

# **Csődelőrejelzés genetikus algoritmus alapján**

- vázlat -

Szebenyi Bálint

## **1. Bevezetés**

### **1.1.Célkitűzés**

A dolgozattal a célom, hogy egy olyan módszert mutassak be, és próbáljak ki, amellyel egyetemi tanulmányaim alatt épp csak érintőlegesen, egy kiselőadás formájában foglalkoztam. A választott feladat keretén belül szeretném azt a kérdést körbejárni, hogy a genetikus algoritmusok mennyire alkalmasak csődelőrejelzésre, képesek-e hasonló, illetőleg jobb eredményt hozni, mint más tanult, és klasszikusnak számító módszerek.

(Alapvetően ez egy meglehetősen komplex kérdés, és mind a csődelőrejelzés, mind a genetikus algoritmusok mélyreható ismerete szükséges ahhoz, hogy a közismert módszerekkel létrehozott modelleknél jobb teljesítményt érjenek el az algoritmus által előállított modellek. Én arra helyezem a hangsúlyt, hogy alkossak egy ilyen modellt, amely képes elfogadható teljesítményt nyújtani, és inkább a módszertant mutatnám be, semmint hogy a teljesítmény kerüljön a középpontba. Mivel a nemzetközi szakirodalomban voltak már próbálkozások ilyen téren, meglehetősen jó eredménnyel, ezért remélhető, hogy a módszer használható eredményt fog felmutatni.)

### **1.2.A témaválasztás megindoklása**

A témával először egy vállalati pénzügy szakirányos tárgy során kerültem kapcsolatba, és itt ragadta meg a figyelmemet a problémakör.

A csődelőrejelzést azért tartom érdekesnek, mert bár főként a múlt adatait használja fel, mégis képes arra, hogy a jövőről valamit eláruljon. Mindez pedig nem kis gazdasági jelentőségű, mert a csődnek költségei vannak. Ha az előrejelzés és a bekövetkezés költségeit szembeállítjuk, akkor intuitívan jól látszik, érdemes megpróbálni előretekinteni a jövőbe.

Természetesen ennek akkor van igazán jelentősége, amikor hitelportfólió-építésről van szó. Egyedi

esetben egy szakértői felmérés sokkal átfogóbb megítélést tesz lehetővé. Nyilvánvalóan egy hitelintézet akkor szeretne egy vállalkozást finanszírozni, hogyha elfogadható kockázat mellett biztosítva látja megtérülését. A vállalatok méretének növekedésével tendenciaszerűen nő a nyújtott hitel összege, és így a kockázat. Ezért a nagyobb vállalatok, strukturált hitelek esetén mindenképp az egyedi elbírálás lesz jellemző, de kisvállalkozások esetén teret nyerhetnek a „dobozos termékek”, és így a modellek használata.

A módszert ugyancsak lehetséges alkalmazni olyan esetben, ahol rendelkezésre állnak adatok a sokaság reprezentációjához, és lehetséges azok alapján valamilyen mintát felfedezni. Ilyen lehet még a fent említetteken kívül egy befektetési stratégia kialakítása olyan tőzsdén, amelyen nagy számosságú vállalat található. További példa lehet az alkalmazásra egy bank lakossági portfólióépítése is.

A fent említettek közül én a magyar vállalatokon szeretnék majd modellt készíteni. Emellett elsősorban az adatok elérhetősége szól.

További személyes érv, amely a témaválasztást indokolja, hogy egyetemi tanulmányaim során szerettem a statisztikát, valamint érdeklődök a számítástechnika iránt is. Mindezek együtt abba az irányba mutattak, hogy érdemes ezt a témakört választanom.

A csődelőrejelzésen belül a genetikus algoritmusokra azért esett a választásom, mert érdekesnek találtam, hogy van egy olyan módszer, amely mesterséges intelligencia segítségével maga keresi meg a legjobb megoldást egy adott problémára, miközben szemléletes és érdekes biológiai hasonlaton alapul.

A mesterséges intelligencián belül a genetikus algoritmusok az egyik új irányt képviselik a csődmodellezés technikái között. Így egy olyan módszertannal volt lehetőségem foglalkozni, amelynek sikeres alkalmazására vannak példák, de itthon nem elterjedt.

