**Startup felvásárlások multikulturális hátterének elemzése,**

**avagy mesterséges intelligencia alapú ellenőrzőszámítás diszkriminancia-elemzéshez**

Barta Gergő, Dr. Pitlik László

1. **Absztrakt**

A kísérlet célja kiválasztott kategóriájú, azaz különböző működési profillal rendelkező startup szervezetek multikulturális hátterének modellezése, melyben a meglévő adatvagyonra vonatkozóan 3 felhasznált algoritmus alapján klasszifikálásra alkalmas mintákat kerestünk a startup cégeket, mint objektumokat leíró attribútumokban. A felhasznált algoritmusok által azt szerettük volna lát(tat)ni, hogy egy-egy módszer mennyiben képes egy már felvásárolt szervezet és felvásárlója közötti pl. ország-azonosságot felismerni, ahol az ország-azonosság bináris változóként lett definiálva. A kísérletet elvégeztük diszkriminancia-elemzéssel, neurális hálóval és hasonlóságelemzéssel.

**Kulcsszavak**: diszkriminancia-elemzés, neurális háló, hasonlóságelemzés, startup, klasszifikáció

1. **Bevezetés – A startupokról**

A startup fogalom teljesen átértékelődött (a válság óta) az elmúlt évtizedben. Hallhatjuk, amint emberek az utcán hipszter amerikai alkalmazásokról, fiatalos és új vállalkozásokról, vagy tech cégekről beszélnek, és a startup fogalmát az előbb említett szavakkal definiálják. Valahol mindenkinek igaza van, valahol mindenki téved. Neil Blumenthal, a Warby Parker alapítója és vezérigazgatója szerint a "*startup nem más, mint egy olyan szervezet, amely olyan problémák megoldásán dolgozik, ahol a megoldás nem nyilvánvaló, és a siker nem garantált*." (Robehmed, 2013) Adora Cheung a Homejoy tulajdonosa pedig úgy gondolja, hogy a "*startup az a szervezet, ahol az emberek folyamatosan fontolóra veszik azt a döntést, hogy megéri-e feladni a stabil állásukat, egy olyan cégért, ahol a siker nem biztosított, de lehetséges a hatalmas növekedés és a gyors hatás elérése.*" (Robehmed, 2013) Harbert szerint "*a startup tulajdonképp két srác egy garázsban, és nagyon dolgoznak valamin.*" (Harbert, 2014) A Magyar Startup Közösség a következő megfogalmazást használja: "*tipikusan egy induló mikro-, vagy kisvállalkozás, amely innovatív ötletet valósít meg, lépésről lépésre halad a termékfejlesztésben és piacra vezetésben, tőkebefektetésre vár és nagyra tör.*" (Magyar Startup Közösség, 2016) A startup cég fogalomra nincs egységes meghatározás, csupán körülírások, és azt is érdekes lehet leszögezni, hogy van-e különbség egy KKV és egy startup között, és ha van, mi? A definíciók ugyanis csak akkor érdemiek, ha az egymástól megkülönböztetendő rokon jelenségek világos és egyidejű lehatárolására képesek (vö. szakértői rendszerek – pl. gombahatátorzó – vö. http://miau.gau.hu/myx-free/ego/).

Az eddigi fogalmak és magyarázatok alapján viszont tehetünk megállapításokat, amelyek közelebb vezethetnek a startup jelenség megértéséhez. Elsősorban azt látjuk, hogy amikor startupról beszélünk, nem egy nagy entitásról esik szó. Másodsorban pedig a "kockázatos" szó juthat eszünkbe a definíciók hallatán. Ha tovább böngésszük az internetes oldalak sorait, találkozhatunk „millióféle” startup céggel, melyek termékpalettája igen színes, az viszont elmondható[[1]](#footnote-1), hogy az esetek többségében egy szellemi termékről van szó, amelynek központjában a szoftver áll. Fontos megjegyezni, hogy nem kizárólag szoftverfejlesztésről, mint szolgáltatásról van szó, hanem a termék maga egy szoftver, melyben általában regisztrálni lehet, majd pénzért vagy ingyen, valamilyen szolgáltatáshoz juthat hozzá annak felhasználója. Összegezve tehát a cikk szakértői rendszer alapon egy önálló cikkben visszaellenőrizendő definíciója a következő: *a startup egy olyan innovatív, vagy innovatívnak gondolt, főleg szellemi terméket fejlesztő társaság, mely a jövő kellően részletes ismerete hiányában kockázatot vállal a tulajdonosok és a befektetők kárára, gyors és nagy profit elérése érdekében.*

