetésre. Röviden elég annyit megjegyezni, hogy a cut-off beállítása a modell által adott helytelen besorolások arányának minimalizálása révén határozandó meg.

A logit modell előnyei:

- Nem támaszt olyan korlátozó feltételeket a változók normalitásával kapcsolatban, mint a diszkriminancia analízis.
- A módszer a besorolási döntések megalapozása során megengedi a diszkrét és folytonos változók egyidejű figyelembevételét.
- A modell együttthatóinak értékei egyben az egyes mutatók fontosságát is jelzik, azaz súlyokként értelmezhetők.
- A logisztikus regresszió függvényében nem csak maguk a változók, hanem azok különféle interakciói is felhasználhatók önálló változókként.
- A logit érték 0 és 1 közötti értéket vehet fel, így gyakorlatilag azonnal valószínűségi mutatóvá alakítható.

Hátrányai viszont a következők:

- A módszer igen érzékeny a független változók közötti korrelációra, illetve a hiányzó vagy szélsőséges értékekre. Ennek kivédése érdekében modellépítés előtt mindenképp adatkorrekciókra és szűrési lépésekre van szükség.
- A logit modellek nagy elemszámú minta esetén jól teljesítenek, ugyanakkor kisebb minták esetén jelentősen romlik az előrejelző képességük.

4.1.2.2. *Probit regresszió*

A probit analízis egyetlen feltételezésben tér el a logisztikus regressziótól. Ahogy a fentiekben látható volt, a logit azt feltételezi, hogy a P valószínűség logisztikus eloszlásfüggvénnyel jellemezhető. A probit analízis feltételezése ezzel szemben az, hogy a P valószínűség standard normális eloszlást követ.

Bár a módszerek a két eloszlás függvény hasonlóságából adódóan nagyon hasonló eredményeket adnak, a probit nehézkesebben alkalmazható. A normális eloszlásfüggvénynek ugyanis a logisztikus eloszlásfüggvénnyel ellentétben nincs zárt alakja, így a logit modell számítása lényegesen egyszerűbb.

Ez az egyik legfőbb oka annak, hogy a gyakorlatban csak a logit regresszió terjedt el, a probit módszert csak elvétve alkalmazzák.

4.1.3. *Rekurzív partícionáló algoritmus (döntési fák)*

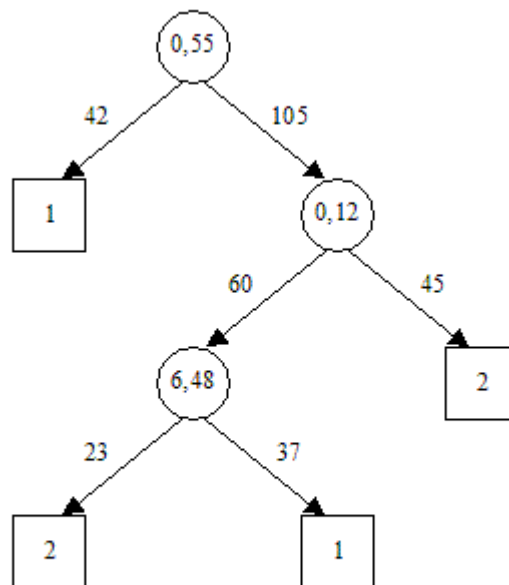
A rekurzív partícionáló algoritmusnak több más elterjedt elnevezése is van: döntési fa, klasszifikációs fa, visszatérő felosztási algoritmus, RPA (Recursive Partitioning Algorithm). Bárhogyan is nevezzük, ugyanazt értjük alatta: egy kifejezetten számítógépes alkalmazásra kifejlesztett osztályozó eljárásról van szó, amelynek végeredménye egy bináris jellegű osztályozó fa struktúra.

A módszer célja az előzőekben bemutatottakhoz teljesen hasonló: egy mintaelem-halmaz besorolása előre meghatározott kategóriákba az elemek bizonyos megfigyelt tulajdonságai alapján. Az eddigiektől eltérően azonban a rekurzív felosztás nem egy parametrikus módszer.

Az algoritmus a mintafelismerésen alapul, vagyis a mintában megfigyelhető jelenségek alapján döntési szabályok sorozatát határozza meg az egyes tulajdonságokra, majd ezekre alapozva osztja csoportokba a minta elemeit. Ezeket a szabályokat binárisan kifejezhető relációkkal fogalmazza meg (igen-nem, kisebb-nagyobb, stb), így teszi lehetővé a szabályok fa struktúrában való ábrázolását. Az egyes döntési pontokon a minta mindig kétfelé válik az adott szempont alapján.

A fentiekből adódóan a módszer furcsa kettőssége figyelhető meg: az egyváltozós elemzésekkel rokon abból a szempontból, hogy minden döntési pontnál csak egy változó alapján osztja ketté a mintát. Összességében viszont több változót is felhasznál, hiszen egy fa több döntési pontból épül fel, amelyek több tulajdonságot vizsgálnak.

Egy tipikus döntési fát mutat be az alábbi ábra. A legelterjedtebb ábrázolási technika szerint körök jelölik a döntési pontokat és négyzetek a besorolási osztályokat. A körökben szereplő számok jelentik a vágási értékeket. Ha egy mintaelem adott változó szerinti értéke kisebb vagy egyenlő, mint a vágási érték, akkor a bal oldali ágra kerül, ha nagyobb, akkor a jobb oldalra. Az ágakon olvasható számok a hasítás utáni részminta elemszámai, vagyis azt mutatják, hogy hány megfigyelési elem tett eleget az adott ágra kerülés feltételének. A fa végpontjai a levelek, amelyek a besorolási csoportokat mutatják.



3. ábra: Egy tipikus döntési fa

A döntési fa előállítása iteratív folyamat. Az algoritmus első lépésként a teljes mintából indul ki. Egyesével megvizsgálja a változókat és az összes lehetséges kétfelé osztást,

majd kijelöli azt a változót, amelyik a legkisebb hibát követi el az osztályok elválasztásakor, vagyis amelyik legjobban növeli a homogenitást.

A homogenitás meghatározásánál a modell azt veszi alapul, hogy a felépítendő modellnél a csoportokon belüli variancia minél kisebb, míg a csoportok közötti variancia minél nagyobb legyen.

A következőkben az algoritmus ugyanilyen logika alapján meghatározza a második, harmadik stb. döntési változót is, egészen addig, amíg a teljes fa lépésről-lépésre fel nem épül.

Az algoritmus által végül kialakított döntési fa így a legjobb besorolási pontosságot eredményezi, azonban a legtöbb esetben egyrészt túl bonyolult, másrészt egy buktatót is rejt magában. A legpontosabb fa ugyanis csak a minta jellegzetességeit követi, túlzottan alkalmazkodik a minta specifikumaihoz. A kialakított modell így előrejelzésre rendszerint alkalmatlan, hiszen az algoritmus által nem ismert adatokon jelentősen visszaesik a besorolási pontossága.

Ezért az első lépésben lefutott algoritmus után szükség van egy második lépésre is, ahol ezt a „legjobb” fát egyszerűsítik. Ennek legelterjedtebb módja, hogy a teljes mintát felosztják egy tanulási és egy tesztelési részmintára. A tanulási minta adatain felépített döntési fát a tesztelési mintán való alkalmazás iterációi során „megnyesik”, azaz csökkentik a komplexitását, amelyet egyszerűen a végpontok számával mérnek.

Ez addig folytatódik, amíg két részmintán mért besorolási hibák körülbelül azonos mértékűek nem lesznek. Az így megnyesett döntési fát tekintjük a végső előrejelzési modellnek, amelyet ezután már megfelelő megbízhatósági szinten lehet alkalmazni a modell számára ismeretlen adatokon is.

A rekurzív felosztási algoritmus előnyei:

- egyesíti az egyváltozós elemzések egyszerűségét a többváltozós megközelítések előnyeivel;

- nem követel meg semmiféle feltevést a változókra vonatkozóan, mivel nem paraméterbecslés alapján dolgozik;
- egyszerű, követhető struktúra, amely könnyen értelmezhető és jól algoritmizálható.

A módszer hátrányai:

- Fennáll a túltanulás veszélye: ilyenkor a döntési fa nem a sokaságban megfigyelhető jelenségeket, hanem magát a mintát modellezi, azaz túlzottan alkalmazkodik a minta specifikumaihoz.

4.1.4. Neurális hálók

Az ismertetett módszerek közül a legújabb, mesterséges intelligencián alapuló modellezési eljárás a neurális háló. Működése leginkább az emberi agy tanulási folyamatához hasonlítható, innen ered az elnevezése is.

Egy neurális háló az alábbi három ismérv alapján határozható meg egyértelműen¹⁷:

- a felhasznált idegsejt modell,
- a neuronok összekötésének módja (topológia),
- a tanulási módszer.

Csakúgy, mint az emberi idegrendszer, a neurális háló modellje is számos egymáshoz kapcsolódó neuronból épül fel, amelyek biológiai megfelelőikhez hasonlóan az információ-feldolgozási egységei, és szoros összeköttetésben vannak egymással.

A hálózat felépítésekor a neuronok csoportjai külön rétegekbe (layer) szerveződnek. Megkülönböztetünk:

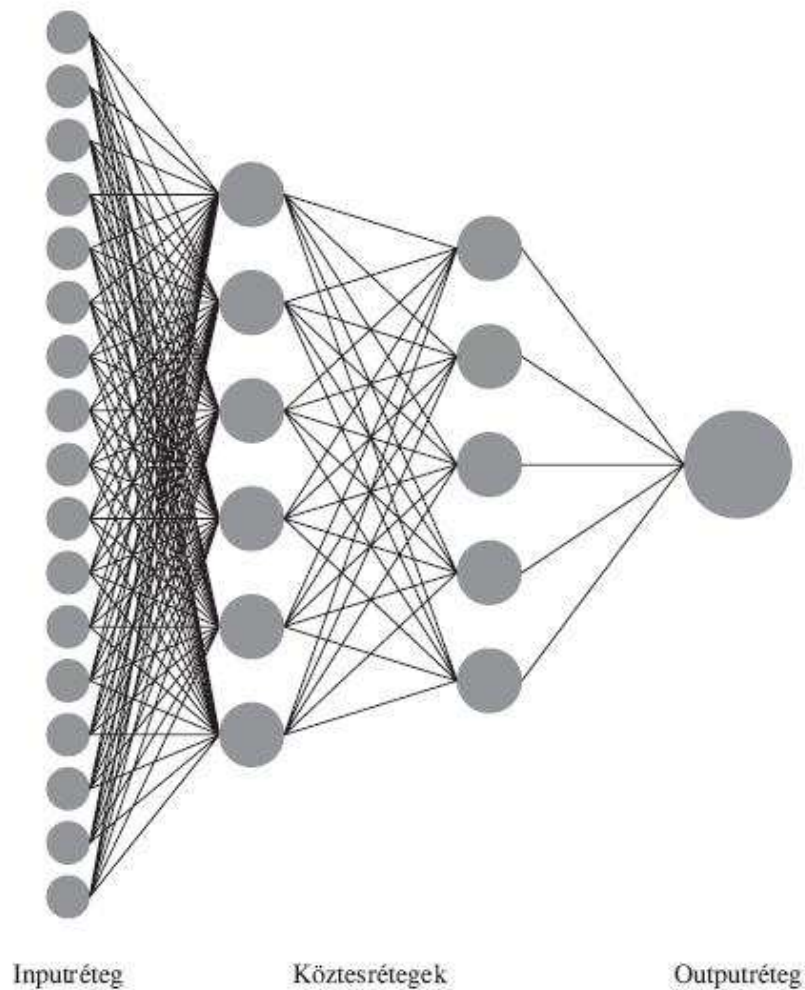
- egy input réteget, amelyek neuronjain keresztül a mintainformációk a modellbe érkeznek,

¹⁷ Kiss (2003)

- a háló felépítésétől függően egy vagy több köztes réteget (vagy rejtett, hidden layer), amelyekben ezek az információk feldolgozásra kerülnek,
- egy output réteget, amelyen keresztül az eredményt kapjuk.

Léteznek olyan modellek is, amelyeknél köztes réteg nem szerepel, az input és output layer között közvetlen kapcsolat van.

Az alábbi ábra egy lehetséges neurális hálót szemléltet, két köztes réteggel.



4. ábra: Neurális háló két köztes réteggel¹⁸

¹⁸ Virág-Kristóf (2005)

A neuronok közötti kapcsolatok alapján megkülönböztethetünk visszacsatolós és visszacsatolás nélküli neurális hálókat. Az utóbbiaknál az információ szigorúan csak az input réteg felől az output réteg felé halad, míg a visszacsatolós hálózatban az információ visszacsatolható egy korábbi rétegbe is. Sőt, a visszacsatolásnál előfordulhat az is, hogy egy neuron önmagához is kapcsolódjon, így erősítve saját hatását.

A neurális háló lényege a tanulási algoritmus. Ennek célja, hogy a modell egy minta alapján képes legyen felismerni a mögötte rejlő jelenségeket. A neuronok rendelkeznek egy saját belső függvényvel, amelyhez adott egy kezdő súlyozás. A modellben e súlyok megváltoztatása jelenti a tanulást. A neuron a bemenetére érkező jeleket súlyozza, majd meghatározza az eredő kimenetet az átviteli függvény felhasználásával. A hálózat eltárolja és megtanulja az inputok és outputok közötti összefüggéseket, és kiépíti a legjobb súlyokat. A tanulási folyamat akkor fejeződik be, ha elér egy bizonyos elfogadható hibaszintet, ekkor a súlyok rögzülnek.

Az egyik legismertebb tanulási algoritmus az ún. back-propagation eljárás. Ennek lényege a hiba-visszacsatolás, vagyis a tanulási folyamat során kapott outputokat a rendszer folyamatosan összehasonlítja a mintában adott elvárt értékkel. Amennyiben eltérést tapasztal, felülvizsgálja és korrigálja a belső súlyokat, egészen az elfogadható hibahatár eléréséig, vagyis az inputok és outputok közötti kapcsolat megtanulásáig. A hibahatár definiáláshoz az átlagos négyzetes hibát (MSE) használja, az optimális súlyozás elérését az MSE minimalizálása jelzi.

Ahogy a rekurzív felosztási algoritmusnál is, úgy a neurális hálók esetében is fennáll a túltanulás veszélye. Ez az a jelenség, amikor a tanulási folyamat során nem az általános összefüggéseket, hanem az adott minta sajátosságait tanulja meg a modell.

Ennek elkerülése érdekében ugyanúgy kettéosztják a mintát tanulási és tesztelési részmintákra. A tanulási mintán megedzett modell eredményeit összehasonlítják a számára ismeretlen tesztelési mintán elért eredményekkel. Ha a besorolási pontosság hasonló a tanulási mintához, akkor a modell véglegesíthető. Ha viszont a tesztelési mintán mért hibaarányok jelentősen meghaladják a tanulási mintán mért szintet, akkor bizonyítást nyer a túltanulás, új modellt kell készíteni.

A tútanulás elkerülését alapvetően a tanuló ciklusok számának helyes megválasztása biztosíthatja. Ez gyakorlatban számos kísérletezést és folyamatos nyomon követést takar. Az eredmény eléréséhez ugyanakkor rendkívül fontos, mivel sem az elégtelenül megedzett, sem a tútanult neurális háló nem alkalmas előrejelzésre.¹⁹

A módszer jellegéből fakadóan számos előnnyel rendelkezik²⁰:

- a neurális háló képes bonyolult mintákat gyorsan és magas megbízhatósági fokkal elemezni, feltárni az azokban rejlő összefüggéseket,
- mindezt bármilyen előreprogramozott tudás nélkül teszi,
- nincsen statisztikai kötöttség az eloszlásokra vonatkozóan,
- kvalitatív ismérveket is képes feldolgozni,
- nem érzékeny a hiányzó, hibás vagy extrém input adatokra,
- a független változók korrelációja sem okoz problémát a módszer számára,
- valódi előrejelző eljárás,
- jellegénél fogva robosztusabb és flexibilisebb a statisztikai modelleknél.

Meg kell említeni ugyanakkor az alábbi hátrányait is:

- a modell által meghatározott súlyok értelmezése nehezebb;
- nagyobb módszertani mérési tapasztalatot és nagy elemszámú tanulási mintát igényel;
- fennáll a tútanulás veszélye.

4.1.5. Módszertani összefoglaló

Összefoglalásként tekintsük át a fentiekben részletesen bemutatott modellek előnyeit és hátrányait!

¹⁹ A neurális hálókról bővebben lásd: Borgulya (1998)

²⁰ Halas (2004)

Módszer	Előnyök	Hátrányok
<i>Diszkriminancia analízis</i>	• többváltozós modell • egyszerű felépítés	• több korlátozó előfeltételezéssel él (linearitás, normalitás, előzetes csoportvalószínűségek) • csak sorrendi mérő • a független változók súlya nem határozható meg
<i>Regressziós modellek</i>	• többváltozós modell • kvalitatív ismérveket is képes felhasználni • a független változók súlya kinyerhető • nem feltételez linearitást	• érzékeny a független változók korrelációjára • érzékeny a kiugró vagy hiányzó értékekre • nagyobb mértékű nemnormális eloszlás torzíthatja • kis mintán nem hatékony
<i>Rekurzív partícionálás</i>	• egyesíti az egyváltozós elemzések egyszerűségét a többváltozós megközelítések előnyeivel, • egyszerű felépítés, • könnyen interpretálható és algoritmizálható eredmények • nem él korlátozó előfeltételekkel	• az éles szétválasztások miatt sarkításokat eredményezhet • fennáll a túltanulás veszélye
<i>Neurális háló</i>	• komplex, robusztus felépítés • előre beépített tudás nélkül dolgozik • flexibilitás • nem él korlátozó előfeltételekkel • kvalitatív ismérvek beépíthetőek • nem érzékeny a korrelációra, a hiányzó és kiugró értékekre • valódi előrejelző eljárás	• fennáll a túltanulás veszélye • nagy mintán hatékony • nehezebben értelmezhető, bonyolultabb súlyok • alkalmazása komolyabb módszertani tudást, több mérést igényel

10. táblázat: Az egyes modellezési módszerek összehasonlító táblázata²¹

Következtetésként leszögezhetjük: az egyes módszerek közötti választás nem lehet egyértelmű mindössze az előnyök és hátrányok áttekintése alapján. Különböző

²¹ Halas (2004), Balcaen-Ooghe (2004) és saját összeállítás alapján

feladatok megoldására, különböző adatbázisokon végrehajtandó elemzésekre más-más módszertani választás hozhat jobb eredményt.

A szakirodalomban fellelhető módszertani összehasonlítások sem állítanak fel egyértelmű sorrendet, szinte minden módszer mellett felsorakoztathatók olyan tanulmányok, amelyek egyértelműen a legpontosabbnak minősítik, és olyanok is, amelyeknél alul maradnak más modellekkel szemben.

A legcélravezetőbb mindenképp az a megközelítés lehet, ha a konkrét adatok és modellezési feladat ismeretében kijelöljük a végrehajtásra alkalmas modellek körét, és több számítás elvégzése után a legjobb teljesítményt nyújtó modellt véglegesítjük.

4.2. A módszerek értékelésére használt technikák²²

4.2.1. Klasszifikációs mátrix

A csődmodelleket a szakirodalomban általában a besorolási pontosságukkal és a hibaarányaikkal jellemzik a szerzők. Ezek alapja a klasszifikációs mátrix, amely a modell által előrejelzett csoporttagságot veti össze a tényleges csoporttagsággal.

Az általános klasszifikációs mátrix a következőképp ábrázolható:

		Tényleges csoport	
		csődbe jutott	túlélő
Előrejelzett csoport	csődveszélyes	TP helyes besorolás (csőd)	FP elsőfajú hiba
	túlélő	FN másodfajú hiba	TN helyes besorolás (túlélő)

11. táblázat: Klasszifikációs mátrix csődmodellekre értelmezve

A besorolási pontosságot a helyes csoportba sorolás mutatja, vagyis a csődbe ment és a túlélő vállalatokra vonatkozó helyes előrejelzések (másképpen true positive vagy TP, illetve true negative, TN) összege adja.

²² A fejezetben összefoglalt eredményeket Imre (2007b) is tartalmazza.

Elsőfajú hibának (type I error, vagy false positive, FP) azt tekintik, amikor a modell csődveszélyesnek minősít egy vállalatot, amely valójában túlélő.

Másodfajú hibáról (type II error, false negative, FN) akkor beszélünk, amikor a modell túlélőnek feltételez egy vállalatot, miközben az csődbe ment.

A mátrix segítségével természetesen több más mérőszám is előállítható. A leggyakoribbak a következők²³:

- TPR (true positive rate, vagy találati arány, érzékenység)

$$TPR = \frac{TP}{TP + FN}$$

- FPR (false positive rate, vagy téves riasztási arány)

$$FPR = \frac{FP}{FP + TN}$$

- PPV (positive predictive value, pozitív előrejelzési érték vagy pontosság)

$$PPV = \frac{TP}{TP + FP}$$

- NPV (negative predictive value, negatív előrejelzési érték)

$$NPV = \frac{TN}{TN + FN}$$

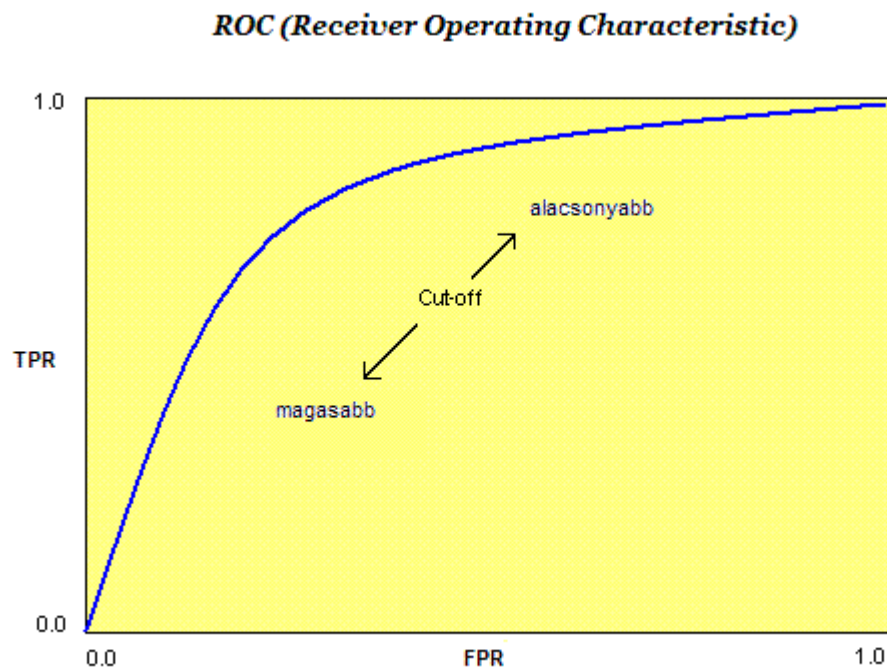
A klasszifikációs mátrix a modellek teljesítményének megítélésére alkalmazott módszerek közül kétségtelenül a legegyszerűbb. Mivel a korai csődmodell irodalomban szinte kizárólag ezzel mérték a modellek teljesítményét, és ezt publikálták, a klasszifikációs mátrix az egyetlen eszköz, amely lehetővé teszi a későbbi modellek összemérését az elődjeikkel. Ezért is maradt mind a mai napig a csődmodellek legelterjedtebb értékelési módszere.