### **1.3.A gondolatmenet ismertetése**

A dolgozatot négy nagy tartalomrészre tagoltam. Az első a bevezető, amely felvezeti a dolgozat problémáját, bemutatja miről fog szólni a tartalmi rész.

A második és a harmadik rész a hipotézis érdemi vizsgálatához szükséges elméleti áttekintést és a tényleges vizsgálatot tartalmazza. A második rész két nagyobb alrészre osztható: a történeti áttekintésre, valamint a módszertan ismertetésére. A történeti áttekintés mind a csődelőrejelzés, mind

a genetikus algoritmusok alkalmazásának előzményeit összegyűjti. A módszertani részre is jellemző lesz ez a kettősség, egy csődmódel felépítésének általános ismertetése, valamint a genetikus algoritmusok koncepciójának, működésének, fogalmainak a leírása kap itt helyet. Ebben a részben kerül majd ismertetésre a választott technológia is.

A harmadik részbe a tényleges futtatás, és annak eredményei fognak kerülni. Ebbe beletartozik az adatok gyűjtésétől, a módel elkészítésén át, a teljesítmény értelmezéséig a módelépítés teljes folyamata. Azért, hogy lehetséges legyen összehasonlítani a genetikus algoritmus klasszifikációs teljesítményét egy logit módellet is készítek, mely a referencia szerepét fogja betölteni. Itt kerül sor a módszer hatékonyságának és eredményességének értékelésére is.

A negyedik rész tartalmazza a dolgozat összefoglalását. Ehhez kapcsolódóan itt kerülnek levonásra a következtetések, valamint a felmerülő további kutatási lehetőségeket is ide gyűjtöm majd össze.

## **2. Elméleti rész**

(A csődelőrejelzés elméleti áttekintésében nagy segítség, hogy erre a Vállalati Válságkezelés és Csődelőrejelzés tárgy keretében sor került. Azonban az újabb technológiák közül, csak a neurális hálókkal foglalkoztunk részletesebben, a support vector machines (svm)-mel, és genetikus algoritmusokkal csak említés szintjén.)

### **2.1.Fogalmi áttekintés a csődelőrejelzésben**

Különbség a felszámolás, végelszámolás, csőd fogalma között.

### **2.2.Történeti áttekintés**

#### **2.2.a) Klasszikusok**

Valamennyi csődmódellet ismertető tudományos cikk Beaverrel kezdi a történeti áttekintést. (Beaver, 1966) Ő volt az első, aki tudományos módszerrel nézte meg, hogy jók-e a pénzügyi mutatók csődelőrejelzésre. Ehhez páros mintavételt használt, 33 jó, és 33 rossz vállalat 30 pénzügyi mutatóját elemezte t-próbával, hogy különböznek-e a hovatartozás szerint. Az elemzés arra a következtetésre jutott, hogy volt szignifikáns eltérés. Ennek ellenére nem egyértelmű az eredmények használata, mivel volt több szignifikáns mutató is, amely egymásnak ellentmondó besorolást sugallt, de a szerző nem tért ki rá, hogy melyiket lenne érdemes használni.

Ugyancsak klasszikus az első többváltozós módel, a Z-Score módel. (Altman, 1968) Altman

többszörös diszkriminancia-analízissel kereste a választ arra, hogy mely pénzügyi mutatókat érdemes szerepeltetni, milyen súllyal, illetve hogy hogyan kerüljenek ezek objektíven megállapításra. Öt változós modelljében minden vállalat egy értéket kap, melyet Z értéknek hív. Ez alapján az érték alapján sorolja be a vállalatokat működő és csődös kategóriákba. Modelljének hátránya, hogy

$$Z \in ]-\infty; \infty[ ,$$

így nem alkalmas a csődkockázat kardinális mérésére, nem lehetséges megmondani, hogy két eltérő érték esetén hányszorosa a bedőlés esélye egyiket a másikhoz viszonyítva.

Zavgren is tovább lépett elődei módszertanában. (Zavgren, 1985) A többszörös diszkriminancia-analízis előfeltevéseit nem hagyta figyelmen kívül, mint tette azt szándékosan Altman, helyette új módszert javasolt. Ez a módszer a logit, amely nem él annyi kikötéssel a változókra, mint a többszörös diszkriminancia-analízis. A klasszifikációhoz, vagyis a működő/csődös besoroláshoz szükséges cut value-t úgy állapítja meg, hogy a félresorolások száma legyen minimális.