A startup jelenségkör triviálisan interdiszciplináris, határokon és kultúrákon átívelő, így tehát ennek apropóján a multikulturalitás, mint olyan szinte minden karakterisztikus aspektusból (nyelv, nemzet, vallás, stb.) vizsgálható. A startup-ok kapcsán rendelkezésre álló adatvagyonok ezen aspektusokat nem kezelik közvetlenül, így a startup-ok földrajzi kötődésének finomhangolt továbbgondolása új perspektívákat nyit a multikulturalitás fogalmának értelmezéséhez. A cikk keretében választ kapunk olyan jellegű kérdésekre, vajon van-e, s ha van milyen irányú és mértékű eltérése a startup-ok gazdasági adatai és ezek kulturális kötődései között?

1. **Adatok forrása és feldolgozása**

A startup cégekre vonatkozó adatok a Crunchbase[[2]](#footnote-2) oldaláról kerültek letöltésre. A Crunch Base a Tech Crunch amerikai startup hírportál online startup adatbázisa. 2007-ben került kidolgozásra, mely kezdetben a startup cégek nyomon követését és adminisztrációját tűzte ki célul. Mára a platform egy kollaborációs térré nőtte ki magát, ahol cégek, befektetők és érdeklődők bányászhatnak a startup adatok között. A startup adatbázis frissen tartása érdekében a szervezet marketingstratégiája egy olyan portál üzemeltetése, mely által a startup cégek maguk regisztrálnak és frissítik profiljukat, hogy szem előtt legyenek a bámészkodó befektetők előtt, tehát egy a közösségi háló mintájára kialakult online felületről van szó. Az oldalon a cikk írásának időpontjában[[3]](#footnote-3) csaknem 400000 regisztrált cég volt megtalálható, tehát az egyik legnagyobb adatbázis[[4]](#footnote-4), mely összesűríti az elemzéshez értékes adatokat, ez miatt is a Crunch Base adatbázisa a forrása kutatómunkának. Azonban az adatok hozzáférése regisztrációhoz kötött, és részletes leírást és adatokat a regisztrált cégekről kizárólag kutatói hozzáféréssel lehet ingyenesen lekérdezni.

A jelenleg feldolgozott adatok 2017. március 15-i időpontot tükrözik. Az adathalmazban, többek között megtalálható:

* a startup szervezetek alapításának helye (ország, város),
* időpontja (év, hónap, nap),
* státusza (még működő, felvásárolt, bezárt, tőzsdén jegyzett),
* kategóriája, azaz a startup cégprofil (pl. cybersecurity, e-commerce, 3D printing stb.),
* dolgozók száma (fő),
* finanszírozási adatok (pl. finanszírozási körök, teljes finanszírozás stb.),
* ha felvásárolt, akkor pedig a felvásárló neve,
* annak székhelye (ország, város), és
* a felvásárlás összege (USD[[5]](#footnote-5)).

A pénzügyi adatok értelmezésénél azonban fontos felhívni a figyelmet, hogy az adatbázis nem kalkulál az éves infláció mértékével, egy startup cég finanszírozása/felvásárlása esetén tehát a statisztikák nem a jelenlegi nettó jelenértéken lettek elkészítve.

Az adatok feldolgozásához a teljes sokaság heterogenitásához képest lényegesen homogénebb csoportokra osztottuk a cégeket kategóriájuk szerint, majd három kategóriát véletlenszerűen kiragadva a szoftver-kötődésűek halmazából („Enterprise Software”, „Cloud Services” és „Fintech”) végeztük el a későbbiekben ismertetett elemzéseket.

* Enterprise Software: olyan szoftvertermékeket szállító cégek, melyek ügyfelei vállalatok (B2B modell), és termékük a vállalati információfeldolgozást hivatott támogatni.
* Cloud Services: olyan termékeket forgalmazó/előállító cégek, melyek felhőszolgáltatást és megoldást nyújtanak partnereik számára.
* Fintech: olyan termékeket forgalmazó/előállító cégek, melyek pénzügyi megoldásokat nyújtanak az információtechnológia felhasználásával (pl. fizetési platformok).