²³ Fawcett (2004)

4.2.2. ROC (Receiver Operating Characteristic)

Az ROC görbét először a II. világháborúban használták radarjelzések elemzésére. Alkalmazása az 50-es évektől terjedt el más területeken is: elsősorban az orvostudomány egyes ágaiban (pszichofizika, radiológia, epidemiológia, stb), illetve osztályozási algoritmusok (amilyenek például a csődmodellek) értékelésénél is.²⁴

A ROC a modellek megkülönböztető képességének grafikus megjelenítése. Két dimenziója a fentiekben ismertetett FPR (false positive rate, azaz téves riasztási arány) és TPR (true positive rate, azaz találati arány). Eredménye az alábbi ábrán látható:



5. ábra: Receiver Operating Characteristic (ROC) grafikon

²⁴ Forrás: wikipedia

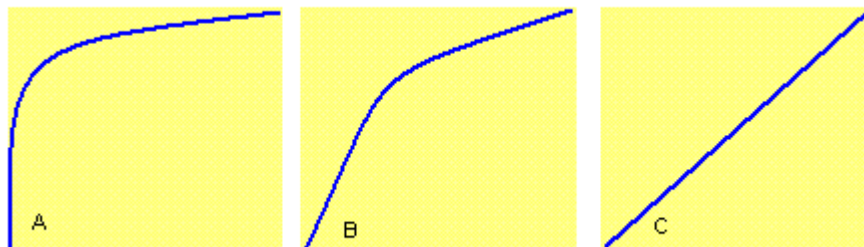
A görbe minden egyes pontja egy lehetséges cut-off értéket jelent, amelyre vonatkozóan ábrázolhatók a TPR és FPR értékek. Az ábráról leolvasható, hogy:

- Az (1;0) és a (0;0) pontok azokat a szélsőséges eseteket reprezentálják, amikor a modell megkülönböztetés nélkül sorolja a fizetéseképtelen illetve a túlélő kategóriába a megfigyelt vállalatokat.
- A teljesen véletlenszerű klasszifikációt jelentő grafikon megegyezik az $y=x$ egyenessel. Az ilyen modell szétválasztó képessége gyakorlatilag zérus.
- A tökéletes görbe átmegy a (0;1) ponton.

A valóságos helyzeteket leíró görbék az ideális és a véletlenszerű között helyezkednek el. Ilyenkor a modell teljesítményét az mutatja meg, hogy mennyire közelít a görbe a grafikon bal felső sarkához, azaz a tökéletes besoroláshoz. Ez a grafikon alatti terület nagyságával mérhető, amelynek a szakirodalomban elterjedt jelölései az AUC (Area Under the Curve) vagy AUROC (Area Under ROC).

Az AUC értéke 50 és 100 közé esik, ahol az előbbi érték a teljesen véletlenszerű csoportba sorolást, vagyis a modell alkalmatlanságát jelzi, míg az utóbbi a tökéletes besorolási pontosságú modellt jelenti. A gyakorlatban úgy tekintik, hogy a 70% körüli AUC esetén a modell teljesítménye elfogadható, 80-90% körül pedig már kiemelkedőnek értékelhető.

A modell előrejelzési erejének megértését segítik az alábbi ROC grafikonok.



6. ábra: Néhány lehetséges ROC-görbe

Az „A” grafikon egy jó előrejelzési pontosságú modell ROC-görbáját mutatja. A cut-off értékek széles köre képes kellő pontossággal szétválasztani a csődveszélyes és a túlélő vállalatokat.

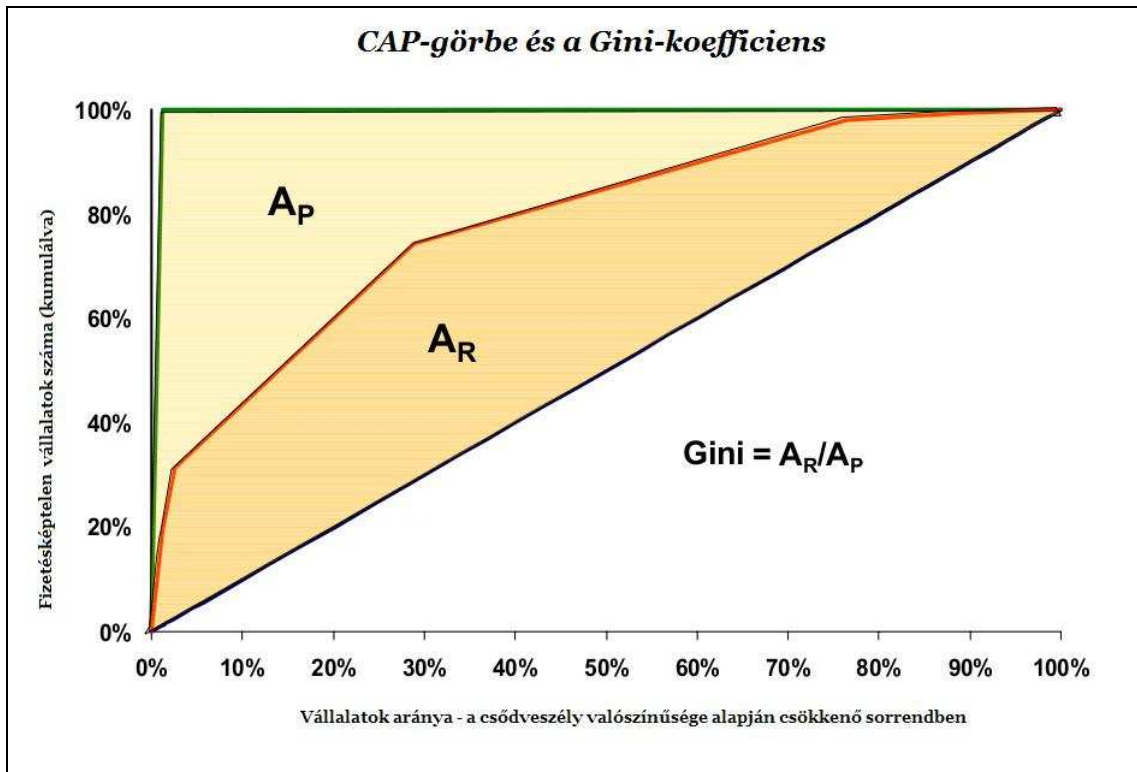
A „B” ábra egy elfogadható modell eredményét szemlélteti. A cut-off egy racionális kompromisszumot jelent a TPR és az FPR arányainak kiegyensúlyozott rögzítésére. Magasabb cut-off esetén alacsonyabb FPR arány is elérhető lenne, de szintén csökkenő TPR mellett. A cut-off csökkentése pedig ezzel ellentétes irányban hat.

A „C” grafikon gyenge minőségű modellt jellemez. A lehetséges cut-off értékek széles köre ad nagyságrendileg ugyanolyan gyenge besorolási képességgel rendelkező modellt.

4.2.3. Gini-koefficiens

A Gini-koefficiens kiindulási alapja a CAP (Cumulative Accuracy Profile, kumulált megfeleléségi profil) grafikon. A CAP a ROC-görbéhez hasonlóan az előrejelzési modellek megkülönböztető erejének ábrázolására szolgál.

A grafikont két dimenzió mentén ábrázoljuk. Az x-tengelyen a csőd bekövetkezésének előrejelzett valószínűségét (pl. a diszkriminancia analízis vagy a regressziós modell függvényértékét) jelenítjük meg. Kiindulásként tehát az előrejelzett csőd-valószínűségek alapján csökkenő sorrendbe állítjuk a vizsgált vállalatokat. Az y-tengelyen ezután az egyes valószínűségekhez tartozó kumulált gyakoriságokat, vagyis az adott valószínűségnél magasabb csődveszélyességű vállalatok számát ábrázoljuk.



7. ábra: A CAP-görbe és a Gini-koefficiens²⁵

A fenti ábráról például az olvasható le, hogy a modell alapján, az összes vizsgált vállalat között a 30% leginkább csődvesztélyesnek minősített cég lefedi a ténylegesen csődöt jelentő vállalatok 70%-át. A csőd valószínűsége alapján a legvesztélyesebb 60% pedig már magában foglalja az összes csődbement vállalat 90%-át.

A CAP-görbe értékelésekor hasonló kitüntetett esetek figyelhetők meg, mint az ROC esetében:

- A teljesen véletlenszerű klasszifikációt jelentő grafikon itt is megegyezik az $y=x$ egyenessel. Az ilyen modell szétválasztó képessége gyakorlatilag zérus.
- A tökéletes görbe átmegy a (0;1) ponton.

A valóságos helyzeteket leíró görbék természetesen a CAP esetén is az ideális és a véletlenszerű között helyezkednek el. A modell teljesítményét itt is az mutatja meg,

²⁵ Glennon (2006) alapján

hogy mennyire közelít a görbe a grafikon bal felső sarkához, azaz a tökéletes besoroláshoz.

A grafikusán értékelés mellett a CAP-görbe eredménye is kifejezhető egyetlen mérőszám segítségével, amely a véletlenszerű szétosztás egyenese és a tökéletes szétválasztás pontja alapján meghatározható háromszög területe (az ábra szerint A_P), illetve a véletlenszerű szétosztás egyenese és a CAP-görbe által határolt terület (az ábrán A_R) hányadosa alapján képezhető. Ezt nevezik Gini-koefficiensnek.

Az ábrán látható jelöléseket használva tehát:

$$Gini = \frac{A_R}{A_P}$$

A Gini-koefficiens értéke 0 és 1 között lehet, amelynek általában százalékos formában használnak. Minél közelebb van a Gini értéke a 100%-hoz, annál jobb a modell megkülönböztető ereje. A gyakorlatban a 40-50% közötti Gini-koefficienssel rendelkező modelleket már elfogadhatónak értékelik, 70% fölötti érték esetén pedig már kifejezetten erős modelltől beszélhetünk.

A fentiek alapján látható, hogy a ROC és a CAP-görbék hasonlóképp viselkednek. A Gini-koefficiens és az AUC között létezik egy egyszerű lineáris kapcsolat, amelynek segítségével egyikből előállítható a másik, és fordítva:

$$Gini = 2 AUC - 1$$

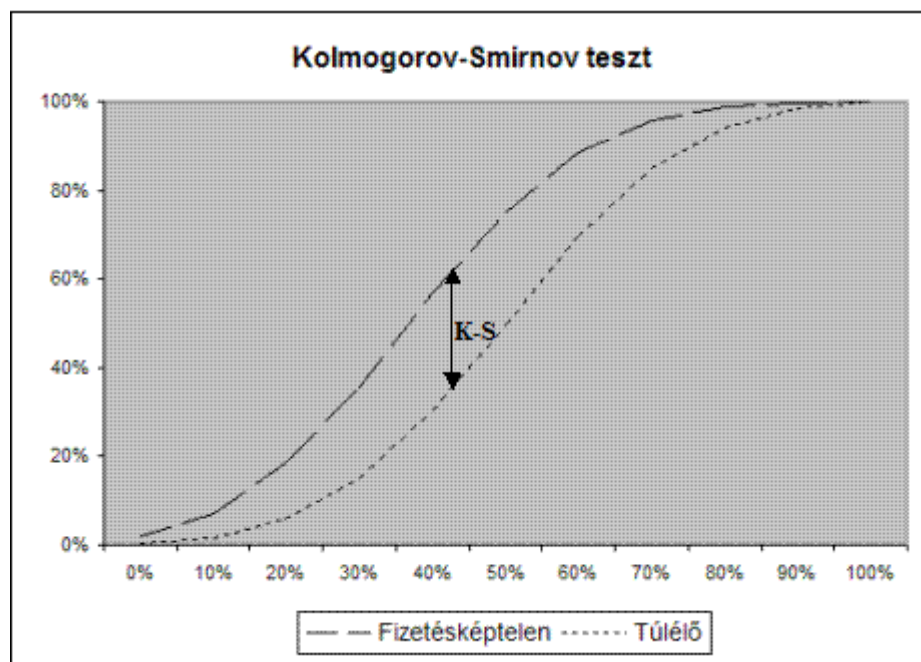
Mind az AUC, mind a Gini-koefficiens igen elterjedt mérőszámai az előrejelzési modellek megkülönböztető erejének, mindkettőt széles körben alkalmazzák a gyakorlatban is.

4.2.4. Kolmogorov-Smirnov teszt (K-S)

Érdemes még megemlíteni a modellek értékelésére használt eszközök között a Kolmogorov-Smirnov (K-S) tesztet is, bár ez kevésbé elterjedt, mint a fenti két módszer.

Ebben az esetben is a modell megkülönböztető erejét mérjük, és a CAP-hoz hasonlóan az előrejelzett csőd-valószínűségek alapján csökkenő sorrendbe állított vállalatok adatsorából indulunk ki.

Az x-tengelyen itt is a csőd bekövetkezésének előrejelzett valószínűségét (pl. a diszkriminancia analízis vagy a regressziós modell függvényértékét) jelenítjük meg, az y-tengelyen pedig mind a fizetéseképtelen, mind a túlélő vállalatok kumulált számát ábrázoljuk:



8. ábra: A Kolmogorov-Smirnov (K-S) teszt

A K-S mérőszámát úgy kapjuk, hogy a két görbe közötti legnagyobb távolságot vesszük. Minél nagyobb ez a távolság, a vizsgált modell annál jobban képes szétválasztani egymástól a fizetéseképtelen és a túlélő vállalatokat.

A K-S értéke 0 és 100% között lehet, a gyakorlatban az elfogadási határa 20%. Amennyiben a K-S 30-40% körül alakul, a modell teljesítménye jónak mondható, 60% fölött pedig kiválónak minősíthető.

Amiért ez a teszt mégis kevésbé elterjedt, mint az AUC és a Gini-koefficiens, az elsősorban a közelítő jellege. A K-S ugyanis csak a legnagyobb távolságot fejezi ki a két grafikon között, de nem veszi figyelembe, hogy ez mely ponton mérhető. Nem mindegy ugyanis, hogy a cut-off környékén, vagy valamelyik szélső tartományban teljesít jól a modell.

Ezért is alkalmazzák a K-S tesztnél szélesebb körben a ROC és CAP-görbéket, valamint az azokból képezhető mérőszámokat, mivel azok általánosabban jellemzik a modelleket.

5. A prediktív modellek kidolgozása

5.1. Csőd vagy default? Definíciós kérdések

5.1.1. *A csődmodellek leggyakoribb kritikái, problémái*

A csődmodellek kritikusai több általános érvényű problémát fogalmaztak meg, amelyek részben a modellezési munkát, részben pedig az alkalmazott definíciókat érintik.

Első és talán legfontosabb a bináris célváltozó, vagyis a mindössze két lehetséges kimenetet figyelembe vevő osztályozás, a fizetéseképtelen és a túlélő kategóriák alkalmazása. Ez pedig nyilvánvalóan durva leegyszerűsítése a valóságnak, hiszen a két végletes állapot között rengeteg köztes pénzügyi helyzet megfigyelhető.

Különösen fontos eleme a két csoportot alkalmazó besorolásnak, hogy maga a csőd jogi fogalom, így a különböző jogrendszerek miatt nemzetközileg nehezen összehasonlíthatóak a különböző modellek, eredményeik is nehezen interpretálhatók más országok vállalataira, piacaira.

Problémát jelent az a tény is, hogy a jogi értelemben vett csőd nem feltétlenül esik egybe a pénzügyi csőddel. A pénzügyi nehézségekkel küzdő cég még húzhatja az időt, és megpróbálhat javítani helyzetén, mielőtt hivatalosan csődöt jelentene. Így előfordulhat, hogy olyan cégek, amelyek profiljukban inkább a csődbe jutott vállalatokkal mutatnak hasonlóságot, jogi helyzetük miatt mégis a túlélők közé kerülnek, amely természetesen a modellépítésnél is torzításhoz vezethet.

Sokan kérdőjelezik meg maguknak a számviteli-pénzügyi mutatóknak a felhasználhatóságát, illetve azok megbízhatóságát is. A számviteli rendszerek torzító tényezőinek, a számviteli kimutatások és az azokból származtatott mutatók hiányosságainak irodalma igen széles, a kérdéskör részletes tárgyalása önmagában is felölelhetne egy külön tanulmányt. A jelen kutatás szempontjából elég annyit leszögezni, hogy egyrésről ezek az adatok törvényi kötelezettségből fakadó és törvényi szabályozásoknak eleget tevő beszámolókból származnak. Másrésről vitathatatlanul jelen van bennük az egyszerűsítés, következésképp a torzítás, és kétségtelen az is, hogy

a kritikusok által megfogalmazott érvek egyáltalán nem alaptalanok. Mindazonáltal egy csődmódel más, objektív adatforrásra nem nagyon tud támaszkodni, és a gyakorlatban alkalmazott modellek is keverve használják fel egyéb, nem számviteli (szubjektív, vagy másképp soft) faktorokkal kombinálva.

Az éves beszámolók továbbá mindössze egy pillanatfelvételt (snap-shot) mutatnak az adott vállalatról, amely a pénzügyi év lezárásakor állt fenn. Ugyanez lesz igaz a beszámolók alapján képzett mutatókra is, amely így nem veszi figyelembe azok időbeli változását, csak statikus képet fest a vállalatról.

A mutatók időállósága is fontos, hiszen amint nagyobb változás áll be akár a makrogazdasági környezetben, akár a számviteli szabályozásban, az addigi mutatókon alapuló módel nem lesz alkalmazható a megváltozott körülmények között. Ennek az esélye természetesen kisebb, de hosszabb távon számolni lehet vele.

Ahhoz, hogy a modellek előrejelzésre is felhasználhatók legyenek, olyan fejlesztési mintát követelnek meg, amely reprezentatív az egész gazdaságra és cégeire vonatkozóan. Ennek a csődmódellek többsége valóban nem tesz eleget, pedig a módel általános kiterjesztése és előrejelző képessége szempontjából lényeges kérdés.

5.1.2. Egy lehetséges megoldás: Bázeli II. definíciókon alapuló default-modellek

A fenti kritikák figyelembe vételével alakítottam ki koncepciómat, amelyben a csődmódellek logikáján alapuló vállalati prediktív modellek kialakítását tűztem ki célul, de más definíciók alapján.

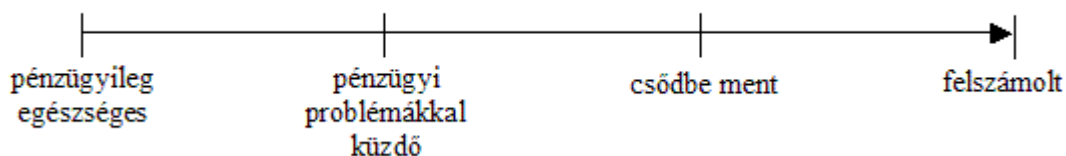
A magyar gazdaság egészére értelmezett Virág Miklós és szerzőtársai több általános csődmódellet is kifejlesztettek már. Az általam kidolgozott modellek ugyanakkor nem a jogi értelemben vett csőd előrejelzését szolgálják, hanem a csőd előtti, pénzügyi nehézségekkel terhelt szakaszát próbálják megfogni és előrejelezni.

Segítségemre volt a bankok életében nagy szerepet játszó Bázeli Tőkeegyezmény, vagyis a Bázeli II. Az új szabályozás megjelenése ugyanis kutatások és fejlesztések sorát indította el a kockázatkezelés, azon belül is az előrejelzési modellek terén. Az egyezmény ugyanakkor nem a csőd, hanem a hitelkockázatok kezelésénél fontos, ún.

default előrejelzését és számszerűsítését követeli meg, amelyet alapvetően a 90 napon túli fizetési késedelembe eséssel határoz meg.

Fontos leszögezni tehát, hogy nem egy újabb csődmodellről van szó, hanem olyan modellről, amelynek célváltozója a default, azaz a nemfizetés, vagy pénzügyi értelemben vett csőd, inszolvencia.

Ezt a különbséget legszemléletesebben Stickney (1990) illusztrálja, amikor a vállalatok pénzügyi problémáinak kialakulását és elmérgesedését mint egy betegség kialakulását és lefolyását ábrázolja, a következő kontinuum mentén:



9. ábra: Vállalatok pénzügyi „egészségi állapotának” romlása²⁶

A default ténye tehát a fenti ábrán a pénzügyi problémákkal küzdő vállalatot jelzi. A csőd előrejelzésével szemben előnye, hogy a pénzügyi problémák egy korábbi fázisánál jelez a döntéshozónak, aki így korábban felkészülhet az esetleges intézkedésekre, vagy hitelfelvétel esetén a magasabb kockázat miatt megtagadhatja a hitel folyósítását.

Modellezési szempontból viszont pont ez jelentheti a nagyobb nehézséget: mivel a pénzügyileg egészséges és a problémás vállalatok profilja valószínűleg még nem különbözik olyan markánsan, mint a csődbe ment és a túlélő vállalatoké, az előrejelzés is valószínűleg nehezebb, pontatlanabb lesz. Ezt a hipotézist vizsgálom meg a modellépítés eredményei alapján.

Tekintsük át most táblázatos formában, hogy milyen megoldásokat adhat a csődmodellek kritikáira egy Bázeli II. definíciókon alapuló default-modell!

²⁶ Stickney (1990)

A csődmodellek említett kritikái	Az általam kidolgozott default-modellek nyújtotta megoldások
a csőd fogalmának jogi megközelítése	egyértelműen gazdasági alapú megközelítés: a 90 napos késedelem előrejelzése által
nehéz összehasonlíthatóság a különböző jogi szabályozások miatt	egyértelmű összehasonlíthatóság, az egyértelmű és könnyebben megfogható definíció miatt
számveteli-pénzügyi mutatók alkalmazása	kombinált minősítési eljárás: a számveteli-pénzügyi mutatók (az ún. hard faktorok) mellett kvalitatív ismérvek (ún. soft faktorok) bevonása az elemzésbe
éves beszámolók pillanatfelvétel-jellegű adatira támaszkodó vizsgálat	tesztelés dinamikus mutatók felhasználásával: a mutatók két időszak közötti változását is input tényezőként alkalmazva
korlátozott felhasználhatóság előrejelzési célokra, a minta reprezentativitásának hiánya miatt	a modellezésnél olyan adatbázissal dolgoztam, amely mind a vállalatok árbevétele, mind iparági megoszlásuk szempontjából reprezentálja a hazai nemzetgazdaságot

12. táblázat: A csődmodellek kritikái és a javasolt modell megfelelő megoldásai

Amint látható, a korábbiakban összefoglalt főbb kritikai észrevételek közül egyedül a bináris célváltozót nem tudtam tovább finomítani, amelynek elsődleges oka a megfelelő országos szintű adatbázis hiánya. A további kutatásoknak mindenképpen ez lehet az egyik iránya.

5.2. A modellezésnél használt adatbázis felépítése

5.2.1. A modellezés alapját képező minta

A modellek kialakításához használt adatbázist az Info-Datex Kft. bocsátotta rendelkezésemre, amely 1000 defaultos és 1000 kontroll vállalatot tartalmaz.