## 2.2.b) Jelenkor

A fent kiemelt módszerekre irányuló műveken kívül meglehetősen sok irányban történtek kutatások a csődelőrejelzésben, olyannyira, hogy a modellépítés egyes fázisainak külön-külön irodalma alakult ki:

1. (Adatgyűjtés)
2. Adatelőkészítés: A létrehozott adatbázist használata előtt elő kell készíteni, az adatokat meg kell tisztítani a hibás értékektől. (Kristóf, Tamás, 2008)
1. Iparági hovatartozás: Az egyes mutatók már az iparági sajátosságok miatt is eltérnek, ezért célszerű az iparági átlaghoz képest relatíve vizsgálni azokat. (Platt, Platt, 1990)
2. (Adattömörítés)
3. (Kategorizálás)
4. Módszerek (2 nagy irány van)
  - Az első család a statisztikai, amelyet javarészt a klasszikusoknál történő történelmi vonal le is fed.
  - A második a mesterséges intelligenciát felhasználó módszerek családja

Itt az általam feldolgozott cikk jelenti a kiindulópontot, amely a Teheráni tőzsde vállalatait elemezte hibrid genetikus algoritmusokkal. (Divsalar, Javid, et al., 2011) A cikk is tömören áttekinti a csődelőrejelzés történeti fejleményeit, majd a genetikus algoritmusok működését ismerteti. A tanulmány hibrid genetikus algoritmusokat használ, azaz a genetikus algoritmust logit modellel ötvözi, illetve szimulált lehűlést (simulated annealing) alkalmazó variánst is bemutat.

Ezen felül az EBSCO adatbázisában találtam még tudományos írásokat ebben a témakörben, amelyeket majd fel tervezek dolgozni: (Shin, Lee, 2002; Etemadi et al., 2009; McKee, 2007; Lensberg et al., 2006; McKee, Lensberg, 2002; Divsalar, Firouzabadi, et al., 2011)

Valamennyi publikáció a genetikus algoritmusokkal foglalkozik, mint kiemelt irány, de megemlítenek más mesterséges intelligenciára építő módszereket is, amelyeket szintén szeretnék érintőlegesen bemutatni.

#### 5. Megbízhatóság-vizsgálat, validáció

Például a cut-off megválasztásának problémája is ide tartozik. A szakirodalomban gyakori, hogy objektív mércét használnak az összehasonlíthatóság kedvéért. A valóságban azonban a téves besorolások költsége is megjelenik a döntést befolyásoló tényezők között.

További kiindulási pont a kutatáshoz a Vállalkozások Pénzügyei Intézet Igazgatójának, Virág Miklósnak a munkássága, illetve a szakirányos tárgy keretében kötelező tankönyv.

### 2.2.c) Genetikus algoritmusok

A genetikus algoritmusok sem az elmúlt tíz évben születtek, ezért itt is indokolt felvázolni, hogy honnan hová jutott el a módszer. Például mikor jelentek meg hibrid futtatások (amelyek más módszerekkel kombinálják), az élet mely területein alkalmazzák (amelyeket eddig olvastam: csatornahálózat tervezés, raktártervezés, útvonaltervezés, építészet, fizika).

## **2.3.Módszertan**

### **2.3.a) Rövid megfogalmazás**

A genetikus algoritmusok a biológiai evolúciót tekintik mintának. Egy kezdeti populációból kiindulva szeretnének eljutni egy olyan egyedhez, amely a legkiemelkedőbb jellemzőkkel bír bizonyos ismérvek alapján.