A cégek kategóriákra való szűrése után a meglévő adatvagyonból saját mutatókat is készítettünk a felvásárolt szervezetek esetén, melyek „-1” és „1” értékeket vehetnek fel, melyet „**multikulturalitási**” mutatócsoportnak neveztünk el. Ennek elemei az alábbiak:

* Belföld/külföld: A felvásárló és felvásárolt cég bejegyzett székhelye a működő ország tekintetében (azonos ország esetén 1, különböző ország esetén -1)
* Kontinens: A felvásárló és felvásárolt cég bejegyzett székhelye a működő kontinens tekintetében (ayonos kontinens esetén 1, különböző kontinens esetén -1)[[6]](#footnote-6)
* Nyelv: A felvásárló és felvásárolt cég bejegyzett székhelyének anyanyelve (azonos nyelv esetén 1, különböző nyelv esetén -1)[[7]](#footnote-7)
* Pénznem: A felvásárló és felvásárolt cég bejegyzett székhelyének hivatalban lévő pénzneme (azonos pénznem esetén 1, különböző pénznem esetén -1)[[8]](#footnote-8)
* Vallás: A felvásárló és felvásárolt cég bejegyzett székhelyének elsődlegesen használt vallása (azonos vallás esetén 1, különböző vallás esetén -1)[[9]](#footnote-9)

Ezen mutatók felhasználásával a cikk hivatott bemutatni az multikulturalitásra vonatkozó összefüggéseket, illetve azok hiányát a felvásárlási tranzakciókra vonatkozó adatok tükrében.

Az adatvagyon minőségének stabilizálása érdekében először meg kellett vizsgálni:

* Az adathalmazban az alapítástól és a legelső finanszírozástól eltelt napok számát. Elvárás, hogy a két dátum között eltelt napok száma legyen pozitív. Az adathalmaz szerencsére csak 29 db olyan rekordot tartalmaz, ahol ez az oszlop értéke negatív, így ezeket a rekordokat kényszerűen ki kellett zárni az elemzésből.
* Ugyanezt a logikát követve kiszűrésre kerültek a negatív értékeket hordozó objektumok az alapítástól és felvásárlástól számítva is (mindösszesen 1 db rekord).
* Az első finanszírozástól és felvásárlástól, illetve az utolsó finanszírozástól és felvásárlástól eltelt napok számát tartalmazó cégek rekordjai (összesen 36 db rekord).

Az előzőekben leírt adattisztítást követően az adathalmaz **összesen 767 startup cég** adatait tartalmazza (tisztítás előtti 833 db céghez képest). A három említett csoport megoszlása ezen belül az alábbi:

**1****. ábra: Végleges adathalmaz, a három kiválasztott startup kategória megoszlása**

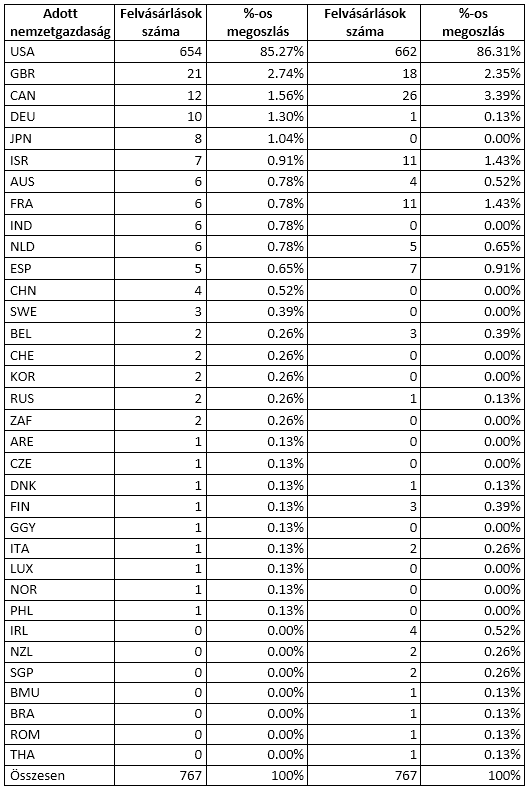
**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját számítások

|  |  |  |
| --- | --- | --- |
| **Startup kategóriák** | **db** | **%** |
| Enterprise Software | 352 | 45.9% |
| Cloud Services | 274 | 35.7% |
| Fintech | 141 | 18.4% |
| **Összesen** | **767** | **100%** |

1. **Az adathalmaz leíró statisztikája**

A három kategóriára vonatkozó konszolidált adatok a felvásárló és felvásárolt cégek bejegyzett székhelyének tekintetében az egyes nemzetgazdaságok közötti megoszlását a 2. és 3. ábra szemlélteti.