A Bázeli II. direktíva kétféleképpen határozza meg a default bekövetkezését:

- 90 napos fizetési késedelem esetén, illetve
- a minősítő bank úgy véli, hogy a rendelkezésre álló információk alapján a tartozás visszafizetése kétségesse válik.

Az utóbbi kategória már a meghatározásból fakadóan is szubjektív döntést takar, így az adatbázis kialakításánál nem lehetett figyelembe venni. A 90 napos késedelem ugyanakkor gazdasági jellegű definíció, jól körülhatárolható és objektív feltétel, ezen alapul a felépített adatbázis is.

Mivel országos szintű adatbázist kellett kialakítani, a 90 napos fizetési késedelemnél a BISZ Rt- által üzemeltetett KHR (Központi Hitelinformációs Rendszer) jelentett megoldást. Ez a központi adatbázis a bankok jelentési kötelezettsége alapján töltődik, és minden, banki és hitelintézeti kölcsön adatait tartalmazza. Az adatbázis nyilvántartja az egyes hitelekhez kapcsolódó fizetési státuszokat is. Egy ügyfél akkor minősül problémásnak, ha tartozása meghaladja a 30 napot. Defaultról tehát akkor beszélhetünk, ha az ügyfél a rendszerben már legalább 60 napja szerepel rossz adósként.

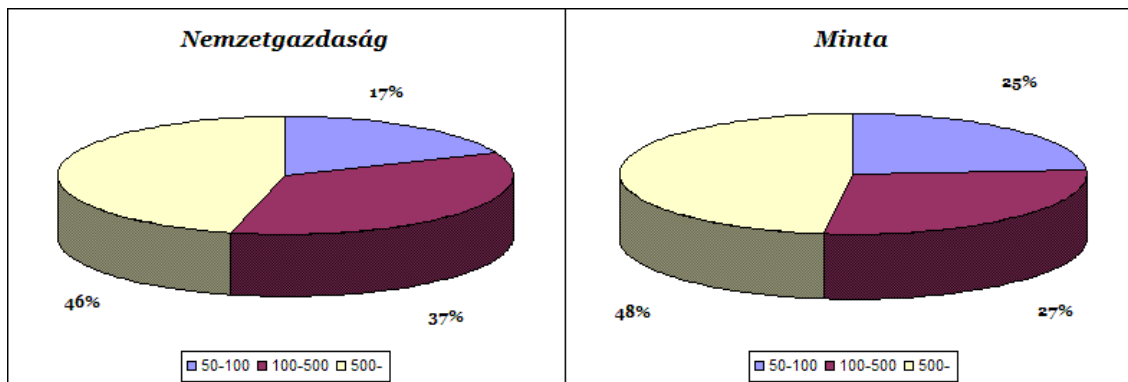
Az adatbázis a fenti információkat 5 éves időintervallumra vonatkozóan vette figyelembe, vagyis 2002 és 2006 közötti defaultos vállalatok, illetve ugyanerre az időszakra vonatkozóan default esemény nélküli cégek adataiból épül fel.

Mivel a bankok hitelezési gyakorlatára is jellemző, hogy megállapítanak egy minimális árbevételi határt, amelynek teljesítését elvárják a cégektől, ezért a mintába csak olyan vállalatok kerültek bele, amelyek legalább 50 millió Ft-os árbevételt értek el az utolsó két lezárt év valamelyikében.

5.2.2. A minta reprezentativitása

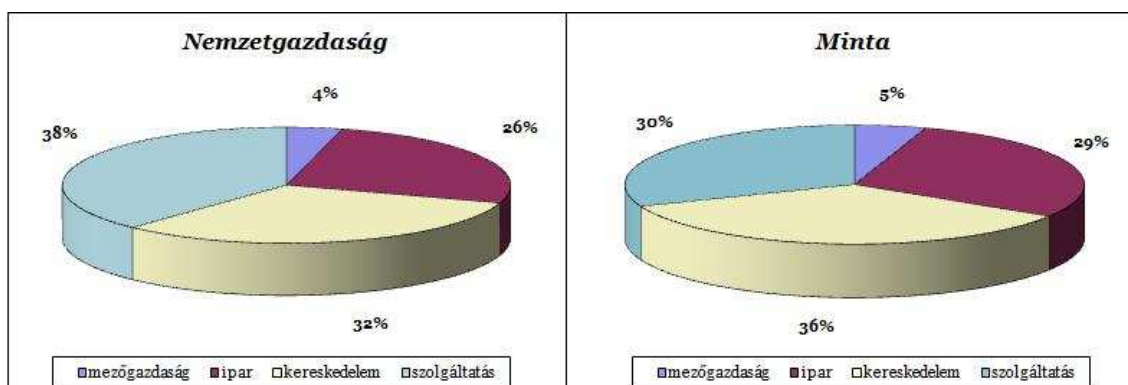
A fejlesztési mintát úgy kellett kialakítani, hogy az minél szélesebb körben alkalmas legyen az előrejelzésre. A minta két szempont alapján képezi le a hazai nemzetgazdaságot: a vállalatok árbevétele és iparági megoszlásuk alapján.

Árbevétel szerinti megoszlás tekintetében összehasonlítva a mintát a nemzetgazdaság egészének adataival:



10. ábra: A fejlesztési minta és a nemzetgazdaság megoszlása árbevétel alapján

Ágazati megoszlás tekintetében is összevetve a mintát és az országos adatokat:



11. ábra: A fejlesztési minta és a nemzetgazdaság megoszlása gazdasági ágazatok alapján

A fentiek alapján tehát elmondhatjuk, hogy ha nem is teljes mértékben reprezentatív, a minta nagymértékben igazodik a nemzetgazdasági megoszlásokhoz.

5.2.3. Adattisztítás, adatkorrekció

A modellépítés egyik kulcsfontosságú tényezője az adatok minősége és megbízhatósága. Ennek elérése érdekében számos problémát kell megoldani.

A modellépítési adatbázis esetében több ilyen adatkorrekciós lépés vált szükségessé:

Pontatlan beszámolók kiszűrése

A beszámolók pontatlanságai az egyes mérlegsorok összefüggései alapján általában ellenőrizhetők (pl. mérlegfőösszeg-egyeztetés, stb). A pontatlan beszámolóval rendelkező cégek adatai nem kerültek a mintába.

Törtévi mérlegek törlése

Mivel törtévi mérlegek nem hasonlíthatók össze a teljes pénzügyi évet záró cégek beszámolóival, ezen vállalatok adatai szintén kimaradtak a mintából.

Negatív saját tőke értékek

A saját tőkénél szereplő negatív értékek nemcsak ideiglenes állapotot tükröznek (mivel a vállalatok törvényi kötelezettsége a tulajdonosi tőkepótlás), hanem torzításhoz is vezethetnek azon mutatók számításánál, ahol ez osztóként szerepel: pl. veszteséges gazdálkodás esetén a ROE számításánál a negatív saját tőke – hibásan – pozitív rátát eredményez. Az egyértelműség miatt ezek a vállalatok sem kerültek be modellezési adatbázisba.

Hiányzó értékek kezelése

Az adatbázis esetén egyetlen esetben vált szükségessé a hiányzó értékek helyettesítése: az adósságszolgálati mutatók esetében. A gondot az osztóban szereplő fizetett kamatok jelentették, ugyanis ez a tétel nem szerepel az egyszerűsített beszámolóknak, így nagy arányban szerepelt nulla érték. Ennek helyettesítésére a mérlegstruktúrában egyvel feljebb szereplő pénzügyi ráfordítások sor adata lehetett egyedül alkalmas.

Kiugró (extrém) értékek kezelése

Mivel egy-egy kiugró érték már torzítást eredményezhet az erre érzékeny modellezési módszereknél, az ilyen értékeket is ki kell szűrni az adatokból. A modellezési adatbázis esetén ez felső és alsó 10-10%-nál való szélsőérték-csonkolással történt.

5.2.4. A modellépítésnél vizsgált mutatók köre

Amint a fejezet elején látható volt, a modellezésnél kétfajta döntési faktorról dolgoztam:

- objektív faktorokkal, azaz pénzügyi-számviteli mutatókkal, illetve
- szubjektív kritériumokkal, vagyis a vállalkozások egyéb, kvalitatív ismérveivel.

5.2.4.1. Pénzügyi-számviteli mutatók (objektív, hard faktorok)

A modellezés során összesen 52 pénzügyi mutatót vizsgáltam, 7 mutatócsoportból:

- likviditási mutatócsoport,
- tőkeszerkezeti mutatócsoport,
- adósságszolgálati mutatócsoport,
- működőtőke mutatócsoport,
- eszközforgási mutatócsoport,
- termelékenység mutatócsoport,
- jövedelmezőségi mutatócsoport.

Az egyes mutatócsoportokból vizsgált mutatókat, kódjaikat és számítási módjaikat mutatják be az alábbi táblázatok.

<i>Likviditási mutatócsoport</i>		
Kód	Mutató	Számítási mód
m ₁	Likviditási ráta	Forgóeszközök / Rövid lejáratú kötelezettségek
m ₂	Likviditási gyorsráta	(Forgóeszközök - Készletek) / Rövid lejáratú kötelezettségek
m ₃	Pénzhányad	(Pénzeszközök + Likvid értékpapírok) / Rövid lejáratú kötelezettségek
m ₄	Dinamikus likviditás	Üzemi (üzleti) tevékenység eredménye / Rövid lejáratú kötelezettségek
m ₅	Időtartam mutató	(Forgóeszközök - Készletek) / ([Anyagjellegű ráfordítások + Személyi jellegű ráfordítások + Egyéb költségek + Egyéb ráfordítások + Fizetett kamatok és kamatjellegű kifizetések + Adófizetési kötelezettség] / 365)

13. táblázat: A vizsgált likviditási mutatók

<i>Tőkeszerkezeti mutatócsoport</i>		
Kód	Mutató	Számítási mód
m ₆	Tőkeellátottság	saját tőke / összes eszköz
m ₇	Eladósodottság aránya	kötelezettségek / összes eszköz
m ₈	Idegen tőke / saját tőke	kötelezettségek / saját tőke
m ₉	Hosszú távú eladósodottság	hosszú lejáratú kötelezettségek / (hosszú lejáratú kötelezettségek + saját tőke)
m ₁₀	Kötelezettségek / tárgyi eszközök	kötelezettségek / tárgyi eszközök
m ₁₁	Kötelezettségek / dologi nettó érték	kötelezettségek / (saját tőke - immateriális javak)

14. táblázat: A vizsgált tőkeszerkezeti mutatók

Adósságszolgálati mutatócsoport		
Kód	Mutató	Számítási mód
m ₁₂	Kamatfedezet I.	üzemi (üzleti) tevékenység eredménye / fizetett kamatok és kamatjellegű kifizetések
m ₁₃	Kamatfedezet II.	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek) / fizetett kamatok és kamatjellegű kifizetések
m ₁₄	EBITDA fedezet I.	(üzemi tevékenység eredménye + amortizáció) / fizetett kamatok és kamatjellegű kifizetések
m ₁₅	EBITDA fedezet II.	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek + amortizáció) / fizetett kamatok és kamatjellegű kifizetések
m ₁₆	Cash-flow fedezet	(adózott eredmény + amortizáció) / fizetett kamatok és kamatjellegű kifizetések
m ₁₇	Adósságvisszafizető-képesség	(adózott eredmény + amortizáció) / kötelezettségek

15. táblázat: A vizsgált adósságszolgálati mutatók

Működőtőke mutatócsoport		
Kód	Mutató	Számítási mód
m ₁₈	Működőtőke ellátottság	(forgóeszközök - rövid lejáratú kötelezettségek) / összes eszköz
m ₁₉	Működőtőke forgása	értékesítés nettó árbevétele / átl.(forgóeszközök - rövid lejáratú kötelezettségek)
m ₂₀	Vevőállomány / szállítóállomány	vevőállomány / szállítóállomány
m ₂₁	Rövid lejáratú kötelezettségek/ dologi nettó érték	rövid lejáratú kötelezettségek / (saját tőke - immateriális javak)

16. táblázat: A vizsgált működőtőke mutatók

<i>Eszközforgási mutatócsoport</i>		
Kód	Mutató	Számítási mód
m ₂₂	Összes eszköz forgási időtartama	átl. összes eszköz / (értékesítés nettó árbevétele / 365)
m ₂₃	Készletek forgási időtartama I.	átl. készletek / (anyagjellegű ráfordítások / 365)
m ₂₄	Készletek forgási időtartama II.	átl. készletek / (értékesítés nettó árbevétele / 365)
m ₂₅	Beszedés időtartama	átl. követelések / (értékesítés nettó árbevétele / 365)
m ₂₆	Vevőállomány forgási időtartama	átl. vevőállomány / (értékesítés nettó árbevétele / 365)
m ₂₇	Szállítóállomány forgási időtartama	átl. szállítóállomány / (anyagjellegű ráfordítások / 365)

17. táblázat: A vizsgált eszközforgási mutatók

<i>Termelékenység mutatócsoport</i>		
Kód	Mutató	Számítási mód
m ₂₈	1 alkalmazottra eső nettó árbevétel	értékesítés nettó árbevétele / alkalmazottak átlagos létszáma
m ₂₉	1 alkalmazottra eső adózott eredmény	adózott eredmény / alkalmazottak átlagos létszáma
m ₃₀	1 alkalmazottra eső üzemi (üzleti) eredmény	üzemi (üzleti) tevékenység eredménye / alkalmazottak átlagos létszáma
m ₃₁	1 alkalmazottra eső EBITDA	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek + amortizáció) / alkalmazottak átlagos létszáma

18. táblázat: A vizsgált termelékenységi mutatók

<i>Jövedelmezőségi mutatócsoport</i>		
Kód	Mutató	Számítási mód
m ₃₂	Saját-tőke-arányos adózott eredmény (ROE)	adózott eredmény / átl. saját tőke
m ₃₃	Árbevétel-arányos adózott eredmény (ROS)	adózott eredmény / értékesítés nettó árbevétele
m ₃₄	Eszközarányos adózott eredmény (ROA)	adózott eredmény / átl. összes eszköz
m ₃₅	Adózott eredmény / működőtőke	adózott eredmény / átl. (forgóeszközök - rövid lejáratú kötelezettségek)
m ₃₆	Adózott eredmény / dologi nettó érték	adózott eredmény / átl. (saját tőke - immateriális javak)
m ₃₇	Saját-tőke-arányos adózás előtti eredmény	adózás előtti eredmény / átl. saját tőke
m ₃₈	Árbevétel-arányos adózás előtti eredmény	adózás előtti eredmény / értékesítés nettó árbevétele
m ₃₉	Eszközarányos adózás előtti eredmény	adózás előtti eredmény / átl. összes eszköz
m ₄₀	Tevékenység haszonkulcsa I.	üzemi (üzleti) tevékenység eredménye / értékesítés nettó árbevétele
m ₄₁	Tevékenység haszonkulcsa II.	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek) / értékesítés nettó árbevétele
m ₄₂	Üzemi eredmény / saját tőke	üzemi (üzleti) tevékenység eredménye / átl. saját tőke
m ₄₃	Üzemi eredmény / összes eszköz	üzemi (üzleti) tevékenység eredménye / átl. összes eszköz
m ₄₄	EBIT / összes eszköz	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek) / átl. összes eszköz
m ₄₅	EBIT / dologi nettó érték	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek) / átl. (saját tőke - immateriális javak)
m ₄₆	EBITDA / összes eszköz	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek + amortizáció) / átl. összes eszköz
m ₄₇	Tőkearányos működési eredmény (ROCE) I.	üzemi (üzleti) tevékenység eredménye / átl. (saját tőke + hosszú lejáratú kötelezettségek)

m ₄₈	Tőkearányos működési eredmény (ROCE) II.	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek) / átl. (saját tőke + hosszú lejáratú kötelezettségek)
m ₄₉	Tőkearányos EBITDA	(szokásos vállalkozási eredmény + fizetett kamatok és kamatjellegű kifizetések - kapott kamatok, kamatjellegű bevételek + amortizáció) / átl. (saját tőke + hosszú lejáratú kötelezettségek)
m ₅₀	Befektetett tőke hozama (ROIC)	üzemi (üzleti) tevékenység eredménye*(1- társasági adókulcs) / átl. (összes forrás - szállítók - passzív időbeli elhatárolások)
m ₅₁	Nyereség-visszaforgatás aránya	mérleg szerinti eredmény / saját tőke
m ₅₂	Bruttó haszonrés I.	(értékesítés nettó árbevétele - anyagjellegű ráfordítások - személyi jellegű ráfordítások) / értékesítés nettó árbevétele

19. táblázat: A vizsgált jövedelmezőségi mutatók

5.2.4.2. Kvalitatív adatok (szubjektív, soft faktorok)

A gyakorlati életben a szubjektív vagy kvalitatív információknak széles körét alkalmazzák a bankok, például: a menedzsment szakértelme, a gazdálkodás színvonala, piaci ismertség, stb. Publikus adatbázisból azonban ilyenek nem nyerhetők ki, ezért helyettük a cég egyéb, nem kvantifikálható adatai sorolhatók ide.

Az felhasznált adatbázis esetén ezek a következők:

- földrajzi elhelyezkedés: megye és régió szintjén;
- iparági hovatartozás (TEÁOR főágazati kód alapján): mezőgazdaság, ipar, kereskedelem, szolgáltatás;
- cégforma: kft, bt, rt, stb.

5.3. A mutatók előzetes áttekintése

Az adatbázis elemzésének első lépéseként érdemes rövid áttekintést készíteni a mutatókról alapvető leíró statisztikák segítségével. Ez a lépés hasonló a kezdeti csődmódel-kutatások technikájához, vagyis azon mutatók meghatározását jelenti,

amelyek átlagos értékei szignifikánsan különböznek a csődbe jutott és a működő vállalatok csoportjai esetén. Esetünkben ez a 90 napos késedelembe eső vállalatok és a kontrollcsoport adatai közötti különbözőséget legjobban tükröző mutatók kiválasztását jelenti.

Az 52 mutató esetén az egyes csoportokhoz tartozó átlagokat, mediánokat és a mutatók szórását vizsgáltam. Alapvetően olyan mutatókat kerestem, amelyek középértékei legjobban különböznek a két csoport esetén, azaz kellőképpen alkalmasak a két csoport szétválasztására. Fontos tényező az is, hogy a mutató szórása kellőképpen nagy legyen, vagyis az értékek ne egy kis sávban szóródjanak, hiszen ilyenkor az adott mutató a szétválasztásra egyáltalán nem lehet alkalmas.

A mutatók áttekintése után hét mutatót tudtam kiválasztani, amelyek leginkább megfelelnek a fenti kritériumoknak:

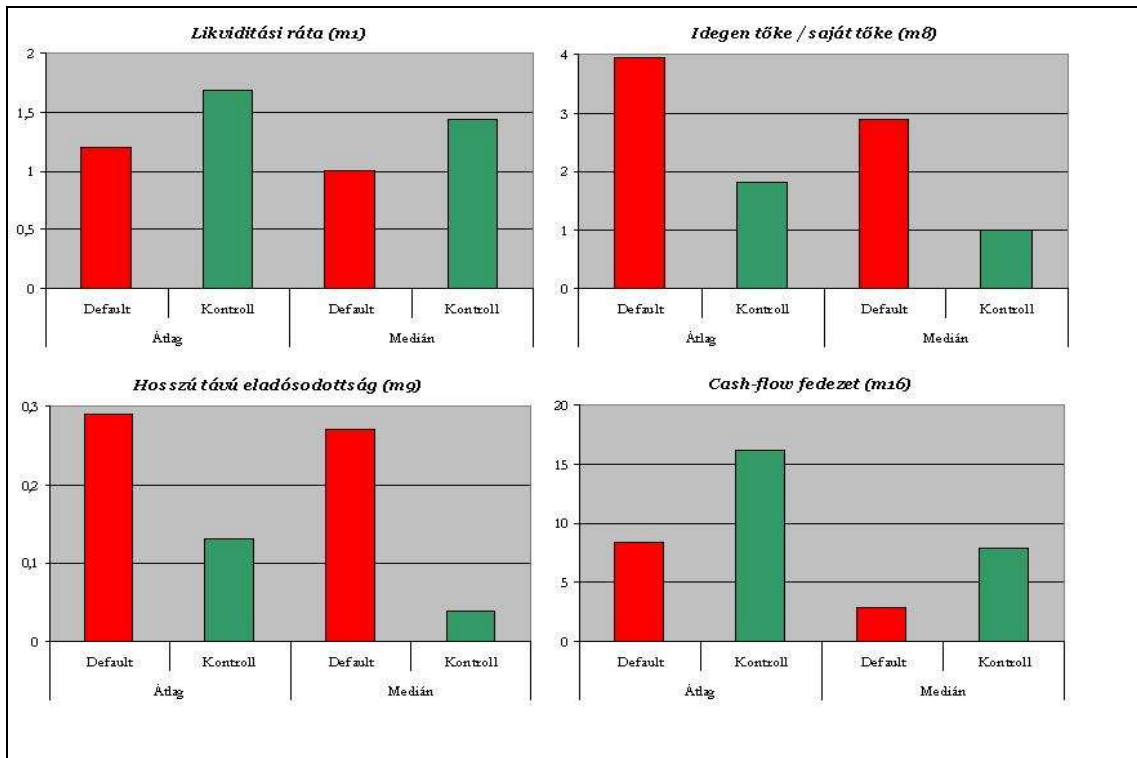
- A likviditási mutatók közül a likviditási rátát (m_1) találtam a legerősebb magyarázó erejűnek.
- A tőkeszerkezeti mutatók közül több mutató is alkalmasnak tűnt a csoportok megkülönböztetése szempontjából, végül úgy döntöttem, hogy két mutatót, az idegen tőke / saját tőke (m_8) és a hosszú távú eladósodottság (m_9) rátákat emelem ki közülük.
- Az adósságszolgálati mutatócsoportból a cash-flow fedezet (m_{16}) mutató tűnt a legalkalmasabbnak a csoportok szétválasztására.
- A működőtőke csoportból a legerősebbnek a működőtőke-ellátottságot (m_{18}) lehet kiemelni.
- Az eszközfordítási mutatók közül kiemelkedik a szállítóállomány forgási időtartama (m_{27}) mutató.
- A jövedelmezőségi mutatócsoportból az adózott eredmény / működőtőke (m_{35}) mutató tűnt legalkalmasabbnak a két csoport szétválasztására.

A fentiekben kiválasztott mutatók leíró adatait foglalja össze az alábbi táblázat.