Az elméleti részhez a tudnivalókat egyelőre – a cikkeken kívül – két könyv jelenti: (Álmos et al., 2012; Coley, 2010)

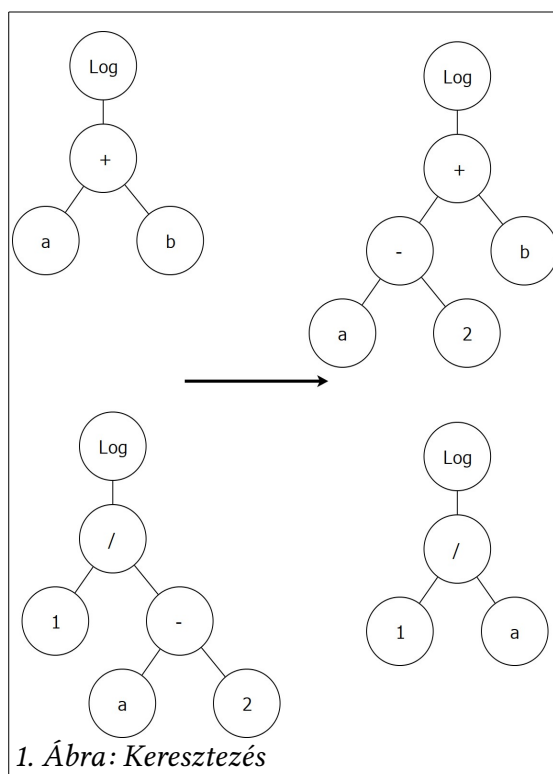
### **2.3.b) Fogalmak az ismertetéshez**

- Kromoszóma, egyed – egy lehetséges megoldás reprezentációja
- Populáció – az egyedek sokasága
- Gray code – az egyed binárisan dekódolt formája
- Fa alapú ábrázolás – az egyed ábrázolása (bináris) fában
- Gén – jellemző, amely lehet 0 vagy 1
- Allél, vagy más néven génváltozat – egy helyen felvett érték, 0 vagy 1
- Locus – pozíció a gray code-ban
- Jósági függvény (fitness függvény) – a kromoszóma minőségét ítéli meg

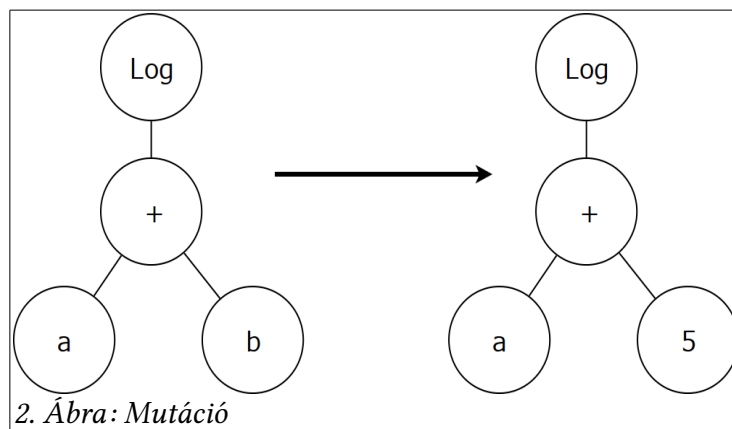
A módszertani részben használt szimbólumok listája is itt kerülne felsorolásra.

### 2.3.c) A folyamat

1. A kezdeti sokaság előállítása véletlen alapon
2. Az egyedek kiértékelése fitness függvény alapján
3. A legjobb egyed kiválasztása
4. Transzformáció:
  - a) Keresztezés:



- b) Mutáció:



c) Szaporodás:

A fenti két módszer használata az új utódok előállításához, valamint döntés arról, hogy az új populáció miképpen jöjjön létre. Tehát, hogy a régi populáció teljes egészében lecserélődjön-e.

5. Ellenőrizni, hogy megállhat-e az algoritmus, ha nem akkor újakezdés

Ez általában a generációk számának meghatározásával történik, de történhet a fitness függvény változásának üteme alapján is.

További elméleti dilemmákat is ki kell emelni, például hogy milyen paramétereket célszerű választani a futtatáshoz, melyik beállításnak milyen hatása lesz. (Mekkora populációt érdemes használni, futásidő, kiválasztás a következő populációkhoz, mekkora legyen a mutáció, kereszteződés valószínűsége.)

### 2.3.d) A választott technológia

A futtatáshoz használni kívánt valamennyi szoftver ingyenes, és nyílt forráskódú.