Az Amerikai Egyesült Államok a legnagyobb felvásárló nemzet, a szűkített adatbázis esetében 85%-ban az USA vásárolt fel cégeket, a felvásárolt cégek esetében, pedig ez a szám 1%-kal több, tehát 86%-a a felvásárolt cégeknek amerikai, ami arra enged következtetni, hogy az USA-n belül történt a legtöbb felvásárlásra irányuló tranzakció, amely kevésbé meglepő, mivel a nemzetgazdaság tekinthető a startup cégek őshazájának. Az Egyesült Államokon belül is a tarnzakciók döntő hányada San Fransisco városhoz köthető. San Francisco-öböl déli részén található a Szilícium-völgy, amelyet mára az információtechnológia fellegvárának neveznek, így nem is kérdéses, miért ebben a városban található meg a legtöbb startup cég. Olyan híres vállalatok központjai és leányvállalatai találhatóak meg itt, mint az Oracle, Apple, Facebook, eBay, Google és a magyar alapítású Prezi, ami jelentheti egy kezdő vállalkozás számára, hogy a világvezető „tech” cégei közül toborozhat új alkalmazottakat, tehát az egyik legfontosabb szempont, a tudástőke gyors integrálása a termékbe, amely kedvezővé teheti a cégalapítást régióban. Ugyancsak ebben az övezetben található meg a legtöbb befektető és startup felkészítő program, ami a fiatal cégeket és vállalatvezetőket motiválhat a költözésre.[[10]](#footnote-10) Amerikát leszámítva mindkét listán előkelő hellyel szerepel még Nagy-Britannia, Kanada és Izrael is.



**2. ábra: Felvásárlások országonként megbontva**

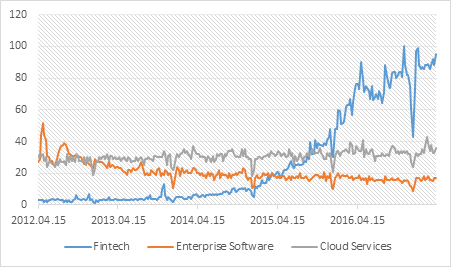
**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját számítások

A 3. ábra szemlélteti a három kategóriára külön bontva átlagosan a startup cégek teljes finanszírozását a felvásárlásig (USA-val és a felvásárolt USA cégek nélkül), azaz, hogy ezen szervezetek átlagosan mennyi külső forrásra tettek szert, amíg meg nem történt a nagyobb hal általi bekebelezés. Átlagosan az Enterprise Software kategóriába tartozó cégek gyűjtötték a legtöbbet, mely értéke 23 millió dollár. Ez a felállás valamelyest meglepő lehet, ha a kulcsszavakra rákeresünk a Google Trends elemzései között, mivel az érdeklődés azt mutatja (4. ábra), hogy az előző években a „Fintech” kulcsszó ugrásszerűen megnövekedett, míg a „Cloud Services” stagnál, az „Enterprise Software”, pedig csökkenő tendenciát mutat. Erre egy lehetséges magyarázatot mutat az 5. ábra, ahol kitűnik, hogy a Fintech cégek esetében a legkevesebb az átlagosan eltelt napok száma az alapítás és felvásárlás között, tehát, ezen cégek esetében valószínűsíthető, hogy a gyorsabb felvásárlás miatt kevesebb finanszírozási kör valósult meg. A gyorsabb felvásárlás, pedig a magas érdekeltségnek tudható be. A másik magyarázat a tudománymarketing zavaros hatásának tettenérése: a cégek élete, napi folyamatai nem változnak olyan mértékben vélelmezhetően, ahogy ezt egyes divatkifejezések felfutása sejtetni véli. Vagyis a hangzatos kifejezések mögött továbbra is a folyamatok automatizálásának mindenkor szükségszerű jellegére való rámutatás áll egyre újabb és újabb kifejezésekkel helyettesítve az „adatfeldolgozást”, mint talán az első és azóta is teljesen értelmezhetőnek megmaradt kifejezést. Ha az USA-t kizárjuk az elemzésből (ennek domináns, a többi ország hatását elnyomó volta miatt), akkor meglepő módon a Cloud Services esetén az átlagos finanszírozás több, mint kétszeresére nő, amely a holland alapítású O3b Networks Inc. –nek köszönhető, mely összesen a felvásárlása előtt csaknem 2 milliárd dollár értékben talált beruházókra. A szervezet kiszűrésével az átlagos finanszírozás a Cloud Services kategóriában 12 millióra visszaesik.