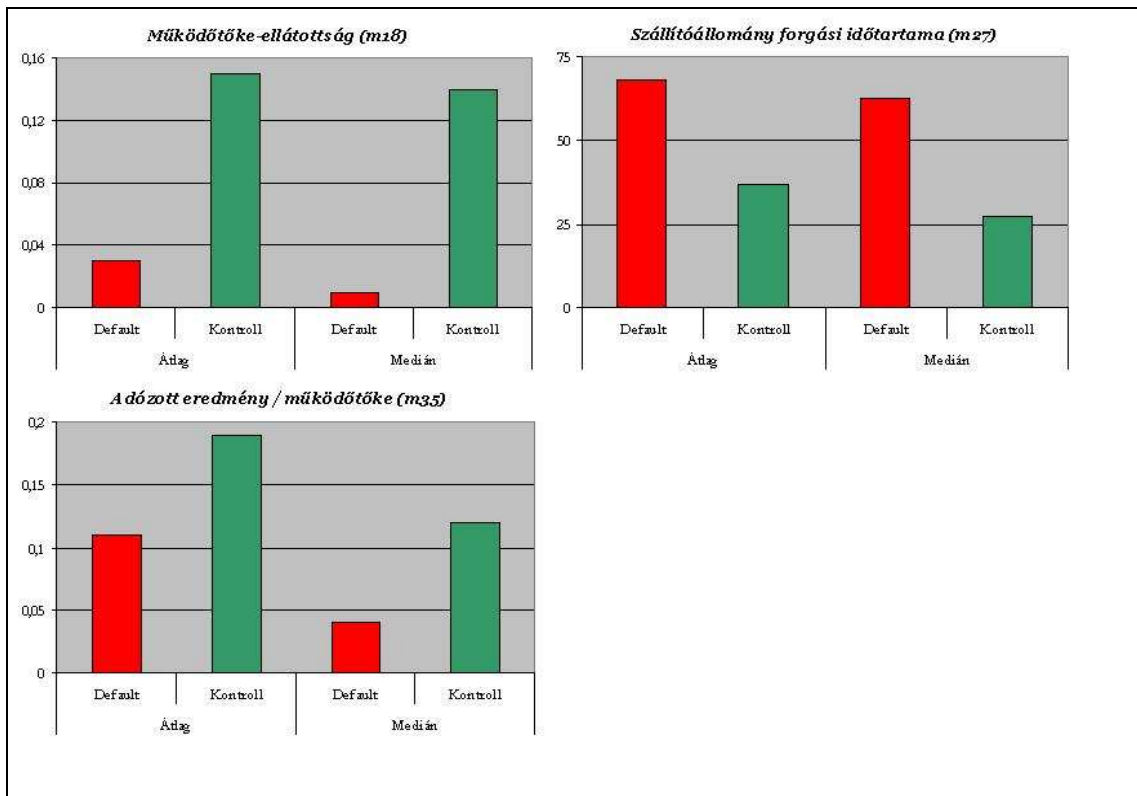
Mutatók	Átlag		Medián		Szórás
	Default	Kontroll	Default	Kontroll	
Likviditási mutatók					
likviditási ráta (m ₁)	1,20	1,69	1,01	1,43	0,79
Tőkeszerkezeti mutatók					
idegen tőke / saját tőke (m ₈)	3,93	1,82	2,88	0,98	2,82
hosszú távú eladósodottság (m ₉)	0,29	0,13	0,27	0,04	0,23
Adósságszolgálati mutatók					
cash-flow fedezet (m ₁₆)	8,41	16,14	2,92	7,92	15,81
Működőtőke mutatók					
működőtőke-ellátottság (m ₁₈)	0,03	0,15	0,01	0,14	0,20
Eszközforgási mutatók					
szállítóállomány forgási időtartama (m ₂₇)	67,96	36,98	62,53	27,33	41,98
Jövedelmezőségi mutatók					
adózott eredmény / működőtőke (m ₃₅)	0,11	0,19	0,04	0,12	0,56

20. táblázat: A leíró statisztika alapján legjobb szétválasztó erejű mutatók

A kiválasztott mutatók átlagait és mediánjait az alábbi ábrák szemléltetik.



12. ábra: Az előzetesen kiválasztott mutatók átlagai és mediánjai I.



13. ábra: Az előzetesen kiválasztott mutatók átlagai és mediánjai II.

A fenti felsorolás nem tartalmaz egyetlen termelékenységi mutatót sem. Ennek az az oka, hogy itt több olyan mutatót is találtam, amelyek a gazdasági racionalitással ellentétesen viselkednek.

Egy példán szemlélítve: az egy alkalmazottra eső üzemi eredmény (m_{30}) mutató átlaga a késedelembe eső vállalatok esetében 1.528 eFt, míg a kontrollcsoport esetén 1.277 eFt. Azaz a nemfizető cégek átlagban 20%-kal jobban teljesítenek a pénzügyi problémák nélküli vállalatoknál, miközben ennek pont az ellenkezőjét várnánk. A mediánokat tekintve még szembetűnőbb a különbség: a defaultos cégek 1.086 eFt-os mediánja 70%-kal haladja meg a kontroll vállalatok 642 eFt-os mediánját.

Itt nagy valószínűséggel a létszám adatok alacsonyabb fokú megbízhatósága merül fel elsődleges magyarázatként. Magyarországon ebben a kérdésben általánosan hallhatunk a fekete illetve szürke foglalkoztatásról, amelynek hatása van a statisztikai létszám adatokra is. A termelékenységi mutatók esetében alighanem ez állhat a háttérben, bár a jelenlegi adatokra támaszkodva ez nem bizonyítható.

Hasonló irracionális viselkedés fedezhető fel néhány további (főképp jövedelmezőségi) mutató esetében is:

- időtartam mutató (m_5),
- üzemi eredmény / saját tőke (m_{42}),
- EBIT/dologi nettó érték (m_{45}),
- ROCE I. (m_{47}),
- ROCE II. (m_{48}),
- Tőkearányos EBITDA (m_{49}).

Ezeket a mutatókat tehát az előzetes elemzések alapján teljesen kizártam a lehetséges szétválasztó erejű faktorok közül, mivel a modellek így a közgazdasági racionalitással ellentétes kapcsolatot feltételezve dolgoznának.

A fentiekre támaszkodva az alábbi hipotézist fogalmaztam meg.

H2. A csődmodellekhez hasonlóan a default esemény is modellezhető ugyanazon statisztikai módszertan felhasználásával. A modellek teljesítménye ugyanakkor – a kétfajta pénzügyi helyzet eltérő jellegéből és súlyosságából fakadóan - várhatóan a csődmodelleké alatt marad.

A deskriptív statisztika segítségével elvégzett előzetes mutatószelekció alapján a modellek várhatóan a következő mutatókat fogják felhasználni:

- likviditási ráta (m_1),
- idegen tőke / saját tőke (m_8),
- hosszú távú eladósodottság (m_9),
- cash-flow fedezet (m_{16}),
- működőtőke-ellátottság (m_{18}),
- szállítóállomány forgási időtartama (m_{27}),
- adózott eredmény / működőtőke (m_{35}).

5.4. A modellfejlesztési munka lépései

5.4.1. A modellező szoftver

A modellek kidolgozásához a SAS Enterprise Miner 5.2. szoftvert használtam, amelyet a SAS Institute Kft. bocsátott rendelkezésemre.

5.4.2. A modellezés folyamata

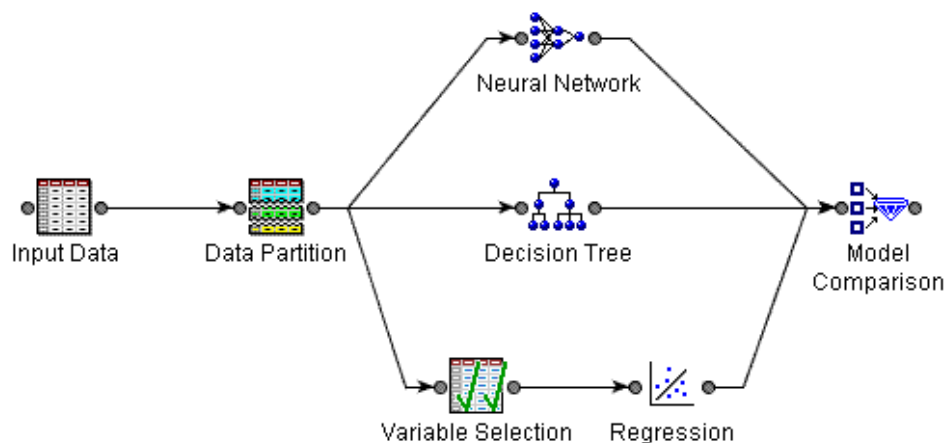
Mivel a szoftver nagyban meghatározza az alkalmazható módszerek körét is, ezért értelemszerűen csak olyan módszerekkel végezhettem el a számításokat, amelyeket a SAS támogat.

Az Enterprise Miner segítségével háromféle statisztikai módszert teszteltem²⁷:

- logisztikus regressziós modell,
- rekurzív partícionáló algoritmus (döntési fa),
- neurális háló.

Az elkészült modellek teljesítményét a korábbiakban ismertetett értékelési módszerek segítségével vettem össze, és kiválasztottam azt, amelyik a legeredményesebben alkalmazható a default előrejelzésére.

Az alábbi ábra mutatja a modellfejlesztés folyamatát a SAS Enterprise Minerben:



14. ábra: A modellfejlesztés folyamatábrája

²⁷ Itt kell megjegyeznem, hogy a modellezési munka során több hazai és nemzetközi szakirodalomra támaszkodtam, amelyekből sok elméleti és gyakorlati útmutatást, ötletet, megoldási javaslatot merítettem. Ezek közül kiemelném az alábbiakat, amelyek véleményem szerint a téma iránt érdeklődőknek mindenképpen nagy segítséget jelentenek: Bodon (2004), Deutsche Bundesbank (2003), Fair Isaac (2006), Hajdu (2003), Hajdu (2004), Hosner-Lemeshow (2000), Hossari (2007), Info-Datax (2006), Mays (2004), SAS Institute (2001, 2003, 2006), Siddiqi (2006), Thomas-Edelman-Crook (2002).

A továbbiakban a folyamatábrát alkotó részfeladatok (a SAS terminológia szerint: ún. „node”-ok) tartalmát és az alkalmazott paramétereket, beállításokat tekintem át.

5.4.2.1. Adatbeolvasás (*Input Data Node*)

Ez a node határozza meg, hogy milyen adatfájl, illetve annak mely adatmezői képezzék az elemzés bemenetét. Pontosán meg kell határozni az egyes adatoszlopok formátumait, valamint a modellezésben betöltött szerepeiket. Ez alapvetően kétféle lehet:

- *célváltozó (target)*: vagyis az előrejelzés kimeneti csoportjait tartalmazó adatoszlop. Esetünkben ez a default bekövetkezését jelző dummy változó oszlopa, ahol 1 jelzi a legalább 90 napos fizetési késedelembe eső vállalatot, 0 pedig ennek ellenkezőjét.
- *bemenő változó (input)*: azok az adatok, amelyek a modellezésnél a független változókat jelentik. Ezek lehetnek:
 - *folyamatos változók*: ezek a beszámolók főbb sorainak adatai: árbevétel, üzemi eredmény, saját tőke, fizetett kamatok, pénzügyi ráfordítások, vállalkozás létszáma, illetve a pénzügyi-számviteli mutatók ($m_1 - m_{52}$).
 - *diszkrét változók*: gazdasági forma, megye, régió, tevékenység (TEÁOR) illetve főágazat.²⁸

5.4.2.2. Mintafelosztás (*Data Partition Node*)

A modellezés során két részminta segítségével dolgoztam. Az eredeti adatbázist tanulási és tesztelési mintákra osztottam. A tanulási mintán megedzett modellek eredményei így összehasonlíthatók lesznek a modellek számára ismeretlen tesztelési mintán elért eredményekkel.

²⁸ A folytonos illetve diszkrét változók fogalmai nem teljesen esnek egybe a szokásos matematikai értelmezéssel. A SAS az input változók definiálásánál lényegében annyit vár el, hogy definiáljuk számára, miként értelmezze az egyes oszlopokban szereplő adatokat: csak az egyes ismerv-kategóriáknak megfeleltethető numerikus és/vagy karakteres kódokat tartalmaz, vagy folytonos numerikus értékeket. A két típus kezelése ugyanis nem ugyanolyan módon történik: az előbbieket a SAS a háttérben dummy változókká alakítva használja fel a modellezési folyamat során.

Ez egyrészt egyfajta kontrol szerepét is betölti, hiszen ha a besorolási pontosság hasonló a tanulási és a tesztelési mintán, akkor nem áll fenn a túltanulás jelensége, a modell véglegesíthető. Ha viszont a tesztelési mintán mért hibaarányok jelentősen meghaladják a tanulási mintán mért szintet, akkor bizonyítást nyer a túltanulás, új modellt kell készíteni. Erre mind a döntési fa, mind a neurális háló modellek esetében kiemelt figyelmet kell fordítani.

A modellek kidolgozásánál a szakirodalomban általában elterjedt, 75% - 25% arányú tanulási és tesztelési részmintákat alkalmaztam.

A Data Partition Node részletes beállításait mutatja az alábbi ábra.

Property	Value
Node ID	Part
Imported Data	
Exported Data	
Variables	
Partitioning Method	Default
Random Seed	123456
☒ Data Set Percentages	
Training	75.0
Validation	25.0
Test	0.0

15. ábra: Az Enterprise Miner Data Partition Node paraméterezése²⁹

5.4.2.3. Döntési fa (Decision Tree Node)

A döntési fa kialakításánál több kísérletet végeztem. A legpontosabbnak az a modell bizonyult, amelyben 6 vágási mélységet engedélyeztem és bináris fát modelleztem, vagyis az egyes elágazási pontoknál csak kétfelé történhet a minta vágása.

Ez a modelltípus is szakértő által vezérelt, azaz a modellezési munka során interaktív módon felügyelni kell a folyamatot. A döntési fa esetében ez azt jelenti, hogy meg kell találni azt a beállítást, amelynél a modell nem lesz elnagyolt, tehát megfelelő

²⁹ Az egyes paraméterek fordítása: beolvasott adatok, exportált adatok, változók, particionálás módszere, véletlenszám-generátor paramétere, fejlesztési minta aránya, ellenőrző minta aránya, tesztelési minta aránya.

pontosságot ér el, de mégis elkerüli a túltanulást, azaz nem alkalmazkodik túlságosan a minta specifikumaihoz, így általános előrejelzésre is alkalmas lesz.

A döntési fa esetében ez legkönnyebben a fa mélységével, illetve az egy levélen engedélyezett minimális megfigyelések számával szabályozható, a túltanulást pedig a tanulási és tesztelési mintán elért eredményeken kísérhetjük figyelemmel. Ha ezek között túl nagy a különbség, vagyis a modell teljesítménye a számára ismeretlen adatbázison túlzottan visszaesik, az túltanulásra enged következtetni. A legjobb modellt tehát a folyamatos felügyelet mellett kialakított, a kedvezőtlen eredmények átparaméterezett újrafuttatások alapján lehet véglegesíteni.

A Decision Tree Node végleges, legjobb modellt eredményező paraméterezését mutatja a következő ábra.

Property	Value
Node ID	Tree
Imported Data	
Exported Data	
Variables	
Interactive	
☐ Splitting Rule	
Criterion	Default
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	2
Maximum Depth	6
Minimum Categorical Size	5
☐ Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	
☐ Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Decision
Assessment Fraction	0.25
☐ Output Variables	
Variable Selection	Yes
Leaf Variable	Yes
Leaf Role	Segment
Performance	Disk

16. ábra: A Decision Tree Node paraméterezése³⁰

5.4.2.4. Logit modell (Regression Node)

A logisztikus regressziós modell esetén a SAS módszertani útmutatója szerint jártam el. A regressziós modell lefuttatása előtt végrehajtottam egy ún. változókiválasztási lépést is. Ez tulajdonképpen egy egyszeres faktoranalízis, amelynek lényege, hogy egyesével megvizsgálja a bemenő változók és a célváltozó közötti kapcsolat erősségét. Ezután csak azokat a változókat használja fel a regressziós modell kialakításánál, amely egy előre definiált szintnél erősebb kapcsolatot mutat a célváltozóval. A logisztikus

³⁰ Az egyes paraméterek fordítása: beolvasott adatok, exportált adatok, változók, interaktív mód, vágási kritériumok, szignifikancia szint, hiányzó értékek, ágak maximális száma, maximális mélység, minimum csoportnagyság, levélméret, szabályok száma, helyettesítő szabályok száma, vágási méret, alfák mőszere, levelek száma, értékelési módszer, értékelési szakaszok, kimenő változók kiválasztása, levél változója, levél szerepe, teljesítmény

regressziós modell pedig feltárja az e mutatók között fennálló kapcsolatokat, és kialakítja a végleges modellt.³¹

A tesztelést elvégeztem a változókiválasztás módszere nélkül is, és valóban azt tapasztaltam, hogy az úgy kapott modell teljesítménye már a tanulási mintán is valamelyest rosszabb volt, a tesztelési mintán azonban már szignifikáns volt a különbség a besorolási pontosság terén.

A legjobb teljesítményt elérő logit modell pontos paraméterezését az alábbi ábra mutatja.

³¹ A SAS az input változók és a célváltozók közötti kapcsolatot kétféleképpen tudja mérni: a determinisztikus együtthatóval (R^2) illetve az egyes bemenő változók szignifikanciájával (Weight of Evidence, WOE). A paraméterezésben ez szabadon választható. A modellek kialakítása során én az utóbbi beállítással dolgoztam.

Property	Value
Node ID	Reg
Imported Data	
Exported Data	
Variables	
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppress Intercept	No
Input Coding	Deviation
Min Resource Use	No
Model Selection	
Selection Model	None
Selection Criterion	Default
Use Selection Default	Yes
Selection Options	
Sequential Order	No
Entry Significance Level	0.05
Stay Significance Level	0.05
Start Variable Number	0
Stop Variable Number	0
Force Candidate Effects	0
Hierarchy Effects	Class
Moving Effect Rule	None
Maximum Number of Steps	0
Optimization Options	
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour

17. ábra: A Regression Node paraméterezése³²

5.4.2.5. Neurális háló (Neural Network Node)

A neurális háló modelljénél a legelterjedtebb, ún. MLP (multilayer perceptron³³) típust alkalmaztam, amely a standard backpropagation tanulási algoritmussal dolgozik.

³² Az egyes paraméterek fordítása: beolvasott adatok, exportált adatok, változók, regresszió típusa, kapcsolat típusa, modell beállítások, input kódolás, minimum erőforrás, modellkiválasztás, kiválasztási kritériumok, kiválasztási alapbeállítások használata, sorrendiség, beválasztási szignifikancia szint, bentmaradási szignifikancia szint, kezdő változó szám, utolsó változó szám, kényszerítés, hierarchia hatások, mozgó hatások szabálya, maximum lépésszám, optimalizációs technika, alapbeállítás, maximális iterációk, maximális függvényhívások, maximális futási idő.

Mivel ez a modelltípus a legérzékenyebb a túltanulásra, többszöri kísérletezés után sikerült csak végleges modellt kialakítani. A neurális hálónál az alkalmazott rétegek számával, illetve az iterációk számával szabályozható a túltanulás elkerülése. Ugyanúgy, ahogy a döntési fa esetében, a túltanulás a fejlesztési és tesztelési mintán elért eredményeken látható: ha a modell teljesítménye a számára ismeretlen adatbázison nagymértékben visszaesik, akkor fennáll a túltanulás jelensége. Ilyenkor a modellt el kell vetni, és újabb paraméterezés után újrafuttatni.

A legeredményesebbnek a 3 köztes réteget tartalmazó MLP modell bizonyult, amely maximálisan 20 iterációval dolgozott.

A Neural Network Node végleges, legjobb modellt eredményező paraméterezését mutatja a következő ábra.

³³ A kifejezést körülbelül „többrétegű érzékelő hálónak” lehetne fordítani, bár a szaknyelvben magyar fordítást nem használnak, az MLP rövidítés az elterjedt megnevezés.

Property	Value
Node ID	Neural
Imported Data	
Exported Data	
Variables	
Use Current Estimates	No
Architecture	MLP
Direct Connection	No
Model Selection Criterion	Profit/Loss
Number of Hidden Units	3
☒ Training Options	
Maximum Iterations	20
Maximum Time	4 Hours
Training Technique	Default
☒ User Defined Network Options	
Randomization Distribution	Normal
Randomization Center	0.0
Randomization Scale	1.0
Input Standardization	Standard Deviation
Hidden Layer	Yes
Hidden Layer Combination Function	Default
Hidden Layer Activation Function	Default
Hidden Bias	Yes
Target Layer Combination Function	Default
Target Layer Activation Function	Default
Target Layer Error Function	Default
Target Bias	Yes
☒ Preliminary Training Options	
Preliminary Training	No
Maximum Iterations	10
Maximum Time	1 Hour
Number of Runs	5

18. ábra: A Neural Network Node paraméterezése³⁴

5.4.2.6. Modellek kiértékelése (Model Comparison Node)

Az utolsó lépés a folyamatban a kiértékelés. A node összegyűjti a kidolgozott modellek eredményeit, és összesítve megjeleníti azokat. Elsődleges kiértékelési szempontja az ROC-görbe illetve az ahhoz tartozó AUROC index. Ezen kívül természetesen több más

³⁴ Az egyes paraméterek fordítása: beolvasott adatok, exportált adatok, változók, aktuális becslések alkalmazása, felépítés, közvetlen kapcsolatok, model kiválasztás kritériumok, rejtett egységek száma, maximális iterációk, maximális futásidő, tanulási technika, véletlenszerű eloszlás, véletlenség középpontja, véletlenségi skála, bemeneti eloszlás, rejtett régegek, rejtett rétegek kombinációs függvénye, rejtett rétegek aktiválási függvénye, rejtett torzítások, célréteg kombinációs függvénye, célréteg aktiválási függvénye, célréteg hibafüggvénye, cél torzítás, előzetes tanulás, maximális iterációk, maximális futási idő, futtatások száma

értékelési szempontot is ismer, amelyet – elég kezelhetetlen formátumú - táblázat formájában foglal össze.

A modellek kiértékelését én a módszertani áttekintésben bemutatott technikák segítségével fogom részletesen bemutatni, a következő fejezetben.

A Model Comparison Node esetében alkalmazott paraméterezést az alábbi ábra mutatja.

Property	Value
Node ID	MdlComp
Imported Data	
Exported Data	
Variables	
Decile Bin	20
Lift Epsilon	1.0E-6
Profit Epsilon	1.0E-6
ROC Chart	Yes
ROC Epsilon	0.01
ScoreDist Bin	20
Selection Statistic	Default

19. ábra: A Model Comparison Node paraméterezése³⁵

5.5. A kidolgozott modellek, eredmények

5.5.1. A döntési fa

A döntési fa modellje összesen 8 mutatót használ: két kvalitatív ismérvet és hat pénzügyi mutatót. Ezek:

- gazdálkodási forma,
- megye,
- hosszú távú eladósodottság (m_9),
- kötelezettségek / dologi nettó érték (m_{11}),

³⁵ Az egyes paraméterek fordítása: beolvasott adatok, exportált adatok, változók, kategóriák száma, lift epsilon, profit epsilon, ROC grafikon, ROC epsilon, score-pontok kategóriáinak száma, kiválasztási statisztika

- cash-flow fedezet (m_{16}),
- rövid lejáratú kötelezettségek/ dologi nettó érték (m_{21}),
- összes eszköz forgási időtartama (m_{22}),
- eszközarányos adózott eredmény - ROA (m_{34}).