R: logit modell készítéséhez (SPSS is lehet, de megnéztem és az R jól használható. Van hozzá leírás az interneten, könyv a könyvtárban, ha kéne, valamint az R előnye, hogy meghívható más alkalmazásból is. Alapvetően tényleg csak a logit modellhez kell.)

Python: a futtató keretrendszer nyelve, ebben a nyelvben készített csomagokkal szeretném a genetikus részt megvalósítani. (Mivel genetikus algoritmus nincsen SPSS-ben, illetve ez mindenképp egyedi „program” lesz, nem lehet igazán összekattintgatni.)

IPython: notebook támogatás, követhetőség, szép felület. A teljes futtatást tárolja, hogy az áttekinthető legyen sorrendiségében. Tulajdonképp egy dinamikusan szerkesztett weblap, hasonló a Mathematica notebookjához, kódszínezéssel. Támogatja a többszálú feldolgozást, ami hasznos a számításigényes feladatokhoz.

Pandas: adatok beolvasása, tárolása és manipulációja.

PyEvolve: genetikus algoritmus keretrendszer. Tartalmazza a kiinduláshoz szükséges osztályokat, hogy ne kelljen mindent a nulláról megírni. Képes a teljes futtatás lefolytatására, közben a populáció elemzésére, vizualizációra.



Graphviz: ábrákat készítő motor

(+SymPy: szimbolikus kalkulációt tesz lehetővé, amely esetleg ahhoz kellhet, hogy a létrejövő függvényeknél kikerülhető legyen az azonos értelmű, de ellentétes irányú műveletek redukálása. (Ugyanannak az elemnek a kivonása és hozzáadása))

(+NumPy, SciPy – nagy teljesítményű mátrix kezelés, és matematikai műveletek, ha szükséges lenne valamihez)

### 3. Gyakorlati alkalmazás futtatással

Itt sorra venném a csődmódel-építés korábban ismertetett lépéseit, és ez alapján mutatnám be, hogy én mit csináltam.

#### 3.1.a) Adatgyűjtés

Egy iparágon belül szeretnék adatokat gyűjteni, magyar vállalatokról. Azt szeretném, ha lenne vagy kétszáz vállalat, úgy hogy lehetőség szerint egy ötöde legyen fizetéseképtelen. Ez egyelőre nehézséget okoz, mert ezekhez az adatokhoz a hozzáférés körülményes. Az e-beszámoló captchai idegőrlők, de végső megoldásként lehet, hogy ezt kell választanom.

Beszéltem azonban egy ismerőssémmel, és lehet hogy tud nekem segíteni adatokat letölteni (Opten, Céginfó). Ehhez egyelőre az kell, hogy utánanézzek milyen adatokra van pontosan szükség, remélem az elméleti irodalomban választ fogok találni.

#### 3.1.b) Adatelőkészítés

#### 3.1.c) Adattömörítés, Kategorizálás

#### 3.1.d) Logit modell

Azért logit modellt készítek referenciának, mert gyorsan és könnyen elkészíthető, valamint elterjedt, így nem kell külön részleteiben ismertetni.

#### 3.1.e) Genetikus algoritmus futtatása

Annak ismertetése, mely irányokban történtek kísérletek, illetve a legjobb részletes ismertetése.

Eddig egy egyszerű példafuttatást csináltam a PyEvolve honlapján található leírás alapján a Pitagorasz-tétel közelítésére. Ezt példaként csatoltam html-ben. A populációból mutat egyedeket a csatolt kép (itt látszik, hogy előfordul, hogy ugyanazt a lépést teszi már meg fordítva az algoritmus, azaz le is von valamit, meg hozzá is ad).

### 3.1.f) A két modell teljesítményének vizsgálata

Itt én is objektív besorolási kritériumot fogok választani, és ez alapján a két modellt összehasonlítani.

### 3.1.g) Következtetések levonása

Arról, hogy melyik modell a jobb, melyiknek milyen előnyei, hátrányai vannak. Mennyire lehet alkalmazni genetikusan algoritmusokat ebből a példából kiindulva.

## 4. Összegzés

### 4.1. Rövid összefoglaló

Az eddig írottak összefoglalása, beleértve a következtetések ismertetését is, hogy aki csak ezt olvassa, az is megismerhesse a dolgozatot.

### 4.2. Kitekintés

További lehetőségek, kérdések, kutatási irányok felvetése.