|  |  |  |
| --- | --- | --- |
| **Startup kategória** | **Átlagos finanszírozás (1000 USD) - Teljes adatvagyon** | **Átlagos finanszírozás (1000 USD) - USA-ban alapított cégek kizárásával** |
| Enterprise Software | 20645 | 19207 |
| Cloud Services | 23105 | 54570  (12259[[11]](#footnote-11)) |
| Fintech | 16753 | 19942 |

**3. ábra: Átlagos finanszírozás kategóriánként (millió USD)**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját számítások



**4. ábra: Érdeklődési trend kategóriánként**

**Forrás:** https://trends.google.hu/trends/explore?date=all&q=Enterprise%20Software,Fintech,Cloud%20Services

Az 5. ábrán látható táblázatos formában kategóriánként elkülönítve az átlagos finanszírozási körök száma, az átlagos egy főre jutó finanszírozás mértéke, illetve az átlagosan eltelt napok száma többféle a cég életében bekövetkező mérföldkőtől számítva. Ahhoz, hogy az egyes csoportok leíró adatai alapján a csoportok különbözőségének szignifikanciáját teszteljük, a varianciaelemzésnél is alkalmazott F próbafüggvényt alkalmaztuk (nullhipotézis tesztelése, azaz a startup kategóriaátlagok szignifikánsan különböznek-e egymástól). Az F próbát torzíthatja a minta csúcsossága, mely a 6. ábra alapján azt sugallja, hogy 1 kivétellel (ahol a csúcsosság értéke negatív) az F próba értékét kellő óvatosság mellett érdemes kezelni. A táblázat alapján azt is le lehet olvasni, hogy a minták nem normál eloszlásúak. Az F próba egy másik feltétele a szórás-homogenitás teljesülése, melyre az SPSS-ben elérhető Levene-tesztet használtuk, melynek értéke pl. az átlagos finanszírozási körök esetén 5.77, 0.003 (kevesebb 5%-nál) szignifikanciaszint mellett, tehát a szórás-homogenitás feltétele ebben az esetben sérül, ugyanakkor kihangsúlyozandó, hogy az ismertetett feltételek sérülése nem gyakorol esszenciális befolyást az F próbával történő első- és másodfajú hiba elkövetési valószínűségére.

A **finanszírozási körök száma** (db) egy startup életében azt mutatja meg, hogy az alapítástól számítva hányszor történt tőkebevonás külső beruházótól. Minél több beruházás történik egy cégbe, annál valószínűbb, hogy a tulajdonosok eredeti részesedése csökken, mivel a beruházó a tőkéjéért cserébe pl. osztalékot vár, mely a részesedésén keresztül realizálódik. A felvásárlások előtt az átlagos körök száma együttesen 2.72, azaz közel három volt együttesen. A legkevesebb finanszírozási kör mindegyik kategóriában értelemszerűen 1 volt, míg a legtöbb 16[[12]](#footnote-12). Az F próba alapján elmondható, hogy a kategóriaátlagok **szignifikánsan** különböznek egymástól (értéke 5% alatti).

Az **átlagosan egy főre jutó finanszírozás** értéke az Enterprise Software kategória esetén kiugró, kétszer akkora, mint a másik két kategória értéke. A kategória-átlagok **szignifikánsan** különböznek egymástól, az F próba szignifikanciája 1%.

Az **átlagosan eltelt napok száma** az alapítástól és felvásárlásig a három a kategóriában kb. 6 és fél év, az Enterprise Software kategória közelít a héthez, míg a másik kettő, pedig inkább csak a 6 évhez. Ennek az egyik magyarázata lehet, hogy az „Enterprise Software” az egyik legrégebben használt kulcsszó, s a kulcsszavak életciklusa vélelmezhetően átlagosan egyre rövidül. A kategória-átlagok **szignifikánsan** különböznek egymástól, az F próba szignifikanciája 1.7%.