Az is látható, hogy az objektív faktorok mutatócsoportok közötti eloszlása egyenletes: a két tőkeszerkezeti mutató mellett egy-egy adósságszolgálati, működőtőke, eszközforgási és jövedelmezőségi mutató került be a modellbe. Nem találunk ugyanakkor likviditási mutatókat a döntési pontok között.

5.5.2. A logit modell

A logisztikus regressziós modell 11 mutatóval dolgozik. Ugyanaz a két szubjektív faktor jelenik meg itt is, mint a döntési fa esetében, és a pénzügyi mutatók között is négy megegyezik.

A logit modell változói:

- gazdálkodási forma,
- megye,
- dinamikus likviditás (m_4),
- tőkeellátottság (m_6),
- hosszú távú eladósodottság (m_9),
- kötelezettségek / tárgyi eszközök (m_{10}),
- cash-flow fedezet (m_{16}),
- adósságvisszafizető-képesség (m_{17}),
- rövid lejáratú kötelezettségek/ dologi nettó érték (m_{21}),
- szállítóállomány forgási időtartama (m_{27}),
- eszközarányos adózott eredmény - ROA (m_{34}).

Az egyes mutatócsoportok közötti megoszlás a logit esetében a következőképpen alakul: a három tőkeszerkezeti és a két adósságszolgálati mutató mellett egy-egy működőtőke, eszközforgási, jövedelmezőségi illetve likviditási mutató került be a modellbe.

5.5.3. A neurális háló

A neurális háló jóval bonyolultabb szerkezetű modellt jelent, mint az előző két modellesalád, mivel a bemenő változók közötti bonyolult kapcsolatok feltárására irányul a tanulási algoritmus. Ennek megfelelően a modell is sokkal bonyolultabb.

Összesen 34 döntési tényezőt használ fel a modell a csoportba sorolás meghatározásánál. Ezek között szerepel a korábbiakban látott gazdálkodási forma és a megye, de bekerült új kvalitatív ismérvként az ágazati kód (TEÁOR) is.

A modell felhasznál két közvetlenül a beszámolóból származó adatot is: a fizetett kamatok és a pénzügyi ráfordítások értékét.

Az összes bemeneti változó listája:

- gazdálkodási forma,
- megye,
- ágazat,
- pénzügyi ráfordítások,
- fizetett kamatok,
- likviditási ráta (m_1),
- likviditási gyorsráta (m_2),
- pénzhányad (m_3),
- dinamikus likviditás (m_4),
- tőkeellátottság (m_6),
- eladósodottság aránya (m_7),

- idegen tőke / saját tőke (m_8),
- hosszú távú eladósodottság (m_9).
- kötelezettségek / tárgyi eszközök (m_{10}),
- kötelezettségek / dologi nettó érték (m_{11}),
- kamatfedezet I. (m_{12}),
- kamatfedezet II. (m_{13}),
- EBITDA fedezet I. (m_{14}),
- EBITDA fedezet II. (m_{15}),
- cash-flow fedezet (m_{16}),
- adósságvisszafizető-képesség (m_{17}),
- működőtőke-ellátottság (m_{18}),
- működőtőke forgása (m_{19}),
- rövid lejáratú kötelezettségek/ dologi nettó érték (m_{21}),
- összes eszköz forgási időtartama (m_{22}),
- készletek forgási időtartama I. (m_{23}),
- készletek forgási időtartama II. (m_{24}),
- beszedés időtartama (m_{25}),
- vevőállomány forgási időtartama (m_{26}),
- szállítóállomány forgási időtartama (m_{27}),
- eszközarányos adózott eredmény - ROA (m_{34}),
- adózott eredmény / működőtőke (m_{35}),
- eszközarányos adózás előtti eredmény (m_{39}),
- nyereség-visszaforgatás aránya (m_{51}).

A neurális háló minden mutatócsoportból tartalmaz pénzügyi mutatókat.

Az eredmények alapján tehát megfogalmazható az alábbi tézis.

T2. A rendelkezésemre álló adatbázis segítségével igazoltam, hogy a csődmodellek mintájára készíthető olyan prediktív modell, amely a Bázel II. definícióknak megfelelően a vállalatok 90 napos késedelmes fizetését (defaultját) jelzi előre, 83%-os megbízhatósággal.

5.5.4. A modellekben szereplő döntési faktorok

Összefoglalásképpen tekintsük át az egyes modellek által felhasznált döntési faktorokat!

<i>Mutatók</i>	<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
Kvalitatív ismérvek			
gazdálkodási forma	X	X	X
megye	X	X	X
ágazat			X
Beszámolóadatok			
pénzügyi ráfordítások			X
fizetett kamatok			X
Likviditási mutatók			
likviditási ráta (m_1)			X
likviditási gyorsráta (m_2)			X
pénzhányad (m_3)			X
dinamikus likviditás (m_4)		X	X
Tőkeszerkezeti mutatók			
tőkeellátottság (m_6)		X	X
eladósodottság aránya (m_7)			X
idegen tőke / saját tőke (m_8)			X
hosszú távú eladósodottság (m_9)	X	X	X
kötelezettségek / tárgyi eszközök (m_{10})		X	X
kötelezettségek / dologi nettó érték (m_{11})	X		X

Adósságszolgálati mutatók			
kamatfedezet I. (m_{12})			X
kamatfedezet II. (m_{13})			X
EBITDA fedezet I. (m_{14})			X
EBITDA fedezet II. (m_{15})			X
cash-flow fedezet (m_{16})	X	X	X
adósságvisszafizető-képesség (m_{17})		X	X
Működőtőke mutatók			
működőtőke-ellátottság (m_{18})			X
működőtőke forgása (m_{19})			X
rövid lejáratú kötelezettségek/ dologi nettó érték (m_{21})	X	X	X
Eszközforgási mutatók			
összes eszköz forgási időtartama (m_{22})	X		X
készletek forgási időtartama I. (m_{23})			X
készletek forgási időtartama II. (m_{24})			X
beszedés időtartama (m_{25})			X
vevőállomány forgási időtartama (m_{26})			X
szállítóállomány forgási időtartama (m_{27})		X	X
Jövedelmezőségi mutatók			
eszközarányos adózott eredmény - ROA (m_{34})	X	X	X
adózott eredmény / működőtőke (m_{35})			X
eszközarányos adózás előtti eredmény (m_{39})			X
nyereség-visszaforgatás aránya (m_{51})			X

21. táblázat: A kidolgozott default-modellek által felhasznált mutatók

T3. Igazolódott, hogy a kvalitatív ismérvek felhasználása is eredményes volt. Két ismerv, a gazdálkodási forma és a megye is bekerült mindhárom végleges modellbe.

A legkockázatosabb megyéknek az alábbiak bizonyultak:

- Baranya megye,
- Békés megye,
- Hajdú-Bihar megye,
- Nógrád megye,
- Pest megye,
- Vas megye,
- Veszprém megye,
- Zala megye.

A legkisebb kockázatot jelentő megyék pedig:

- Bács-Kiskun megye,
- Csongrád megye,
- Heves megye,
- Komárom-Esztergom megye.

A gazdálkodási formák közül kiemelkedően kockázatosak a korlátolt felelősségű társaságok (kft). A részvénytársaságok inkább a közepes kockázati kategóriába tartoznak, míg a betéti társaságok (bt) illetve egyéb cégformák jelentik a legkisebb kockázatot.

A beszámolóadatokat közvetlenül csak a neurális háló alkalmazza, a másik két modell egyáltalán nem.

Az egyes mutatócsoportokat tekintve látható, hogy a likviditási mutatók azok, amelyek a legkevesebbszer szerepeltek a modellekben. Szinte csak a neurális háló számol ezekkel a mutatócsoportokkal, a döntési fa egyáltalán nem, a logit modell is csak egyetlen likviditási mutatót tartalmaz.

Legfontosabbnak ugyanakkor a tőkeszerkezeti és adósságszolgálati mutatók tűnnek: ezekből mindegyik modell többet is használ.

Az egyes mutatók alkalmazását tekintve megállapítható, hogy a kvalitatív ismérvekkel együtt összesen hat mutató szerepel mindhárom modellben.

T4. A nemfizetés előrejelzése szempontjából legfontosabb mutatók az alábbiak:

- 1. hosszú távú eladósodottság,**
- 2. cash-flow fedezet,**
- 3. rövid lejáratú kötelezettségek/ dologi nettó érték,**
- 4. eszközarányos adózott eredmény (ROA).**

Ezeket a mutatókat az általam készített modellek mindegyike felhasználja.

A mutatók, amelyeket két modell alkalmaz:

- dinamikus likviditás (m_4),
- tőkeellátottság (m_6),
- kötelezettségek / tárgyi eszközök (m_{10}),
- kötelezettségek / dologi nettó érték (m_{11}),
- adósságvisszafizető-képesség (m_{17}),
- összes eszköz forgási időtartama (m_{22}),
- szállítóállomány forgási időtartama (m_{27}).

Ha a fenti mutatókat összevetjük az előzetes mutatókiválasztásnál meghatározott ráták listájával, az eredmény felemás.

A hétből három mutató nemcsak az előzetes várakozások alapján tűnt jó szétválasztó erejűnek, azokat több modellezési technika felhasználta:

- hosszú távú eladósodottság (m_9),
- cash-flow fedezet (m_{16}),
- szállítóállomány forgási időtartama (m_{27}).

A fennmaradó négy mutatót viszont csak a neurális háló modellje használja fel, mind a döntési fa, mind a logisztikus regresszió más rátákat részesített előnyben az adott mutatócsoportokból.

5.5.5. Az egyes modellek teljesítményének összehasonlítása, értékelése

A kidolgozott modelleket a módszertani áttekintésnél ismertetett technikák alapján értékeltem. Először tekintsük a klasszikus értékelési módszert, a klasszifikációs mátrix alapján mért besorolási pontosságokat! Ezeket az alábbi táblázatban foglaltam össze.

		<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
Besorolási pontosság	<i>Fejlesztési minta</i>	82,9%	83,5%	83,5%
	<i>Tesztelési minta</i>	78,0%	81,8%	81,4%
Elsőfajú hiba	<i>Fejlesztési minta</i>	11,8%	8,6%	8,3%
	<i>Tesztelési minta</i>	12,8%	9,4%	10,4%
Másodfajú hiba	<i>Fejlesztési minta</i>	5,3%	7,9%	8,1%
	<i>Tesztelési minta</i>	9,2%	8,8%	8,2%

22. táblázat: A kidolgozott modellek besorolási pontosságai és hibaarányai

Látható, hogy a fejlesztési mintán mindhárom modell 83% körüli pontosságot ért el, ami jónak mondható. Igazolva látjuk ugyanakkor azt a hipotézist is, hogy míg a csődöt 90%-ot is meghaladó pontossággal lehet előre jelezni, a default már nem ennyire egyértelmű.

T5. A default modellek előrejelzési pontossága némileg elmarad a csődmodellekétől. Ennek a magyarázata az lehet, hogy a pénzügyileg egészséges és a problémás vállalatok profilja valószínűleg még nem különbözik olyan markánsan egymástól, mint a csődbe jutott és a túlélő vállalatoké, ezért az előrejelzés is nehezebb, pontatlanabb.

A tesztelési mintán mért pontosságok esetén megállapítható, hogy mindhárom modell teljesítménye kissé visszaesik az ismeretlen adatokon való alkalmazás során, de amíg a neurális háló és a logit modell csak néhány százalékpontos csökkenést mutat, addig a döntési fa 78%-ra esik vissza. Összességében azonban mindenképpen az látható, hogy a modellek stabilan működnek, és előrejelzési célokra is megfelelően alkalmazhatóak.

Megállapítható továbbá, hogy a logisztikus regresszió és a neurális háló teljesítménye ugyanolyan jó, bármelyik mintán is vizsgáljuk.

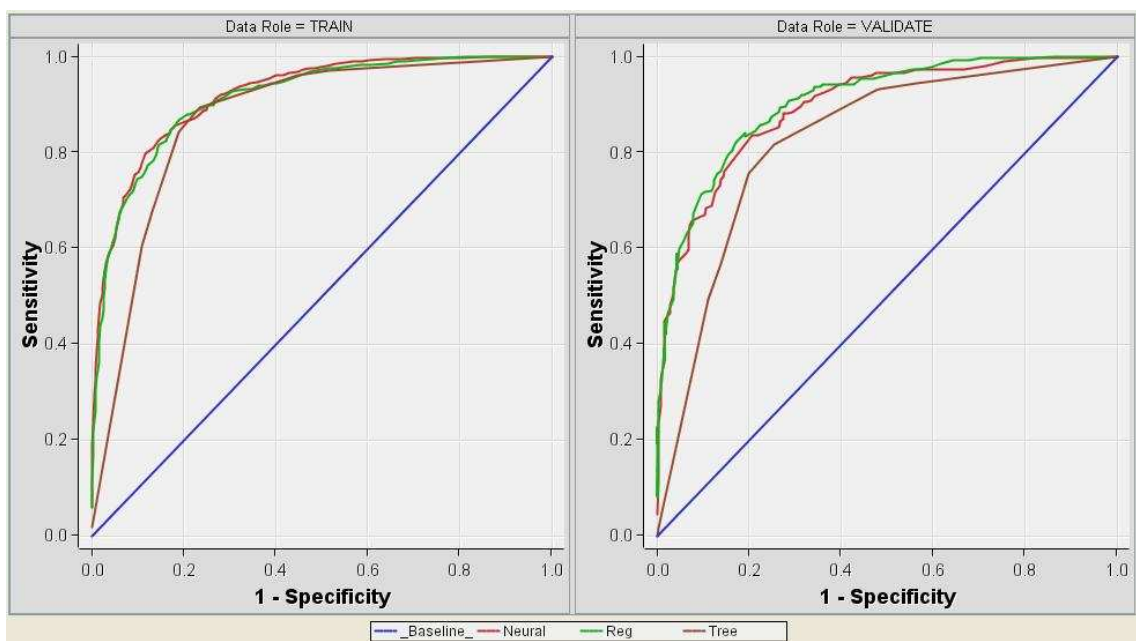
További következtetéseket vonhatunk le, ha részletesebben megvizsgáljuk az egyes modellek által elkövetett besorolási hibák jellegét. A bank szempontjából nem mindegy ugyanis, hogy milyen irányban téved az előrejelzés: egy defaultosnak feltételezett vállalat esetén ugyanis a hitel elutasítása nem jelent közvetlen veszteséget a banknak, a következménye „csupán” piacvesztés, illetve alternatív költsége, elmaradt haszna van. Egy problémás vállalat jónak minősítése és meghitelezése ugyanakkor tényleges pénzügyi veszteséggel jár a bank számára, ezért a másodfajú hibák következményei sokkal komolyabbak.

Elsőfajú hibák tekintetében az látható, hogy a fejlesztési mintán mind a logisztikus regresszió, mind a neurális háló 8% körüli hibaarányal dolgozik, a döntési fa e tekintetben jóval elmarad a másik két modelltől a maga 11,8%-os hibaarányával. A tesztelési mintán ez az érték kissé tovább emelkedik, 12,8%-ra, míg a másik két modell esetén az elsőfajú hibák 9-10% körüli szintre emelkednek, jóval a döntési fa hibaarányai alatt maradnak. E tekintetben a logit regresszió teljesít jobban, eredménye egy százalékponttal jobb a neurális hálónál.

A bank számára lényegesebb másodfajú hibák arányát tekintve hasonló eredmények láthatók. A fejlesztési mintán a neurális háló és a logit regresszió 8% körüli

hibaarányal dolgozik, azonban a döntési fa itt mindössze 5,3%-ban hibázik. A tesztelési mintán ugyanakkor a másodfajú hibaarány kisimul: mindhárom modell 8-9%-os eredményt mutat.

Ha az ROC-görbe grafikus megjelenítését nézzük, a különbség a döntési fa és a két másik modell között jól láthatóvá válik. Míg a neurális háló és a logit görbéi szinte egybeesnek, a döntési fa grafikonja ezek alatt fut, ezzel jelezve a gyengébb teljesítményt.



20. ábra: A kidolgozott modellek ROC-görbéi

Az alábbi táblázat a modellek további értékelési technikákkal mért teljesítményi adatait foglalja össze.

		<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
AUROC	<i>Fejlesztési minta</i>	87,2	91,3	92,0
	<i>Tesztelési minta</i>	83,0	90,3	89,3
Gini-koefficiens	<i>Fejlesztési minta</i>	74,4	82,6	84,0
	<i>Tesztelési minta</i>	65,9	80,6	78,6
K-S	<i>Fejlesztési minta</i>	65,7	67,7	68,1
	<i>Tesztelési minta</i>	56,0	65,2	62,8

23. táblázat: A kidolgozott default-modellek teljesítménymutatói

A fenti adatok is megerősítik azt az értékelést, amelyet a klasszifikációs mátrix és a ROC-grafikon is jelzett: a logisztikus regresszió és a neurális háló teljesítménye közel ugyanolyan, a döntési fa ha nem is jelentősen, de elmarad mögöttük.

Az AUROC indexet tekintve szinte elhanyagolható a különbség a két legjobb modell között. A Gini-koefficiensnél látható ugyanakkor egy hajszálnyi előny a logit oldalán: a fejlesztési mintán ugyan csak 82,6-os értéket mérhetünk a 84-gyel szemben, a tesztelési mintán viszont nem esik annyira vissza a logit együttthatója, mint a neurális hálóé: a koefficiens 80,6, míg a neurális hálóé 78,6. Ez a logit modell kismértékű előnyét jelzi a stabilitás szempontjából. Hasonló figyelhető meg a Kolmogorov-Smirnov teszt esetén is, a logit esetében ez 67,7-ről csak 65,2-re esik vissza, míg a neurális hálónál 68,1-ről 62,8-ra.

Ha végleges konklúziót kell levonni a fenti értékelésből, akkor ez a minimális előny nem lenne elegendő érv a logit modell mellett. Azt mondhatjuk, hogy a döntési fa az abszolút harmadik a modellek sorában, míg a logit és a neurális háló teljesítménye ugyanolyan jó, közöttük csak elhanyagolható a különbség.

A leglényegesebb szempontból, a gyakorlati alkalmazhatóság oldaláról nézve a két modell közötti különbség szembetűnő. Nem mindegy ugyanis, hogy ugyanazt a teljesítményt 11 vagy 34 döntési tényező felhasználásával éri el a modell. Ez a különbség egyértelműen a modellek természetéből fakad, de kétségtelen, hogy egy logit modell implementálása a döntési folyamatba lényegesen egyszerűbb feladat.

A végkövetkeztetés tehát az lehet, hogy bár teljesítményben nem, de megvalósíthatóság terén a logit modell mindenképp felülmúlja a neurális hálót.

A fenti megállapításokat az alábbi tézisben foglaltam össze.

T6. Az alkalmazott modellezési technikák közül a döntési fa bizonyult a legkevésbé hatékony módszernek. A logisztikus regresszió és a neurális háló modelljei körülbelül ugyanúgy teljesítettek, ugyanakkor a könnyebb megvalósíthatóság és alkalmazhatóság a logit mellett szólnak.

6. További tesztek, elemzések

Az adatbázis segítségével két további hipotézist is megvizsgáltam:

- Növelhető-e a modellek besorolási pontossága, ha iparági középértékekhez viszonyított pénzügyi mutatókat vonunk be a vizsgálandó változók körébe?
- Növelhető-e a modellek pontossága dinamikus mutatók felhasználásával?

6.1. Iparági rátákon alapuló mutatók felhasználása a modellezésben

A szakirodalomban több kutatási eredmény³⁶ igazolja, hogy az iparági ráták javítják a csődmodellek besorolási pontosságát. A feltevést megpróbáltam ellenőrizni a rendelkezésemre álló default adatbázis segítségével.

Az elemzés előtt a következő hipotézist fogalmaztam meg:

H3. A szakirodalomban fellelt kutatási eredmények alapján feltételezem, hogy a modellek besorolási pontossága növelhető iparági ráták felhasználásával.

A teszt során azt az eredményt várom, hogy az új modellekbe az eredeti változókat kiszorítva bekerülnek az új, nagyobb információs erővel bíró iparági mutatók, és ezáltal növekszik az új modellek besorolási pontossága.

A cél tehát az volt, hogy bizonyítsam: az általam alkalmazott definíciók és modellezési technikák esetén is mérhető a modellek teljesítménynövekedése, amennyiben iparági mutatókat is felhasználok bemenő változókként.

A tesztelést úgy végeztem, hogy lefuttattam ugyanazon paraméterekkel a korábbi modelleket, de a bemenő adatok közé felvettem 52 újabb mutatót, amelyek az $m_1 \dots m_{52}$ mutatók iparági mediánjához való viszonyulását tükrözik.

A teszt akkor minősíthető sikeresnek, ha az új modellekben az eredeti változókat kiszorítva bekerülnek új, nagyobb információs erővel bíró mutatók, és ezáltal növekszik az új modellek besorolási pontossága.

³⁶ lásd pl. Platt-Platt (1990), Virág (1996), Virág-Kristóf (2006)

Ennek a fordítottja is igaz: ha az eredeti változók jobb magyarázó erővel rendelkeznek, akkor az új mutatók nem kerülnek be a modellekbe, a modellek besorolási pontossága pedig gyakorlatilag változatlan szinten marad.

A tesztek – némi meglepetésre - nem hoztak pozitív eredményt. A logisztikus regresszió esetében egyetlen új mutató sem került be a modellbe. A döntési fa annyiban módosult, hogy a rövid lejáratú kötelezettségek / dologi nettó érték (m_{21}) mutató helyére annak iparági változata lépett, de a modell teljesítménye semmit sem változott: mind a fejlesztési mintán, mind a tesztelési mintán ugyanúgy performált, mint az alapmodell.

A neurális háló annyiban módosult, hogy 34 helyett 65 mutatót vett figyelembe a csoportba sorolásnál, ugyanakkor a pontossága csak kb. másfél százalékpontot javult a fejlesztési mintán (85,3% a korábbi 83,5%-kal szemben). Az előrejelzési képesség becslésére használt tesztelési mintán ellenben nem volt látható semmilyen különbség: a korábbi 81,4% helyett 81,2%-ra még némileg csökkent is a pontossága.

A tesztelés során kapott eredményeket foglalja össze az alábbi táblázat.

		<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
Besorolási pontosság	<i>Fejlesztési minta</i>	82,9%	83,5%	85,3%
	<i>Tesztelési minta</i>	78,0%	81,8%	81,2%
Elsőfajú hiba	<i>Fejlesztési minta</i>	11,9%	8,6%	6,2%
	<i>Tesztelési minta</i>	12,8%	9,4%	8,8%
Másodfajú hiba	<i>Fejlesztési minta</i>	5,3%	7,9%	8,5%
	<i>Tesztelési minta</i>	9,2%	8,8%	10,0%

24. táblázat: Az iparági ráták tesztelése során kapott modellek besorolási pontosságai és hibaarányai

A tesztelést elvégeztem oly módon is, hogy nem az eredeti mutatók mellé vettem fel az új mutatókat, hanem az eredeti változókat teljesen lecseréltem, és csak az iparági ráták felhasználásával próbáltam modelleket készíteni.