## 5. Mellékletek

Kód, nagyobb ábrák a futtatásból

## 6. Hivatkozásjegyzék

ÁLMOS, Attila – GYŐRI, Sándor – HORVÁTH, Gábor – VÁRKONYINÉ KÓCZY, Annamária (2012): *Genetikus algoritmusok*. 2. kiadás, S.l.: Typotex Elektronikus Kiadó Kft., ISBN 978-963-279-107-4,

ALTMAN, Edward I. (1968): Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. In: *The Journal of Finance*. 1968. szeptember 1, Vol. 23, no. 4, pp. 589–609, DOI 10.2307/2978933,

BEAVER, William H. (1966): Financial Ratios As Predictors of Failure. In: *Journal of Accounting Research*. 1966. január 1, Vol. 4, pp. 71–111, DOI 10.2307/2490171,

COLEY, David A (2010): *An Introduction to Genetic Algorithms for Scientists and Engineers*. 5., Singapore: World Scientific Publishing Co. Pte. Ltd., ISBN 981-02-3602-6,

DIVSALAR, Mehdi — FIROUZABADI, Ali Khatami — SADEGHI, Meisam — BEHROOZ, Amir Hossein — ALAVI, Amir Hossein (2011): Towards the prediction of business failure via computational intelligence techniques. In: *Expert Systems*. 2011. július, Vol. 28, no. 3, pp. 209–226, DOI 10.1111/j.1468-0394.2011.00580.x,

DIVSALAR, Mehdi — JAVID, Mohamad Rezi — GANDOMI, Amir Hosein — SOOFI, Jahaniar Bamdad — MAHMOOD, Majid Vesali (2011): Hybrid Genetic Programming-Based Search Algorithms for Enterprise Bankruptcy Prediction. In: *Applied Artificial Intelligence*. 2011. szeptember, Vol. 25, no. 8, pp. 669–692, DOI 10.1080/08839514.2011.595975,

ETEMADI, Hossein — ANVARY ROSTAMY, Ali Asghar — DEHKORDI, Hassan Farajzadeh (2009): A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. In: *Expert Systems with Applications*. 2009. március, Vol. 36, no. 2, Part 2, pp. 3199–3207, DOI 10.1016/j.eswa.2008.01.012,

KRISTÓF, TAMÁS (2008): A csődelőrejelzés és a nem fizetési valószínűség számításának módszertani kérdéseiről. In: *Közgazdasági Szemle*. 2008., Vol. 55. évfolyam, no. 5. szám, pp. 441–461,

LENSBERG, Terje — EILIFSEN, Aasmund — MCKEE, Thomas E. (2006): Bankruptcy theory development and classification via genetic programming. In: *European Journal of Operational Research*. 2006. március 1, Vol. 169, no. 2, pp. 677–697, DOI 10.1016/j.ejor.2004.06.013,

MCKEE, Thomas E. (2007): Altman's 1968 Bankruptcy Prediction Model Revisited via Genetic Programming: New Wine from an Old Bottle or a Better Fermentation Process? In: *Journal of Emerging Technologies in Accounting*. 2007. január, Vol. 4, pp. 87–101,

MCKEE, Thomas E. — LENSBERG, Terje (2002): Genetic programming and rough sets: A hybrid approach to bankruptcy classification. In: *European Journal of Operational Research*. 2002. április 16, Vol. 138, no. 2, pp. 436–451, DOI 10.1016/S0377-2217(01)00130-8,

PLATT, Harlan D. — PLATT, Marjorie B. (1990): Development of a Class of Stable Predictive

Variables: The Case of Bankruptcy Prediction. In: *Journal of Business Finance & Accounting*. 1990. Spring, Vol. 17, no. 1, pp. 31–51,

SHIN, Kyung-Shik — LEE, Yong-Joo (2002): A genetic algorithm application in bankruptcy prediction modeling. In: *Expert Systems with Applications*. 2002. október 1, Vol. 23, no. 3, pp. 321–328, DOI 10.1016/S0957-4174(02)00051-9,

ZAVGREN, Christine V. (1985): Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis. In: *Journal of Business Finance & Accounting*. 1985. Spring, Vol. 12, no. 1, pp. 19–45,