Az is látható, hogy az **utolsó finanszírozás és felvásárlás között** átlagosan 2 évet kell várakozniuk a cégeknek, valószínűsíthető, hogy ez azaz időtartam mire az utolsó beruházás is meghozza gyümölcsét és a kidolgozott termék elég piacérett lesz, hogy egy másik vállalat termékpalettáját színesítse. A kategória-átlagok **szignifikánsan** különböznek egymástól, az F próba szignifikanciája 0.6%.

Érdemes még egy pillantást vetni az **átlagosan eltelt napok számát** illetően **az alapítástól és az első finanszírozástól**. Összesen átlagosan valamivel kevesebb, mint két és fél évet kell a szervezeteknek működnie, a terméken dolgoznia és kiépítenie a márkáját a piacon, hogy a befektetők észrevegyék őket. Ez az érték a Fintech esetében a legalacsonyabb kb. 2 év, mely a Fintech cégekben való nagyobb érdeklődést is mutatja. A kategória-átlagok **szignifikánsan** különböznek egymástól, az F próba szignifikanciája 3.4%.

Az F próba az átlagosan eltelt napok száma az első finanszírozás és a felvásárlás között mutat egyedül nem szignifikáns eltérést, azaz **a startup kategóriák ezen átlagai között nem mutatható ki szignifikáns eltérés** (értéke az 5%-ot meghaladja), ami 17%.

Fontos kiemelni, hogy **szignifikancia** kizárólag a három kategória együttes vizsgálatával mutatható ki a vizsgált változók tekintetében (kivéve 1 említett esetben), azonban **bármely két kategóriát** külön vizsgálva, létezik olyan változó, ahol a kategóriátlagok között **nem** mutatható ki szignifikáns eltérés. Ennek az ellentmondásosnak tűnő eredménynek a végig gondolása jelen cikknek nem tárgya, azonban a szerzők folytatják az erre vonatkozó kutatást.

A fejezet összefoglalásaként elmondhatjuk, hogy a startup felvásárlás jelenséget földrajzi vonatkozásban legfőképp az USA-ban folytatott tranzakcióra tudjuk értelmezni, ezért kézenfekvő az elemzések elvégzése az USA kizárásával is. Továbbá, a kiválasztott kategóriák a többváltozós elemzésekben független változóként használt attribútumukra, 1 változó kivételével, szignifikáns eltéréseket mutatnak (a kategóriaátlagok szignifkánsan különböznek egymástól), mely azt sugallja, hogy az általunk véletlen kiválasztott kategóriák (startup cégprofilok) hatással bírnak a változókra, összefüggés lelhető fel közöttük. Abban a látszólagos anomáliában, melyben azt tapasztaltuk, hogy nem minden változó között mutatható ki szignifikáns eltérés a harmadik kategória kizárásával, vita indítható a módszer teljesítőképességéről.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Startup kategóriák | Átlagos finanszírozási körök (db) | Átlagos egy főre jutó finanszírozás értéke (1000USD) | Átlagosan eltelt napok száma az alapítás és felvásárlás között | Átlagosan eltelt napok száma az első finanszírozás és a felvásárlás között | Átlagosan eltelt napok száma az utolsó finanszírozás és felvásárlás között | Átlagosan eltelt napok száma az alapítás és az első finanszírozás között | Átlagosan eltelt napok száma az alapítás és utolsó finanszírozás között |
| Enterprise Software | 2.89 | 634 | 2,549 | 1,627 | 753 | 922 | 1,796 |
| Cloud Services | 2.64 | 264 | 2,295 | 1,470 | 669 | 826 | 1,627 |
| Fintech | 2.44 | 318 | 2,232 | 1,528 | 888 | 704 | 1,344 |
| Összesen/Átlagosan | **2.72** | **444** | **2,400** | **1,553** | **748** | **848** | **1,652** |

**5. ábra: Startup cégeket leíró adatok kategóriánként**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jelmagyarázat: sötétzöld = legnagyobb érték, élénkzöld = középérték, halványzöld = legalacsonyabb érték