A teszt itt még rosszabb eredményeket hozott: valamennyi modell teljesítménye drasztikusan visszaesett, a fejlesztési mintán is csak 70% körüli pontosságot értek el, a tesztelési mintán pedig még ennél is rosszabb eredményeket mutattak.

		<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
Besorolási pontosság	<i>Fejlesztési minta</i>	68,90%	70,23%	70,35%
	<i>Tesztelési minta</i>	64,30%	68,21%	67,70%
Elsőfajú hiba	<i>Fejlesztési minta</i>	19,28%	14,98%	12,22%
	<i>Tesztelési minta</i>	20,19%	15,66%	13,97%
Másodfajú hiba	<i>Fejlesztési minta</i>	11,82%	14,79%	17,43%
	<i>Tesztelési minta</i>	15,51%	16,13%	18,33%

25. táblázat: Az iparági ráták második tesztelése során kapott modellek besorolási pontosságai és hibaarányai

A szakirodalomban fellelhető példák ellenére ezt a hipotézist a jelen adatokon és definíciós háttérrel nem sikerült igazolnom, tehát el kellett vetnem.

T7. A szakirodalomban több kutatási eredmény igazolja, hogy az iparági ráták javítják a csődmodellek besorolási pontosságát, azonban ezt a rendelkezésemre álló adatbázis segítségével és az alkalmazott definíciós háttérrel nem sikerült igazolni: az iparági ráták felhasználásával képzett mutatók bevonása egyik modellen sem javított.

6.2. A modellek tesztelése dinamikus mutatók bevezetésével

A másik hipotézis során azt próbáltam megvizsgálni, hogy milyen hatással vannak a modellek teljesítményére az ún. dinamikus mutatók. Ezek az egy év beszámolóinak adataiból számított ún. statikus mutatók helyett az egyes mutatók éves változásait fejezik ki. Vagyis a vizsgált, tárgyévi mutatók értékeit ugyanazon mutatók tárgyévét megelőző évi értékéhez hasonlítjuk, tehát az új változók ezen mutatók éves változásait, dinamikáját mutatják.

Az elemzés előtt a következő hipotézist fogalmaztam meg:

H4. Feltételezhető, hogy a modellek besorolási pontossága növelhető a dinamikus mutatók felhasználásával.

A teszt során várhatóan azt tapasztaljuk, hogy az új modellekbe az eredeti változókat kiszorítva bekerülnek az új, nagyobb információs erővel bíró dinamikus mutatók, és ezáltal növekszik az új modellek besorolási pontossága.

A tesztelést ugyanúgy végeztem, mint az előző hipotézis esetében: lefuttattam ugyanazon paraméterekkel a korábbi modelleket, de a bemenő adatok közé felvettem 52 újabb mutatót, amelyek az $m_1 \dots m_{52}$ mutatók éves változásait mutatják.

A teszt itt is akkor minősíthető sikeresnek, ha az új modellekben az eredeti változókat kiszorítva bekerülnek új, nagyobb információs erővel bíró mutatók, és ezáltal növekszik az új modellek besorolási pontossága.

A hipotézis cáfolata is ugyanaz: ha az eredeti változók jobb magyarázó erővel rendelkeznek, akkor az új mutatók nem kerülnek be a modellekbe, a modellek besorolási pontossága pedig gyakorlatilag változatlan marad.

6.2.1. A dinamikus mutatók hatása a modellekre

A dinamikus mutatók tesztelése már látható eredményeket hozott.

T8. A rendelkezésre álló adatbázis segítségével igazoltam, hogy az egyes pénzügyi mutatók éves változásait kifejező ún. dinamikus mutatók felhasználásával növelhető az előrejelzési modellek besorolási pontossága.

A döntési fa ugyanúgy nyolc mutatót használt fel, mint az eredeti modell esetén, de nem ugyanazokat. Új mutatóként bekerültek a modellbe az alábbiak:

- kötelezettségek / tárgyi eszközök (m_{10}),
- 1 alkalmazottra eső nettó árbevétel éves változása (d_{28}),
- befektetett tőke hozamának (ROIC) éves változása (d_{50}).

A modellből kikerült változók az alábbiak:

- hosszú távú eladósodottság (m_9),
- összes eszköz forgási időtartama (m_{22}),
- eszközarányos adózott eredmény - ROA (m_{34}).

Megmaradt ugyanakkor az alábbi öt mutató:

- gazdálkodási forma,
- megye,
- kötelezettségek / dologi nettó érték (m_{11}),
- cash-flow fedezet (m_{16}),
- rövid lejáratú kötelezettségek/ dologi nettó érték (m_{21}).

A logit modell is átalakult a dinamikus mutatók bekerülésének hatására. A korábbi 11 mutató helyett már 12 változót használ a modell, amelyből 7 szerepel az eredeti modellben is:

- gazdálkodási forma,
- megye,
- dinamikus likviditás (m_4),
- hosszú távú eladósodottság (m_9),
- cash-flow fedezet (m_{16}),
- rövid lejáratú kötelezettségek / dologi nettó érték (m_{21}),
- szállítóállomány forgási időtartama (m_{27}).

Kikerültek a modellből az alábbi pénzügyi mutatók:

- tőkeellátottság (m_6),
- kötelezettségek / tárgyi eszközök (m_{10}),

- adósságvisszafizető-képesség (m_{17}),
- eszközarányos adózott eredmény - ROA (m_{34}).

Új döntési faktorok lettek a következő mutatók:

- eszközarányos adózás előtti eredmény (m_{39}),
- pénzhányad éves változása (d_3),
- rövid lejáratú kötelezettségek / dologi nettó érték éves változása (d_{21}),
- szállítóállomány forgási időtartamának éves változása (d_{27}),
- befektetett tőke hozamának (ROIC) éves változása (d_{50}).

A neurális háló által figyelembe vett mutatók köre a dinamikus mutatók felhasználásával 48-ra emelkedett. Jól mutatja a neurális háló tanuló algoritmusának viselkedését, hogy azok a mutatók, amelyeket a modell korábban is meghatározónak minősített az előrejelzés szempontjából, ugyanúgy részt vesznek az új besorolási algoritmusban is. Ezekhez adódtak hozzá új elemként az alábbi dinamikus mutatók:

- likviditási gyorsráta éves változása (d_2),
- pénzhányad éves változása (d_3),
- tőkeellátottság éves változása (d_6),
- eladósodottság arányának éves változása (d_7),
- cash-flow fedezet éves változása (d_{16}),
- működőtőke forgásának éves változása (d_{19}),
- rövid lejáratú kötelezettségek/ dologi nettó érték éves változása (d_{21}),
- összes eszköz forgási időtartamának éves változása (d_{22}),
- beszedés időtartamának éves változása (d_{25}),
- vevőállomány forgási időtartamának éves változása (d_{26}),
- szállítóállomány forgási időtartamának éves változása (d_{27}),
- 1 alkalmazottra eső nettó árbevétel éves változása (d_{28}),

- 1 alkalmazottra eső EBITDA éves változása (d_{31}),
- befektetett tőke hozamának (ROIC) éves változása (d_{50}).

Ha a statikus modellekhez hasonlóan összevetjük az egyes modellek által felhasznált mutatók körét, az alábbi táblázatot kapjuk.

<i>Mutatók</i>	<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
Kvalitatív ismérvek			
gazdálkodási forma	X	X	X
megye	X	X	X
ágazat			X
Beszámolóadatok			
pénzügyi ráfordítások			X
fizetett kamatok			X
Likviditási mutatók			
likviditási ráta (m_1)			X
likviditási gyorsráta (m_2)			X
pénzhányad (m_3)			X
dinamikus likviditás (m_4)		X	X
Tőkeszerkezeti mutatók			
tőkeellátottság (m_6)			X
eladósodottság aránya (m_7)			X
idegen tőke / saját tőke (m_8)			X
hosszú távú eladósodottság (m_9)		X	X
kötelezettségek / tárgyi eszközök (m_{10})	X		X
kötelezettségek / dologi nettó érték (m_{11})	X		X

Adósságszolgálati mutatók			
kamatfedezet I. (m_{12})			X
kamatfedezet II. (m_{13})			X
EBITDA fedezet I. (m_{14})			X
EBITDA fedezet II. (m_{15})			X
cash-flow fedezet (m_{16})	X	X	X
adósság visszafizető-képesség (m_{17})			X
Működőtőke mutatók			
működőtőke-ellátottság (m_{18})			X
működőtőke forgása (m_{19})			X
rövid lejáratú kötelezettségek/ dologi nettó érték (m_{21})	X	X	X
Eszközforgási mutatók			
összes eszköz forgási időtartama (m_{22})			X
készletek forgási időtartama I. (m_{23})			X
készletek forgási időtartama II. (m_{24})			X
beszedés időtartama (m_{25})			X
vevőállomány forgási időtartama (m_{26})			X
szállítóállomány forgási időtartama (m_{27})		X	X
Jövedelmezőségi mutatók			
eszközarányos adózott eredmény - ROA (m_{34})			X
adózott eredmény / működőtőke (m_{35})			X
eszközarányos adózás előtti eredmény (m_{39})		X	X
nyereség-visszaforgatás aránya (m_{51})			X
Dinamikus likviditási mutatók			
likviditási gyorsráta éves változása (d_2)			X
pénzhányad éves változása (d_3)		X	X
Dinamikus tőkeszerkezeti mutatók			
tőkeellátottság éves változása (d_6)			X
eladósodottság arányának éves változása (d_7)			X

Dinamikus adósságszolgálati mutatók			
cash-flow fedezet éves változása (d_{16})			X
Dinamikus működőtőke mutatók			
működőtőke forgásának éves változása (d_{19})			X
rövid lejáratú kötelezettségek/ dologi nettó érték éves változása (d_{21})		X	X
Dinamikus eszközforgási mutatók			
összes eszköz forgási időtartamának éves változása (d_{22})			X
beszedés időtartamának éves változása (d_{25})			X
vevőállomány forgási időtartamának éves változása (d_{26})			X
szállítóállomány forgási időtartamának éves változása (d_{27})		X	X
Dinamikus termelékenység mutatók			
1 alkalmazottra eső nettó árbevétel éves változása (d_{28})	X		X
1 alkalmazottra eső EBITDA éves változása (d_{31})			X
Dinamikus jövedelmezőségi mutatók			
befektetett tőke hozamának (ROIC) éves változása (d_{50})	X	X	X

26. táblázat: A dinamikus default-modellek által felhasznált mutatók

Az egyes mutatócsoportok szerepe is átalakult némileg: a statikus modellek esetén a legjelentősebbek a tőkeszerkezeti és adósságszolgálati mutatók voltak. A dinamikus mutatók megjelenésével az adósságszolgálati mutatók szerepe csökkent leginkább, míg a tőkeszerkezeti mutatók fontossága továbbra is megfigyelhető. Sőt, e mutatócsoport szerepe még növekedett is, hiszen két modell is felhasznál dinamikus tőkeszerkezeti mutatót (d_{21}).

T9. A dinamikus ráták közül kiemelkedik a ROIC éves változása mutató, amely mindhárom modellbe bekerült új változóként.

Kimaradt ugyanakkor a ROA (m_{34}) mutató, amely a statikus modellek mindegyikében szerepelt, a dinamikus modellek közül viszont csak a neurális háló alkalmazza. Ez a kettő valószínűleg összefüggésben áll egymással, hiszen a ROA mint statikus

jövedelmezőségi mutató helyét pótolja valószínűleg a nagyobb információértékkel bíró dinamikus ROIC mutató.

A dinamikus modellek közös változói az alábbiak:

- gazdálkodási forma,
- megye,
- cash-flow fedezet (m_{16}),
- rövid lejáratú kötelezettségek/ dologi nettó érték (m_{21}),
- befektetett tőke hozamának (ROIC) éves változása (d_{50}).

A mutatók, amelyeket két modell alkalmaz:

- dinamikus likviditás (m_4),
- hosszú távú eladósodottság (m_9),
- kötelezettségek / tárgyi eszközök (m_{10}),
- kötelezettségek / dologi nettó érték (m_{11}),
- szállítóállomány forgási időtartama (m_{27}),
- eszközarányos adózás előtti eredmény (m_{39}),
- pénzhányad éves változása (d_3),
- rövid lejáratú kötelezettségek/ dologi nettó érték éves változása (d_{21}),
- szállítóállomány forgási időtartamának éves változása (d_{27}),
- 1 alkalmazottra eső nettó árbevétel éves változása (d_{28}).

6.2.2. A dinamikus modellek értékelése

A dinamikus modellek besorolási pontosságait foglalja össze az alábbi táblázat.

		<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
Besorolási pontosság	<i>Fejlesztési minta</i>	85,7%	84,5%	85,7%
	<i>Tesztelési minta</i>	79,6%	82,2%	83,6%
Elsőfajú hiba	<i>Fejlesztési minta</i>	7,6%	8,3%	8,0%
	<i>Tesztelési minta</i>	9,0%	9,8%	9,8%
Másodfajú hiba	<i>Fejlesztési minta</i>	6,7%	7,2%	6,3%
	<i>Tesztelési minta</i>	11,4%	8,0%	6,6%

27. táblázat: A dinamikus mutatók tesztelése során kapott modellek besorolási pontosságai és hibaarányai

Látható, hogy a döntési fa esetén volt a legnagyobb a teljesítménynövekedés: a korábbi 82,9%-os pontosság 85,7%-ra nőtt a fejlesztési mintán, a tesztelési mintán pedig 78%-ról 79,6%-ra emelkedett. A döntési fa fejlesztési mintán elért javulása arra is elég volt, hogy megelőzze a logit modellt, bár a tesztelési mintán mutatott teljesítmény tekintetében még mindig alatta maradt.

A dinamikus modell pontossága a logit esetén is felülmúlja a statikus modellét: a fejlesztési mintán egy százalékpontot javult, 84,5%-os pontosságot ért el. A tesztelési mintán ennél kisebb mértékű volt a javulás, 81,8%-ról 82,2%-ra emelkedett a besorolási pontosság.

A dinamikus mutatókat is felhasználó neurális háló teljesítménye az új változókkal több mint két százalékponttal javult: 85,7%-ra a fejlesztési mintán, illetve 83,6%-ra a tesztelési mintán.

Ha a modellek elsőfajú hibaarányait vizsgáljuk, szembeötlő a döntési fa javulása. A korábbi 12% körüli szintről 7,6%-ra csökkent a fejlesztési mintán, illetve 9%-ra a tesztelési mintán, a modell javulásának oka tehát elsősorban az elsőfajú hibák csökkenéséből ered.

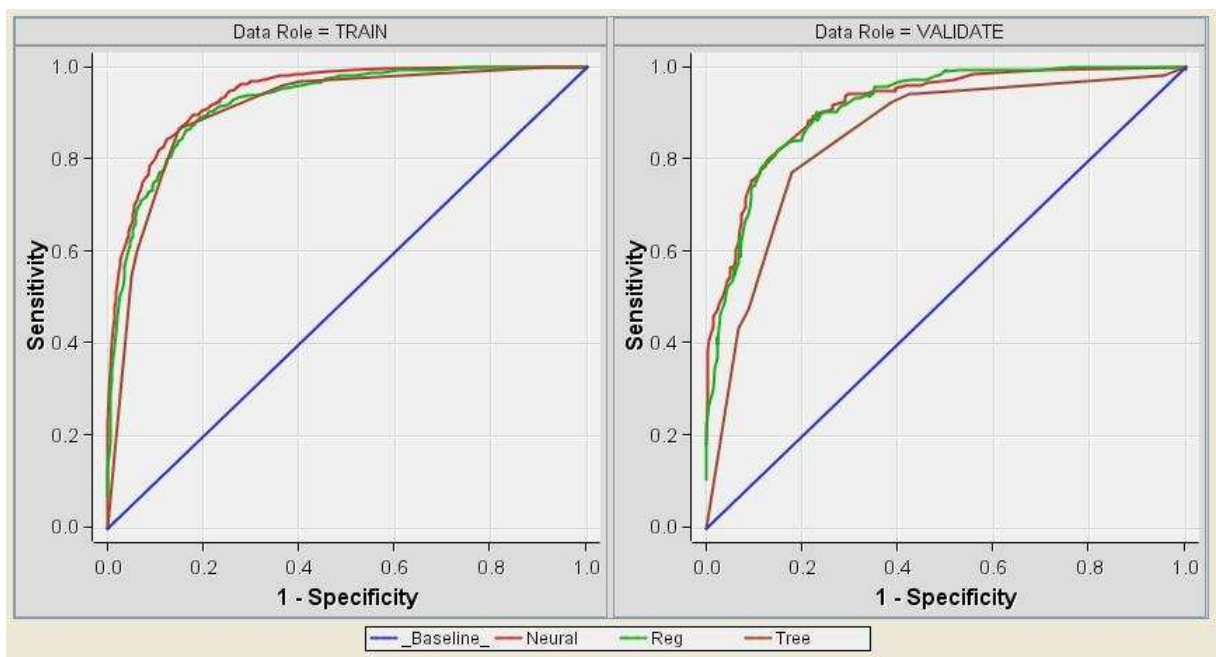
Fontos megjegyezni ugyanakkor, hogy a döntési fa elsőfajú hibáinak csökkenésével párhuzamosan a másodfajú hibák kismértékű, 1-2 százalékpontos emelkedése következett be. Ez pedig a bank szempontjából a nagyobb veszteséggel járó besorolási

hiba, így az általános teljesítménynövekedést elismerve, ez a tény kedvezőtlennek értékelhető.

Ennek pont a fordítottja mondható el a neurális hálóról: a besorolási pontosság növekedésének elsődleges oka a másodfajú hibák kedvezőbb alakulása volt. A korábbi 8%-os szintről 6,3%-ra esett vissza e hibák aránya a fejlesztési mintán, ami a tesztelési mintán is csak kis mértékben, 6,6%-ra emelkedett.

A regressziós modell esetében nem látható hasonló markáns különbség, mindkét fajta hiba aránya csökkenést mutat.

Az ROC-görbéket tekintve a legszembetűnőbb a döntési fa teljesítményjavulása: a fejlesztési mintán a görbéje követi a másik két modell grafikonját, és a tesztelési mintán is megközelítette azokat, bár ott még mindig alattuk marad.



21. ábra: A dinamikus default-modellek ROC-görbéi

Az új modellek további értékelési technikákkal mért teljesítményi adatai az alábbiakban foglalhatók össze.

		<i>Döntési fa</i>	<i>Logit regresszió</i>	<i>Neurális háló</i>
AUROC	<i>Fejlesztési minta</i>	90,5	92,1	93,7
	<i>Tesztelési minta</i>	84,3	91,1	91,5
Gini-koefficiens	<i>Fejlesztési minta</i>	81,0	84,0	87,0
	<i>Tesztelési minta</i>	69,0	82,0	83,0
K-S	<i>Fejlesztési minta</i>	71,3	69,6	71,9
	<i>Tesztelési minta</i>	59,2	67,2	67,2

28. táblázat: A dinamikus default-modellek teljesítménymutatói

Mindhárom értékelési technika megerősíti azt, amit a klasszifikációs mátrix is jelzett: a döntési fa egyértelműen erősödött a dinamikus mutatóknak köszönhetően. A fejlesztési mintán abszolút versenytársa lett a két, korábban egyértelműen jobb modellnek. A tesztelési mintán ugyanakkor nem volt ekkora mértékű a teljesítménynövekedés, így előrejelzési modellként a döntési fa továbbra is a harmadik helyre szorul.

A neurális háló értékelési indexei is azt mutatják, hogy erősödött a modell teljesítménye. Itt is látszik ugyanakkor, hogy a tesztelési mintán mért teljesítménynövekedés elmarad a fejlesztési mintán mért adatokhoz képest.

A logit modell mutatja a legkiegyensúlyozottabb viselkedést: mindhárom mutató esetén másfél-két százalékpontos stabil emelkedést eredményezett a dinamikus mutatók felhasználása. És ami a legfontosabb: az előrelépés nemcsak a fejlesztési, hanem a tesztelési adatokon is megmutatkozik.

Az egyes modelleket összehasonlítva tehát ismét oda értünk, ahol a statikus modelleknél álltunk: adott két, egymással teljesen hasonló pontosságot és előrejelző képességet mutató modell, a logit és a neurális háló, tőlük kissé lemaradva a döntési fa.

Következtetésként ugyanaz mondható el, ami a korábbiakban is: szigorú értelemben véve talán a neurális háló jobbnak bizonyult a logisztikus regressziónál, de előnye annyira elhanyagolható, hogy a nagyságrendekkel több mutató felhasználásból adódó bonyolultság mégis ellene szól.

A logit ezzel szemben stabilabbnak mutatkozott, a fejlesztési mintán mutatott pontosság és a tesztelési minta segítségével becsült előrejelző képesség közötti különbség a logit esetén a legkisebb, és mindezt 12 mutató felhasználásával éri el. A gyakorlati megvalósíthatóság szempontja miatt tehát ismét a logit modellt tekinthetjük győztesnek. A jóval bonyolultabb neurális háló nem tud annyival jobb teljesítményt felmutatni, hogy ezt a tényezőt figyelmen kívül lehetne hagyni.

7. A kutatások folytatásának lehetséges irányai

Bár az elvégzett modellezési munka és a hipotézisek tesztelése meghozta az eredeti célkitűzéseknek megfelelő eredményeket, mégis maradtak olyan nyitott kérdések, amelyek további kutatások tárgyát képezhetik.

Legfontosabbnak tartom ezek között megemlíteni a bináris célváltozó felváltását több lehetséges kimeneteli lehetőséggel. Ezt már korábban, a csődmodellekkel kapcsolatban felsorolt főbb kritikai észrevételek között is kiemeltem.

A vállalatok valóságos pénzügyi helyzete sokkal többféle lehet, mint amennyit ezzel a két szélsőséges esettel jellemezhetünk. Különösen igaz ez a csődbe jutott és túlélő vállalatokra, de ugyanúgy fennáll a default és nem default kategóriák esetén is. Érdemes lenne tehát a célváltozót tovább finomítani, hogy a szélsőségek közötti szürke zóna is feltárható legyen.

Ennek természetesen feltétele egy megfelelő minőségű, országos szintű adatbázis kialakítása. Megtartható lenne a reprezentativitás két szempontú biztosítása, árbevétel és ágazati megoszlás alapján.

A besorolási kategóriáknál legalább három lehetőséget kellene megkülönböztetni, a két szélsőség között húzódó szürke zóna definiálásával. Ez a default modellek esetében a következőképpen képzelhető el, elsősorban a most is felhasznált KHR adatait figyelembe véve:

- a default definíciója változatlan maradna, vagyis a 90 napos lejárt tartozást vennénk figyelembe;
- azok a cégek, amelyeknél a regisztrált késedelem nem éri el a 90 napot, de meghaladja a 30 napot, a köztes kategóriába kerülnének;
- azok a vállalatok, amelyek esetében nem regisztráltak 30 napos késedelmet meghaladó tartozást, képeznék a megfelelően működő kategóriát.

Természetesen ez a megközelítés megkövetelné az alkalmazott modellezési technikák adaptációját is, ugyanis nem minden módszertan alkalmas kettőnél több célváltozó kezelésére (ilyen például a legeredményesebbnek bizonyult logit modell is).

Ezt a módszertani kérdést át lehetne hidalni olyan megoldással, hogy a köztes kategóriába tartozó vállalatokat egyszerűen kivonnánk a vizsgált mintából. Ily módon alkalmazhatók lennének a bináris célváltozóval dolgozó modellek, de az elemzés mégis finomodna azáltal, hogy két, egymástól markánsabban elkülönülő csoportot próbál meg szétválasztani, mivel a nem defaultos kategóriát egyértelműbben definiáltuk, finomítottuk az oda sorolt vállalatok profilját.

Érdemes lenne további teszteket végezni arra vonatkozóan is, hogy hol húzhatók meg a szürke zóna határai. A problémamentes kategória 30 napos definícióját azért tekintem egyértelműnek, mivel a rendszer által adott: egy ügyfél a KHR-ben akkor minősül problémásnak, ha tartozása meghaladja a 30 napot. A default meghatározása úgy történt, hogy az ügyfél a rendszerben már legalább 60 napja szerepeljen rossz adósként. Ezt a határt a bázeli egyezmény rögzíti.

A köztes kategóriára ugyanakkor nincsenek ilyen értelemben vett szabályok: megvizsgálható lenne, hogy pl. 30-90 napos kategória helyett máshol (pl. 45, 60, 75 napos határpontokat alkalmazva lenne-e érdemes köztes kategóriába sorolni a tartozó vállalatokat.

A modellépítés során természetesen elvégezhető lenne az általam is vizsgált iparági, illetve dinamikus mutatók felhasználásával nyerhető teljesítménynövekedés is.

A több kimeneti kategóriát figyelembe vevő modellek további lehetőségeket is rejtenek: később vizsgálható és modellezhető lenne az egyes fizetőképességi kategóriák közötti változások, mozgások előrejelzése is.

Érdekes tesztet lenne elvégezhető, ha az adatbázis keresztmetszeti felosztása helyett annak időszakok szerinti particionálását alkalmaznánk, vagyis az adatbázis fejlesztési és tesztmintára való szétválasztása helyett az adatokat a megfigyelési időszak alapján

osztánánk két részre. Azaz pl. a 2002-2004 közötti időszak adatain kerülnének kidolgozásra a modellek, amelyeket aztán a 2005-2006 közötti adatokon tesztelnénk.³⁷

A fenti vizsgálatok véleményem szerint megvalósíthatók lennének a KHR adatbázis alapján, ha lehetőséget biztosítanak a hozzáféréshez és az adatok kinyeréséhez.

³⁷ Megjegyzendő ugyanakkor, hogy ez nem teljesen illeszkedik a feldolgozott szakirodalomban megismert technikákhoz, illetve maga a Bázeli Tőkeegyezmény is a keresztmetszeti mintákat támogatja. Ennek oka elsősorban azok nagyobb fokú stabilitása, mivel a keresztmetszeti minták alkalmazásával csökkenthetők az egyes időszakok közötti külső változások hatásai. Egy teszt elvégzése mindettől függetlenül érdekes lenne.

A teszt elvégzésére sajnos a rendelkezésre álló adatok nem adnak lehetőséget, mivel a fizetési késedelem bekövetkezését csak dummy változóként tartalmazza az adatbázis, amely nem ad lehetőséget az időszakok szerinti felosztásra, ez pedig torzítást eredményezne.

8. Összefoglalás

Értekezésem első részében a csődmodellek történetét kíséreltem meg átfogóan bemutatni, az alkalmazott módszertan szerinti megközelítéssel.

A csődmodellezés szakirodalmának feldolgozása során összesen 126 modellt vizsgáltam, amelyeket az alkalmazott módszertan szempontjából csoportosítottam. A kutatások időrendiségét és módszertanát figyelembe véve a csődmodellek fejlődéstörténetének 4 szakaszát különböztettem meg:

0. *A csődkutatások kialakulásának korszaka (1931-67)*, amely nem igazán eredményezett mai értelemben vett csődmodellt, inkább a csőd, mint gazdasági esemény előrejelzésének vizsgálata köthető ehhez az időszakhoz, amolyan nulladik szakasznak tekinthető.
1. *A diszkriminancia analízis korszaka (1968-79)*. A korszakot az első csődmodell kidolgozója, Altman neve és a diszkriminancia analízis elterjedése jellemzi. Bár még előfordulnak egyváltozós elemzések, a domináns módszertan egyértelműen a diszkriminancia analízis.
2. *A logisztikus regresszió térhódításának korszaka (1980-1989)*. A módszer első úttörője Ohlson volt, és igen gyorsan elterjedt, mind a mai napig az egyik leggyakrabban alkalmazott modellezési módszernek, „best practice”-nek tekinthető.
3. *A mesterséges intelligencia megjelenésének korszaka (1990-től napjainkig)*. A mesterséges intelligencia a neurális hálók technikájával van jelen a csődmodell-irodalomban. 1990-es bevezetése Odom és Sharda nevéhez kötődik, és napjainkban is ez a legjelentősebb témája a csődmodellek módszertani kutatásainak. A kutatások többsége azt a következtetést vonja le, hogy a neurális hálók eredményei és előrejelző képessége jobb, mint a korábbi módszerekkel készült modelleké.

Azonosítottam továbbá két fejlődési mellékágot: a probit és a rekurzív felosztás módszereinek alkalmazásával készült modelleket. Ezek a módszerek nem terjedtek el

annyira a szakirodalomban, bár időről-időre kísérleteztek velük, mégsem tudták megtörni a mainstream modellezési technikák dominanciáját.

Az értekezés második részében a csődmodellekhez hasonló előrejelzési modellekkel kísérleteztem. Célom olyan prediktív modellek kialakítása volt, amelyek a Bázeli II. definícióknak megfelelően a vállalatok nemfizetését, 90 napos késedelembe esését jelzik előre.

A rendelkezésemre álló adatbázis 2000 magyarországi vállalat adatait tartalmazta a 2002-2006 közötti időszakra vonatkozóan. Az adatbázis felhasználásával több vizsgálatot is elvégeztem, három különböző modellezési technika felhasználásával (döntési fa, logisztikus regresszió és neurális háló).

Megállapítottam, hogy a csődmodellek mintájára készíthető olyan prediktív modell, amely a vállalatok 90 napos fizetési késedelmét (defaultját) jelzi előre. Az előrejelzés pontossága ugyanakkor némileg elmarad a csődmodellektől, amelynek magyarázata az lehet, hogy a pénzügyileg egészséges és a problémás vállalatok profilja valószínűleg még nem különbözik olyan markánsan egymástól, mint a csődbe ment és a túlélő vállalatoké, ezért az előrejelzés is nehezebb, pontatlanabb. A rendelkezésemre álló adatbázis felhasználásával a legjobb előrejelző modell is csak 82% körüli pontosságot ért el, míg a hasonló magyarországi csődmodellek 90%-ot is meghaladó pontossággal jelzik a csődvesztélyt.

Az általam készített modellek mindegyike felhasználja az alábbi hat mutatót:

- gazdálkodási forma,
- megye,
- hosszú távú eladósodottság,
- cash-flow fedezet,
- rövid lejáratú kötelezettségek/ dologi nettó érték,
- eszközarányos adózott eredmény (ROA).

A vizsgált modellezés technikák közül a döntési fa bizonyult a legkevésbé hatékony módszernek, a logisztikus regresszió és a neurális háló modelljei körülbelül ugyanúgy teljesítettek. A két modell közül a könnyebb megvalósíthatóság és alkalmazhatóság mégis a logit mellett szólnak, mivel ugyanazt a teljesítményt a logit modell 11 döntési tényező felhasználásával éri el, szemben a jóval bonyolultabb, 34 faktort alkalmazó neurális hálóval.

A rendelkezésre álló adatbázis segítségével igazoltam továbbá, hogy az egyes pénzügyi mutatók éves változásait kifejező ún. dinamikus mutatók felhasználásával növelhető az előrejelzési modellek besorolási pontossága.

A dinamikus mutatók révén az egyes modellek teljesítménye 1-2 százalékponttal javult, a besorolások pontossága 85% körülire emelkedett a fejlesztési mintán, az előrejelzések becslésére szolgáló tesztelési mintán pedig kb. 83%-os szintre nőtt.

A dinamikus ráták közül kiemelkedik a ROIC éves változása mutató, amely mindhárom modellbe bekerült új változóként.

Az egyes modellezési technikák összehasonlításakor továbbra is a döntési fa maradt a harmadik helyen, a logit és a neurális háló közül az utóbbi kissé jobban teljesített. Továbbra is fennáll ugyanakkor, hogy a logisztikus regressziós modell stabilabbnak mutatkozott, és a gyakorlati megvalósíthatóság szempontja is emellett szól.

Nem sikerült azonban igazolnom azt a több szakirodalom által is bemutatott eredményt, hogy az iparági ráták javítják a csődmodellek besorolási pontosságát. A rendelkezésemre álló adatokon és az alkalmazott definíciós háttérrel azt tapasztaltam, hogy az iparági ráták felhasználásával képzett mutatók nem kerültek be a modellekbe, besorolási pontosságaik pedig gyakorlatilag változatlan szinten maradtak.

Egyedül a neurális háló esetén volt megfigyelhető némi változás, pontossága kb. másfél százalékpontot javult a fejlesztési mintán. Az előrejelzési képesség becslésére használt tesztelési mintán ellenben nem volt látható semmilyen különbség, még némileg csökkent is a modell pontossága. A másik két modell esetén még ekkora változást sem hoztak az iparági mutatók.

9. Felhasznált irodalom

Alici, Y. (1995): *Neural Networks in Corporate Failure Prediction: The UK Experience*, Mimeo, 1995.

Altman, E. I. (1968): *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*. *Journal of Finance*, September 1968.

Altman, E. I. – Haldeman, R. – Narayanan, P. (1977): *Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations*. *Journal of Banking & Finance*, 1, 1977.

Altman, E.I. – Izan, H.Y. (1984): *Identifying corporate distress in Australia: An industry relative analysis*. Working Paper, New York University, 1984.

Altman, E.I. - Levallee, M.Y. (1980): *Business Failure Classification in Canada*. *Journal of Business Administration*, vol. 12, no. 1, pp. 147-164.

Altman, E.I. - Lorris, B. (1976): *A financial early warning system for over-the counter broker-dealers*. *Journal of Finance*, 31.4. (September 1976) pp. 1201-1217.

Altman, E. I. - Marco, G. - Veretto, F. (1994): *Corporate Distress Diagnosis: Comparisons using Linear Discriminant Analysis and Neural Networks*. *Journal of Banking and Finance*, vol.18, No.3, pp. 505-529.

Andersen, H. (2008): *Failure prediction of Norwegian banks: A logit approach*. Norges Bank Working Paper, No. 2., 2008.

Atiya, A.F. (2001): *Bankruptcy prediction for credit risk using neural networks: A survey and new results*. *IEEE Transactions on Neural Networks*, Vol. 12. Nr. 4, July 2001. pp. 929-935.

Awh, R.Y. – Waters, D. (1974): *A Discriminant Analysis of Economic, Demographic, and Attitudinal Characteristics of Bank Change-Card Holders: A Case Study*. *The Journal of Finance*, Vol. 29, pp. 973-980.

Aziz, A. – Emanuel, D.C. – Lawson, G.H. (1988): *Bankruptcy prediction – An investigation of cash flow based models*. Journal of Management Studies, Vol. 25, nr. 5, pp. 419-437.

Back, B. – Laitinen, T. – Sere, K. - van Wezel, M. (1996): *Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms*. Turku Centre for Computer Science Technical Report 40 (September 1996).

Bahnson, P.R. - Bartley, J.W. (1992): *The Sensitivity of Failure Prediction Models to Alternative Definitions of Failure*. Advances in Accounting, vol. 10, pp. 255-278.

Balcaen, S. - Ooghe, H. (2004): *35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and their Related Problems*. Vlerick Leuven Gent Working Paper Series 2004/15.

Bardos, M. – Zhu, W. (1997): *Comparison of discriminant analysis and neural networks: Application for the detection of company failures*. In: Biometric approaches in management science, Kluwer Academic Publishers, 1997. pp.1-25.

Barniv, R. - Raveh, A. (1989): *Identifying Financial Distress: A New Nonparametric Approach*. Journal of Business Finance and Accounting, vol. 16, no. 3, pp. 361-383.

Barth, J.R. – Brumbaugh, R.D. – Sauerhaft, D. – Wang, G.H.K. (1985): *Thrift-Institution Failures: Causes and Policy Issues*. Bank Structure and Competition, Federal Reserve Bank of Chicago, 1985. pp. 184-216.

Beaver, W. H (1966): *Financial Ratios as Predictors of Failures*. Empirical Research in Accounting, Selected Studies, 1966.

Berg, S.A. - Hexeberg, B. (1994): *Early warning indicators for Norwegian banks: A logit analysis of the experiences from the banking crisis*. Norges Bank Working Paper, No. 1., 1994.

Bernhardsen (2001): *A model of bankruptcy prediction*. Working Papers, Norges Bank, 2001/10.

Betts, J. – Belhoul, D. (1987): *The effectiveness of incorporating stability measures in company failure models*. Journal of Business Finance and Accounting, Vol. 14. Nr. 3., pp. 323-334.

Bilderbeek, J. (1973): Financieel-economische indices ten behoeve van de bedrijfsbeoordeling. Economisch en Sociaal Tijdschrift, Vol 27. Nr. 2., April 1973. pp. 141-155.

Bilderbeek, J. (1979): *An empirical study of the predictive ability of financial ratios in the Netherlands*. Zeitschrift Für Betriebswirtschaft, May 1979, pp. 388-407.

Blum, M. (1974): *Failing company discriminant analysis*. Journal of Accounting Research, Vol. Spring pp.1-21.

Bodon F. (2004): *Adatbányászati algoritmusok*.
<http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf>,
2007. július 17-én.

Bongini, P. - Ferri, G. - Hahm, H. (2000): *Corporate Bankruptcy in Korea: Only the Strong Survive?* Financial Review, vol. 35, no. 4, pp. 31-51.

Borgulya I. (1998): *Neurális hálók és fuzzy-rendszerek*. Dialóg Campus Kiadó, Budapest-Pécs, 1998.

Boritz, J.E. - Kennedy, D.B. - Albuquerque, A.M. (1995): *Predicting Corporate Failure Using a Neural Network Approach*. Intelligent Systems in Accounting, Finance and Management, vol. 4, no. 2, pp. 95-111.

Cadden, D. (1991): *Neural networks and the mathematics of chaos – An investigation of these methodologies as accurate predictions of corporate bankruptcy*. Working Paper, New York, 1991.

Casey, C. - Bartczak, N. (1985): *Using Operating Cash Flow Data to Predict Financial Distress: Some Extensions*. Journal of Accounting Research, vol. 23, pp. 384-402.

Castagna, A.D. – Matolcsy, S.P. (1981): *The Prediction of Corporate Failure: Testing the Australian Experience*. Australian Journal of Management, June 1981. pp.23-50.

Charitou, A. - Neophytou, E. - Charalambous, C. (2004): *Predicting Corporate Failure: Empirical Evidence for the UK*. European Accounting Review, vol. 13, no. 3, pp. 465-497.

Chung, H.M. – Tam, K.Y. (1993): *A comparative analysis of inductive-learning algorithms*. International Journal of Intelligent Systems in Accounting, Finance and Management, Vol. 2, Nr. 1, January 1993. pp. 3-18.

Coats, P.K. – Fant, L.F. (1991): *A neural network approach to forecasting financial distress*. The journal of Business Forecasting, Vol. 10. Nr. 4, Winter 1991-1992. pp. 9-12.

Coats, P.K. – Fant, L.F. (1993): *Recognising financial distress patterns using a neural network tool*. Financial Management, Vol. 22. Nr. 3. pp. 142-155.

Collins, R. - Green, R. (1982): *Statistical methods for bankruptcy prediction*. Journal of Economics and Business, 34(4), pp. 349–354.

Dambolena, I. – Khoury, S. (1980): *Ratio stability and corporate failure*. Journal of Finance, Vol. 33, nr. 4, pp. 1017-1026.

Dambolena, I. - Shulman, J.M. (1988): *A Primary Rule for Detecting Bankruptcy: Watch the Cash*. Financial Analyst Journal, vol. 44, no. 5, pp. 74-79.

Darayseh, M. - Waples, E. - Tsoukalas, D. (2003): *Corporate Failure for Manufacturing Industries using Firms Specific and Economic Environment with Logit Analysis*. Managerial Finance, vol. 29, no. 8, pp. 23-36.

Deakin, E.B. (1972): *A Discriminant analysis of predictors of business failure*. Journal of Accounting Research, 10-1 (Spring 1972) pp. 167-179.

Deutsche Bundesbank (2003): *Approaches to the validation of internal rating systems*. Monthly Report, September 2003.

Edmister, R. (1972): *An empirical test of financial ratio analysis for small business failure prediction*. Journal of Financial and Quantitative Analysis, March 1972, pp. 1477-1493.

El Hennawy, R. - Morris, R. (1983): *The significance of base year in developing failure prediction models*. Journal of Business Finance & Accounting, Summer 1983. pp. 209-223.

Elam, R. (1975): *The Effect of Lease Data on the Predictive Ability of Financial Ratios*. Accounting Review, vol. 50, no. 1, pp. 25-43.

Fair Isaac (2006): *Introduction to Model Builder Scorecard*. Fair Isaac White paper, November 2006.

Fawcett, T. (2004): *ROC Graphs: Notes and Practical Considerations for Researchers*. HP Laboratories, Technical Report, Palo Alto, CA USA. March 2004.

http://home.comcast.net/~tom.fawcett/public_html/papers/ROC101.pdf, 2008.07.21-én.

Fitzpatrick, P. (1932): *A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies*. The Accountants Publishing Company, 1932.

Flagg, J.C. - Giroux, G.A. (1991): *Predicting Corporate Bankruptcy Using Failing Firms*. Review of Financial Economics, vol. 1, pp. 67-78.

Fletcher, D. - Goss, E. (1993): *Forecasting with Neural Networks: An Application Using Bankruptcy Data*. Information and Management, vol. 24, pp. 159-167.

Frydman, H. – Altman, E.I. – Kao, D-L. (1985): *Introducing recursive partitioning for financial classification: The Case of financial distress*. The Journal of Finance XI-1 (March 1985).

Gentry, J.A. – Newbold, P. – Whitford, D.T. (1985): *Classifying bankrupt firms with funds flow components*. Journal of Accounting Research, Vol. 23, nr. 1, Spring 1985, pp. 146-160.

- Gilbert, L.R. - Menon, K. - Schwartz, K.B. (1990): *Predicting Bankruptcy for Firms in Financial Distress*. Journal of Business Finance and Accounting, vol. 17, no. 1, pp. 161-171.
- Ginoglou, D. - Agorastos, K. - Hatzigagios, T. (2002): *Predicting Corporate Failure of Problematic Firms in Greece with LPM, Logit, Probit and Discriminant Analysis Models*. Journal of Financial Management and Analysis, vol. 15, no. 1, pp. 1-16.
- Glennon, D. (2006): *Building and Validating Credit Rating and Scoring Models*. OCC Workshop Session 1, Washington, February 2006.
www.occ.treas.gov/vcrsm/doc/Session1_Glennon.pdf, 2007. augusztus 11-én.
- Gloubos, G. – Grammatikos, T. (1988): *The success of bankruptcy prediction models in Greece*. Studies in Banking and Finance, Vol. 7, pp. 37-46.
- Gombola, M. – Haskins, M. – Ketz, J. – Williams, D. (1987): *Cash flow in bankruptcy prediction*. Financial Management, Winter 1987, pp. 55-65.
- Gritta, R.D. - Wang, M. - Davalos, S. - Chow, G. (2000): *Forecasting Small Air Carrier Bankruptcies Using a Neural Network Approach*. Journal of Financial Management and Analysis, vol. 13, no. 1, pp. 44-50.
- Hajdu O. (2003): *Többváltozós statisztikai számítások*. Központi Statisztikai Hivatal, Budapest, 2003.
- Hajdu O. (2004): *Statisztika III*. BKÁE Statisztika tanszék, Budapest, 2004.
- Halas G. (2004): *Hogyan mérjük nemfizetési valószínűséget? Módszertani összehasonlítás*. Mérés Határ Műhely, 2004. szeptember.
- Hamer, M.M. (1983): *Failure Prediction: Sensitivity of Classification Accuracy to Alternative Statistical Methods and Variable Sets*. Journal of Accounting and Public Policy, vol. 2, no. 4, pp. 289-307.
- Hámori G. (2001): *A fizetésképtelenség előrejelzése logit-moddal*. Bankszemle, 45. évf, 1-2 szám, 65-87. old.

Hill, N.T. - Perry, S.E. (1996): *Evaluating Firms in Financial Distress: An Event History Analysis*. Journal of Applied Business Research, vol. 12, no. 3, pp. 60-72.

Hopwood, W - McKeown, J.C. - Mutchler, J.F. (1989): *A Test of the Incremental Explanatory Power of Opinions Qualified for Consistency and Uncertainty*. Accounting Review, vol. 64, no. 1, pp. 28-48.

Horrigan, J. L. (1966): *The determination of long term credit standing with financial ratios*. Journal of Accounting Research, Vol.4. 1966.

Hosner, D. - Lemeshow, S. (2000): *Applied Logistic Regression*. John Wiley, New York, 2000.

Hossari, G. (2007): *Benchmarking New Statistical Techniques in Ratio-Based Modelling of Corporate Collapse*. International Review of Business Research Papers, Vol. 3. No. 3. August 2007. pp.141-161.

Hunter, J. - Isachenkova, N. (2001): *Failure risk: A comparative study of UK and Russian firms*. Journal of Policy Modeling, Elsevier, vol. 23(5) July, pp. 511-521.

Imre B. (2007a): *A csődmodellek története és fejlődése*. E-tudomány, 2007. évfolyam, 3. szám. pp. III-1-42.

Imre B. (2007b): *Prediktív modellek értékelési módszerei*. E-tudomány, 2007. évfolyam, 4. szám. pp. I-1-10.

Info-Datex (2006): *Módszertani elemzés a nemfizetési valószínűség modellezéséhez*. PSzÁF pályázati tanulmány, 2006. július.

http://www.pszaf.hu/resource.aspx?ResourceID=pszafhu_palyamunka_nemfizetes, 2007. július 10-én.

Johnsen, T. - Melicher, R.W. (1994): *Predicting Corporate Bankruptcy and Financial Distress: Information Value Added by Multinomial Logit Models*. Journal of Economics and Business, vol. 46, no. 4, pp. 269-286.

Keasey, K. - McGuinness, P. (1990): *The failure of UK industrial firms for the period 1976-1984: Logistic analysis and entropy measures*. Journal of Business Finance and Accounting, vol. 17. no. 1. pp. 119-135.

Keasey, K. - Watson, R. (1987): *Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti's Hypothesis*. Journal of Business Finance and Accounting, vol. 14, no. 3, pp. 335-355.

Kerling M. - Poddig T. (1994): *Klassifikation von Unternehmen mittels KNN*. In: Rehkugler, H. - Zimmermann, H. G. (szerk) *Neuronale Netze in der Ökonomie*, München, 1994.

Ketz, J.E. (1978): *The Effect of General Price-Level Adjustments on the Predictive Ability of Financial Ratios*. Journal of Accounting Research, vol. 16, no. 3, pp. 273-285.

Kiss F. (2003): *A credit scoring fejlődése és alkalmazása*. PhD. doktori értekezés, Budapesti Műszaki Egyetem, 2003.

Kiviluoto, K. (1998): *Predicting Bankruptcies with the selforganizing map*. Neurocomputing 21 (1998) pp. 191-201.

Lacher, R.C. - Coats, P.K. - Sharma, S.C. - Fant, L.F. (1995): *A Neural Network for Classifying the Financial Health of a Firm*. European Journal of Operational Research, vol. 85, no. 1, pp. 53-65.

Laitinen, T. - Kankaanpää, M. (1999): *Comparative analysis of failure prediction methods: the finnish case*. The European Accounting Review 8:1 (1999) pp. 67-92.

Lane, S. (1972): *Submarginal Credit Risk Classification*. Journal of Financial and Quantitative Analysis, January 1972.

Lau, A.H. (1987): *A Five-State Financial Distress Prediction Model*. Journal of Accounting Research, vol. 25, no. 1, pp. 127-139.

Lennox, C. (1999): *Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches*. Journal of Economics and Business, vol. 51, no. 4, pp. 347-364.

Leshno, M. – Spector, Y. (1996): *Neural network prediction analysis: The bankruptcy case*. Neurocomputing, Vol. 10. March 1996, pp. 125-147.

Levitan, A.S. - Knoblett, J.A. (1985): *Indicators of Exceptions to the Going Concern Assumption*. Auditing: A Journal of Practice and Theory, Fall 1985. pp. 26-39.

Libby, R. (1975): *Accounting ratios and the prediction of failure: Some behavioral evidence*. Journal of Accounting Research, Spring 1975, pp. 150-161.

Lin, F.Y. - McClean, S. (2001): *A Data Mining Approach to the Prediction of Corporate Failure*. Knowledge-Based Systems, vol. 14, no. 3, pp. 189-195.

Lincoln, M. (1984): *An Empirical Study of the Usefulness of Accounting Ratios to Describe Levels of Insolvency Risk*. Journal of Banking and Finance, vol. 8, no.2, pp. 321-340.

Lo, A.W. (1986): *Logit versus Discriminant Analysis: A Specification Test and Application to Corporate Bankruptcies*. Journal of Econometrics, vol. 31, no. 2, pp. 151-178.

Lussier, R.N. (1995): *A nonfinancial business success versus failure prediction model for young firms*. Journal of Small Business Management, Vol. 33, nr. 1, pp. 8-20.

Martin, D. (1977): *Early-warning of Bank Failure: A Logit Regression Approach*. Journal of Banking and Finance, 1977, no.1, pp. 249-276.

Mays, E. (2004): *Credit Scoring for Risk Managers. The Handbook for Lenders*. South-Western Publishers, Ohio, 2004.

McGurr, P.T. - Devaney, S.A. (1998): *A Retail Failure Prediction Model*. International Review of Retail, vol. 8, no. 3, pp. 259-277.

McKee, T.E. - Greenstein M. (2000): *Predicting bankruptcy using recursive partitioning and a realistically proportioned data set*. Journal of Forecasting, 2000, no.19. pp. 219-230.

Mears, P.K. (1966): *Discussion of Financial Ratios As Predictors of Failure*. Journal of Accounting Research, Vol. 4. 1966., pp. 119-122.

Merwin, C. (1942): *Financing small corporations: in five manufacturing industries, 1926-36*. National Bureau of Economic Research, New York, 1942.

Micha, B. (1984): *Analysis of business failures in France*. Journal of Banking and Finance, Vol. 8, pp. 281-291.

Neophytou, E. - Mar-Molinero, C. (2001): *Predicting Corporate Failure in the UK: A Multidimensional Scaling Approach*. Discussion Paper in Accounting and Management Science. University of Southampton, 2001. pp. 99-150.

Neophytou, E. - Mar-Molinero, C. (2004). *Predicting Corporate Failure in the UK: A Multidimensional Scaling Approach*. Journal of Business Finance and Accounting, vol. 31, no. 5, pp. 677-710.

Neter, J. (1966): *Discussion of financial ratios as predictors of failure*. Journal of Accounting Research, Vol.4. 1966, pp. 112-118.

Norton, C.L. - Smith, R.E. (1979): *A Comparison of General Price-Level and Historical Cost Financial Statements in the Prediction of Bankruptcy*. Accounting Review, vol. 54, no. 1, pp. 78-94.

Odom, M.D. - Sharda, R. (1990): *A Neural Network Model for Bankruptcy Prediction*. IJCNN International Joint Conference on Neural Networks, Vol. 2, San Diego, CA, 1990. pp. 163-167.

Ohlson, J.A. (1980): *Financial Ratios and the Probabilistic Prediction of Bankruptcy*. Journal of Accounting Research, Vol. 18, No. 1 (Spring 1980), pp. 109-131.

Olmeda, I.– Fernandez, E. (1997): *Hybrid Classifiers for Financial Multicriteria Decision Making: The Case of Bankruptcy Prediction*. Computational Economics, Vol. 10. Issue 4. pp. 317–352.

Ooghe, H. - Joos, P. - De Vos, D. (1991): *Failure prediction models*. PhD thesis, Department of Corporate Finance, Ghent University, Belgium, 1991.

Ooghe, H. - Verbaere, E. (1982): *Determinanten van faling: verklaring en predictie*. Whitepaper, Ghent University, Belgium, 1982.

Pantalone, C.C., - Platt, M.B. (1987): *Predicting Commercial Bank Failure Since Deregulation*. New England Economic Review, 1987, Jul-Aug, pp. 37-47.

Peel, M.J. - Peel, D.A. (1987): *Some Further Empirical Evidence on Predicting Private Company Failure*. Accounting and Business Research, vol. 18, no. 69, pp. 57-66.

Piramuthu, S. - Raghavan, H. - Shaw, M. (1998): *Using feature construction to improve the performance of neural networks*. Management Science, 44. pp. 416-430.

Platt, H.D. - Platt, M.B. (1990): *Development of a Class of Stable Predictive Variables: The Case of Bankruptcy Prediction*. Journal of Business Finance and Accounting, vol. 17. no. 1. pp. 31-44.

Platt, H.D. - Platt, M.B. - Pedersen, J.G. (1994): *Bankruptcy Discrimination with Real Variables*. Journal of Business Finance and Accounting, vol. 21, no. 4, pp. 491-515.

Pompe, P.P.M. - Bilderbeek, J. (2000): *Prestaties van modellen en ratio's bij het voorspellen van faillissementen*. Maandblad voor Accountancy en Bedrijfseconomie, Oktober 2000. pp. 465-472.

Pompe, P.P.M. – Feelders, A.J. (1997): *Using Machine Learning, Neural Networks, and Statistics to Predict Corporate Bankruptcy*. Computer-Aided Civil and Infrastructure Engineering, Vol. 12. No. 4, July 1997. pp. 267-276.

Ramster, J. – Foster, L. (1931): *A Demonstration of Ratio Analysis*. 1931.

Richardson, F.M. - Kane, G.D. - Lobinger, P. (1998): *The Impact of Recession on the Prediction of Corporate Failure*. Journal of Business Finance and Accounting, vol. 25, no. 1-2, pp. 167-186.

Rudolph, P.M. - Hamdan, B. (1988): *An Analysis of Post-Deregulation Savings and Loan Failures*. AREUEA Journal 16.1. pp. 17 – 33.

Salchenberger, L.M. – Cinar, E.M. - Lash, N.A. (1992): *Neural networks: a new tool for predicting thrift failures*. Decision Science Vol 23, No. 4, July/August, pp. 899-916.

SAS Institute (2001): *Building Consumer Credit Scoring Models with SAS Enterprise Miner*. SAS Institute Inc, Cary, NC, 2001.

SAS Institute (2003): *Data Mining Using SAS Enterprise Miner. A Case Study Approach, Second Edition*. SAS Institute Inc, Cary, NC, 2003.

SAS Institute (2006): *Getting Started with SAS Enterprise Miner 5.2*. SAS Institute Inc, Cary, NC, 2006.

<http://support.sas.com/documentation/onlinedoc/miner/getstarted.pdf>

Sharma, S. - Mahajan, V. (1980): *Early Warning Indicators of Business Failure*. Journal of Marketing, vol. 44, no. 4, pp. 80-89.

Sheppard, J.P. - Fraser, S. (1994): *The Dilemma of Matched Pairs and Diversified Firms in Bankruptcy Prediction Models*. Mid-Atlantic Journal of Business, vol. 30, no. 1, pp. 9-26.

Siddiqi, N. (2006): *Credit Risk Scorecards. Developing and Implementing Intelligent Credit Scoring*. John Wiley & Sons, Hoboken, New Jersey, 2006.

Sinkey, J.F. (1975): *A Multivariate analysis of the characteristics of problem banks*. The Journal of Finance, Vol. 30, March 1975, pp. 21-36.

Sjovoll, E. (1999): *Assessment of credit risk in the Norwegian business sector*. Norges Bank Working Paper, August 1999.

Skogsvik, K. (1990): *Current Cost Accounting Ratios as Predictors of Business Failure: The Swedish Case*. Journal of Business Finance and Accounting, vol. 17, no. 1, pp. 137-160.

Stickney, C. (1990): *Financial Statement Analysis: A strategic perspective*. Harcourt Brace Jovanovich Publishers, San Diego, CA, 1990.

Taffler, R.J. (1983): *The assessment of company solvency and performance using a statistical model*. Accounting and Business Research, Vol. 15, nr. 52, Autumn 1983, pp. 295-307.

Taffler, R.J. – Tisshaw, H. (1977): *Going, Going, Gone - Four Factors Which Predict*. Accountancy, Vol. 88, March 1977, pp. 50-54.

Takahashi, K. - Kurokawa, Y. (1984): *Corporate Bankruptcy Prediction in Japan*. Journal of Banking and Finance, vol. 8, no. 2, pp. 229-247.

Tam, K.Y. - Kiang, M.Y. (1992): *Managerial Applications of Neural Networks - The Case of Bank Failure Predictions*. Management Science, Vol. 38, No. 7. pp. 926-947.

Tan, C.N.W. (1999): *An Artificial Neural Networks Primer with Financial Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System*. Bond University, Australia, 1999.

Thomas, L.C. – Edelman, D.B. – Crook, J.N. (2002): *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics, Philadelphia, 2002.

Udo, G. (1993): *Neural network performance on the bankruptcy classification problem*. Computers and Industrial Engineering, Vol. 25, September 1993. pp. 377-380.

van Frederikslust, R.A.I. (1978): *Predictability of corporate failure: models for prediction of corporate failure and for evaluation of corporate debt capacity*. Martinus Nijhoff Social Science Division, Leiden/Boston, PhD thesis in Economic Sciences, Erasmus University, Rotterdam, The Netherlands.

Virág M. (1996): *Pénzügyi elemzés, csődelőrejelzés*. Kossuth Könyvkiadó, Budapest, 1996.

Virág M. (2004): *A csődmodellek jellegzetességei és története*. Vezetéstudomány, 35. évf. 10. sz. 24-32. old.

Virág M. – Hajdu O. (1993): Pénzügyi viszonyszámokon alapuló vállalatminősítés többváltozós statisztikai módszerek felhasználásával. Ipar-Gazdaság, 1993. 7 sz. pp. 23-32.

Virág M. – Hajdu O. (1996): *Pénzügyi mutatószámokon alapuló csődmodell-számítások*. Bankszemle, 15. évf. 5. sz. 42-53. old.

Virág M. – Hajdu O. (1998): *Pénzügyi viszonyszámok és a csődelőrejelzés*. Bankról, pénzről, tőzsdéről. Válogatott előadások a Bankárképzőben 1988-1998. Budapest, 1998. 440-457. old.

Virág M. – Kristóf T. (2005): *Az első hazai csődmodell újraszámítása neurális hálók segítségével*. Közgazdasági Szemle, 52. évf. 2. sz. 144-162. old.

Virág M. – Kristóf T. (2006): *Iparági rátákon alapuló csődelőrejelzés sokváltozós statisztikai módszerekkel*. Vezetéstudomány, XXXVII. évfolyam, 2006. 1. szám, 25-35. old.

Ward, T.J. (1994): *Cash Flow Information and the Prediction of Financially Distressed Mining, Oil and Gas Firms: A Comparative Study*. Journal of Applied Business Research, vol. 10, no. 3, pp. 78-86.

Wikipedia: *Receiver Operating Characteristic*.

http://en.wikipedia.org/wiki/Receiver_operating_characteristic, 2007.05.12-én.

Wilcox, J. W. (1971). *A simple theory of financial ratios as predictors of failure*. Journal of Accounting Research, Autumn 1971, pp. 389-395.

Wilcox, J. W. (1973): *A prediction of business failure using accounting data*. Journal of Accounting Research, Spring 1973, pp. 163-179.

Wilson, R.L. – Sharda, R. (1994): *Bankruptcy Prediction Using Neural Networks*. Decision Support Systems, vol. 11, no. 5, pp. 545-557.

Winakor, A. – Smith, R. (1935): *Changes in the Financial Structure of Unsuccessful Industrial Corporations*. Bureau of Business Research, University of Illinois, Urbana, IL, Bulletin 51.

Yang, Q.G. – Temple, P. (2000): *The hazard of Chinese enterprises under restructuring*. Working Paper, Heriot-Watt University, Edinburgh, August 2000. pp. 1-37.

Yang, Z. (2001): *A new method for company failure prediction using probabilistic neural networks*. Department of Computer Science, Exeter University, Exeter.

Zapranis, A. - Ginoglou, D. (2000): *Forecasting Corporate Failure with Neural Network Approach: The Greek Case*. Journal of Financial Management and Analysis, vol. 13, no. 2, pp. 11-21.

Zavgren, C. (1985): *Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis*. Journal of Business Finance and Accounting, vol.12, no. 1, pp. 19-45.

Zhang, G. – Hu, M. – Patuwo, B. (1999): *Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis*. European Journal of Operational Research, Vol. 116. pp. 16–32.

Zmijewski, M.E. (1984): *Methodological Issues Related to the Estimation of Financial Distress Prediction Models*. Journal of Accounting Research, 1984, no. 22, pp. 59-82.

10. Mellékletek

1. melléklet: A csődmódel-kutatások szerzőinek és módszereinek összefoglaló táblázata

Év	Szerző	Mutatóelemzés	Diszkriminancia analízis	Probit	Logit	Döntési fa	Neurális háló
1931	Ramser, Foster	X					
1932	Fitzpatrick	X					
1935	Winakor, Smith	X					
1942	Merwin	X					
1966	Beaver		X				
1966	Mears	X					
1966	Horrigan	X					
1966	Neter	X					
1968	Altman		X				
1971	Wilcox	X					
1972	Deakin		X				
1972	Lane		X				
1973	Bilderbeek		X				
1972	Edmister		X				
1973	Wilcox	X					
1974	Awh, Waters		X				
1974	Blum		X				
1975	Sinkey		X				
1975	Libby	X					
1975	Elam		X				
1976	Altman, Lorris		X				
1977	Altman, Haldeman, Narayan		X				
1977	Taffler, Tisshaw		X				
1978	van Frederikslust		X				
1978	Ketz		X				
1979	Bilderbeek		X				
1979	Norton, Smith		X				
1980	Dambolena, Khoury		X				
1980	Ohlson				X		
1980	Altman, Levallee		X				
1980	Sharma, Mahajan		X				
1981	Castagna, Matolcsy		X				
1982	Ooghe, Verbaere		X				
1982	Collins, Green		X		X		
1983	Taffler		X				

1983	El Hennawy, Morris		X				
1983	Hamer		X		X		
1984	Altman, Izan		X				
1984	Zmijewski			X			
1984	Micha		X				
1984	Lincoln		X				
1984	Takahashi, Kurokawa		X				
1985	Zavgren				X		
1985	Barth, Brumbaugh, Sauerhaft, Wang				X		
1985	Gentry, Newbold, Whitford		X		X		
1985	Casey, Bartczak		X		X		
1985	Frydman, Altman, Kao		X			X	
1985	Levitan, Knoblett		X				
1986	Lo		X		X		
1987	Keasey, Watson		X		X		
1987	Pantalone, Platt		X		X		
1987	Betts, Belhoul		X				
1987	Gombola, Haskins, Ketz, Williams		X				
1987	Peel, Peel		X		X		
1987	Lau		X		X		
1988	Rudolph, Hamdan				X		
1988	Gloubos, Grammatikos		X		X		
1988	Aziz, Emanuel, Lawson		X		X		
1988	Dambolena, Shulman				X		
1989	Barniv, Raveh		X	X	X		
1989	Hopwood, McKeown, Mutchler				X		
1990	Odom, Sharda		X				X
1990	H.D. Platt, M. Platt				X		
1990	Keasey, McGuinness				X		
1990	Gilbert, Menon, Schwartz				X		
1990	Skogsvik			X			
1991	Cadden		X				X
1991	Coats, Fant		X				X
1991	Ooghe, Joos, De Vos				X		
1991	Flagg, Giroux				X		
1992	Tam, Kiang		X		X		X
1992	Salchenberger, Cinar, Lash				X		X
1992	Bahnson, Bartley				X		
1993	Coats, Fant		X				X
1993	Fletcher, Goss				X		X
1993	Udo				X		X
1993	Chung, Tam						X
1993	Virág, Hajdu		X		X		
1994	Kerling, Poddig		X				X
1994	Altman, Marco, Varetto		X				X
1994	Berg, Hexeberg				X		
1994	Johnsen, Melicher				X		

1994	Platt, Platt, Pedersen				X		
1994	Sheppard, Fraser				X		
1994	Ward				X		
1994	Wilson, Sharda		X				X
1995	Lussier		X		X		
1995	Alici		X		X		X
1995	Boritz, Kennedy, Albuquerque		X	X	X		X
1995	Lacher, Coats, Sharma, Fant		X				X
1996	Back, Laitinen, Sere, Wesel		X		X		X
1996	Leshno, Spector						X
1996	Virág, Hajdu		X				
1996	Hill, Perry				X		
1997	Bardos, Zhu		X		X		X
1997	Pompe, Feelders		X			X	X
1997	Olmeda, Fernandez		X		X	X	X
1998	Kivilioto		X				X
1998	Piramuthu, Raghavan, Shaw						X
1998	McGurr, Devaney		X				
1998	Richardson, Kane, Lobinger				X		
1999	Laitinen, Kankaanpaa		X		X	X	X
1999	Zhang, Hu, Patuwo				X		X
1999	Lennox		X	X	X		
1999	Sjovoll			X			
1999	Tan			X			X
2000	Mckee, Greenstein		X			X	X
2000	Pompe, Bilderbeek		X				X
2000	Yang, Temple				X		X
2000	Bongini, Ferri, Hahm			X			
2000	Gritta, Wang, Davalos, Chow		X				X
2000	Zapranis, Ginoglou		X				X
2001	Hámori				X		
2001	Neophytou, Mar-Molinero				X		X
2001	Atiya						X
2001	Bernhardsen			X			
2001	Hunter, Isachenkova				X		
2001	Yang		X		X		X
2001	Lin, McClean		X		X		X
2002	Ginoglou, Agorastos, Hatzigagios		X	X	X		
2003	Darayseh, Waples, Tsoukalas				X		
2004	Neophytou, Mar-Molinero				X		
2004	Charitou, Neophytou, Charalambous				X		X
2005	Virág, Kristóf		X		X		X
2006	Virág, Kristóf		X		X	X	X
2008	Andersen				X		

2. melléklet: Az egyes pénzügyi mutatók főbb leíró statisztikai fizetési késedelembe esett és kontroll vállalatokra

Mutató	Átlag		Medián		Szórás
	Default	Kontroll	Default	Kontroll	
Likviditási mutatók					
m1	1,20	1,69	1,01	1,43	0,79
m2	0,78	1,19	0,66	1,06	0,66
m3	0,20	0,41	0,07	0,24	0,37
m4	0,16	0,21	0,11	0,14	0,24
m5	91,93	88,97	78,84	72,30	60,91
Tőkeszerkezeti mutatók					
m6	0,29	0,45	0,25	0,45	0,21
m7	0,68	0,47	0,72	0,44	0,23
m8	3,93	1,82	2,88	0,98	2,82
m9	0,29	0,13	0,27	0,04	0,23
m10	2,64	2,27	1,73	1,21	2,23
m11	3,91	1,85	2,81	0,99	2,86
Adósságszolgálati mutatók					
m12	5,18	8,62	1,78	3,38	10,11
m13	4,91	8,34	1,71	3,38	9,66
m14	9,17	16,00	3,87	8,31	15,43
m15	8,85	15,42	3,73	8,17	14,70
m16	8,41	16,14	2,92	7,92	15,81
m17	0,17	0,29	0,11	0,22	0,23

Mutató	Átlag		Medián		Szórás
	Default	Kontroll	Default	Kontroll	
Működőtőke mutatók					
m18	0,03	0,15	0,01	0,14	0,20
m19	1,22	5,62	1,16	4,88	19,04
m21	3,09	1,55	2,17	0,72	2,45
Eszközforgási mutatók					
m22	309,01	299,14	251,35	204,38	212,01
m23	56,22	44,14	32,74	27,02	55,84
m24	45,71	34,21	28,16	22,06	43,99
m25	67,54	56,14	59,39	42,81	45,16
m26	41,40	32,66	34,44	23,64	31,90
m27	67,96	36,98	62,53	27,33	41,98
Termelékenységi mutatók					
m28	21 856,79	20 085,67	15 281,85	12 254,75	17 564,60
m29	533,00	623,02	225,62	223,99	869,18
m30	950,68	813,70	583,52	294,76	1 166,28
m31	1 528,04	1 276,83	1 085,54	642,18	1 461,00

Mutató	Átlag		Medián		Szórás
	Default	Kontroll	Default	Kontroll	
Jövedelmezőségi mutatók					
m32	0,14	0,14	0,10	0,09	0,19
m33	0,02	0,03	0,01	0,02	0,04
m34	0,04	0,05	0,02	0,03	0,06
m35	0,11	0,19	0,04	0,12	0,56
m36	0,14	0,14	0,10	0,09	0,20
m37	0,16	0,16	0,11	0,10	0,21
m38	0,02	0,03	0,01	0,02	0,04
m39	0,04	0,05	0,02	0,03	0,06
m40	0,04	0,04	0,03	0,03	0,06
m41	0,04	0,03	0,03	0,03	0,06
m42	0,27	0,18	0,24	0,12	0,27
m43	0,06	0,06	0,05	0,05	0,07
m44	0,06	0,06	0,05	0,05	0,07
m45	0,25	0,18	0,22	0,12	0,27
m46	0,11	0,11	0,10	0,09	0,09
m47	0,17	0,15	0,14	0,10	0,18
m48	0,16	0,14	0,13	0,09	0,18
m49	0,31	0,25	0,27	0,20	0,23
m50	0,00	0,00	0,00	0,00	0,00
m51	0,12	0,09	0,09	0,06	0,17
m52	0,18	0,19	0,15	0,16	0,12