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Minimum | Maximum | Átlag | Szórás | Ferdeség | Csúcsosság | Levene | Szig. | F próba | Szig. |
| Átlagos finanszírozási körök (db) | 1 | 16 | 2.72 | 1.74 | 1.81 | 6.94 | 5.77 | 0.003 | 3.9 | **2.1%** |
| Átlagos egy főre jutó finanszírozás értéke (1000USD) | 0 | 29,429 | 444 | 1,610 | 12.01 | 178.53 | 12.05 | 0.000 | 4.6 | **1%** |
| Átlagosan eltelt napok száma az alapítás és felvásárlás között | 228 | 6,113 | 2,400 | 1,345 | 0.73 | -0.35 | 0.34 | 0.711 | 4.11 | **1.7%** |
| Átlagosan eltelt napok száma az első finanszírozás és a felvásárlás között | 62 | 5,447 | 1,553 | 1,049 | 1.09 | 0.78 | 0.4 | 0.669 | 1.8 | 17% |
| Átlagosan eltelt napok száma az utolsó finanszírozás és felvásárlás között | 0 | 4,967 | 748 | 670 | 2.46 | 8.58 | 10.1 | 0.000 | 5.1 | **0.6%** |
| Átlagosan eltelt napok száma az alapítás és az első finanszírozás között | 0 | 5,146 | 848 | 862 | 1.80 | 3.70 | 1.67 | 0.188 | 3.39 | **3.4%** |
| Átlagosan eltelt napok száma az alapítás és utolsó finanszírozás között | 0 | 5,539 | 1,652 | 1,193 | 0.96 | 0.32 | 2,61 | 0.074 | 7.5 | **0.1%** |

**6. ábra: Startup cégeket leíró adatok kategóriánként**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jelmagyarázat: A Szig. oszlopban leolvasható számok az F próba szignifikanciáját ábrázolják, a vastaggal szedett számok esetében a kategóriátlagok közötti eltérés szignifikáns (Szig. < 5%)

1. **Multikulturalitásra vonatkozó elemzések**

**5.1. Diszkriminancia-elemzés**

A soron következő elemzések már a szerzők által kalkulált multikulturalitási mutatókra fókuszálnak, és a szervezetek földrajzi elhelyezkedése, nyelve valamint vallása kerül az elemzés középpontjába. Mivel az ismertetett multikulturalitási mutatók kategorikus változók (értékkészlet: 1 egyezőség, -1 nem egyezőség esetén), melyek a kutatás függő változói, és a vizsgált független változók metrikus skálán mért változók, ezért legkézenfekvőbb a diszkriminancia-elemzés alkalmazása, mely során megvizsgálhatjuk, hogy az egyes kategóriákhoz való tartozás kapcsán milyen mértékben képesek a független változók „előrejelezni” a függő, multikulturalitásra vonatkozó változókat. A diszkriminancia-elemzés használatához számos feltételt meg kell vizsgálni annak bizonyítására, hogy a modell valóban betölti a hozzá fűzött reményeket. Amennyiben a diszkriminancia-elemzés feltételei sérülnek, úgy az eredményeket kellő óvatosság mellett érdemes kezelni. Elvárások:

1. **Korreláció feltárása a független változók között**: A diszkriminancia-elemzés során azzal a feltételezéssel élünk, hogy a független változók a függő változóra adnak magyarázatot, a független változók között nincs összefüggés. Amennyiben van, abban az esetben ugyanaz az információ redundánsan fordul elő többször a modellben, így torzított eredményeket kaphatunk. Az összefüggés-vizsgálatnál 0.7 értéket határoztunk meg felső határnak, tehát a 0.7 és 1 közé eső értékek esetén úgy véltük, hogy a független változók között már torzító erejű összefüggés van, ezért az elemzésből elhanyagolható. A számítás elvégzése után az figyeltük meg (a teljes adathalmaz és USA kizárása esetén is), hogy az „Eltelt napok száma az alapítás és felvásárlás között” változó erős összefüggést mutat a „Eltelt napok  száma az első finanszírozás és a felvásárlás között” és a „Eltelt napok száma az alapítás és utolsó finanszírozás között” változókkal (rendre 0.77, 0.87, 0.78, 0.84), ezért ezt a változót kizárjuk az elemzésből. A teljes (a kiértékelés szempontjából trianguláris) korrelációs mátrixot az 1. melléklet tartalmazza.
2. **Normalitásvizsgálat**: A normalitási feltétel egyrészt függ a független változók skálájától és a kiugró értékektől. Ha a független változók dichotóm változók, akkor az a normalitás sérülését jelenti. A mi esetünkben a független változók nem dichotóm változók (mivel a változók nem kizárólag két értéket vehetnek fel). A másik feltétel a kiugró értékek kiszűrése, mivel a diszkriminancia-elemzés érzékeny az adathalmazban található kiugró értékekre. A kiugró értékek szűrését a Mahalanobis-távolsággal mértük (többváltozós kiugró értékek vizsgálata), melyben az összes a modellben alkalmazott változó szerepel. Minél nagyobb a mutató értéke, annál nagyobbak a változók közötti távolságok, így annál nagyobb az esélye, hogy eltérünk a normalitástól. Sajtos – Mitev (2007) alapján 100 feletti minta esetén a Mahalanobis távolság értékének és az alkalmazott változók hányadosának kisebbnek kell lennie 4-nél, hogy a normalitás feltétele ne sérüljön. A teljes adathalmazon elvégezve a számítást 15 esetben tapasztaltunk többváltozós kiugró értékek (mely 3 amerikai céget érint), így ezen rekordokat kizártuk az elemzésből. A kizárt szervezetek listáját és a hozzájuk tartozó Mahalanobis-távolságot a 2. melléklet szolgáltatja.
3. **Variancia-homogenitás:** Ha fenn áll a variancia-homogenitás a független változók között a függő változó fényében, akkor annak eredményeképp a diszkriminancia-elemzés szignifikancia-tesztjei nem tekinthetőek megbízhatónak. A variancia-homogenitás minden egyes függő változó esetében külön lett letesztelve, így az eredmények ismertetésénél kitérünk a variancia-homogenitás szignifikanciájára.

A diszkriminancia-elemzés feltételeinek vizsgálata, majd a feltételeket sértő adatok kiszűrése után a következők állapíthatóak meg, melyet az elemzésbe bevont függő változók tükrében a következő alfejezetek ismertetnek.

5.1.1. A felvásárlás földrajzi sajátossága (ország)

Az első vizsgált függő változó a startup szervezet és az azt felvásárló cég földrajzi (ország) egyezőségét, illetve nem egyezőségét írja le. A kutatói kérdés az volt, hogy a fentebb ismertetett független változók, milyen mértékben képesek magyarázni a függő változónkat. A 7. ábra szemlélteti, hogy az egyes független változóknak, milyen hatása van a földrajzi felvásárlásra (ugyanaz az országból történik tranzakció, vagy nem). A Wilks’ Lambda értéke tud erre a kérdésre magyarázattal szolgálni. A Wilks’ Lambda értékkészlete 0 és 1 közé esik, ahol az egyhez közeli érték azt reprezentálja, hogy a független változóknak egyáltalán nincs hatása a két csoportra. Ahogy ez a táblázatból is kitűnik, ezen változóknak nincs szignifikáns hatása a csoportba tartozásra (sem a teljes adathalmazra, sem az USA kizárásával) egyetlen egy változó kivételével (Finanszírozási körök (db)), mivel az F próba szignifikanciája, ezen kívül, minden esetben meghaladja az 5%-ot. A variancia-homogenitás a Box’s M mutatóval történt kiértékelésre, ahol annak szignifikanciája a teljes adathalmaz esetén 0.67, USA kizárásával 0 (a teszt azt vizsgálja, hogy a független és függő változóból képzett mátrix, mennyire homogén), tehát a teszt nem szignifikáns a teljes adathalmazra, így a variancia-homogenitás feltétele teljesül, azonban az USA kizárásával a feltétel sérül, ezért a mutatók nem tekinthetőek megbízhatónak.

**7. ábra: Csoportátlagok szignifikanciája**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jelmagyarázat: A Szig. oszlopban leolvasható számok az F próba szignifikanciáját ábrázolják, a vastaggal szedett számok esetében a kategóriátlagok közötti eltérés szignifikáns (Szig. < 5%)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Felvásárlás (ország azonosság) | **Teljes adathalmaz** | | | **USA kizárásával** | | |
|  | Wilks' Lambda | F | Szig. | Wilks' Lambda | F | Szig. |
| Finanszírozási körök (db) | 0.993 | 4.941 | **2.7%** | 0.978 | 2.293 | 13.3% |
| Egy főre jutó finanszírozás értéke (1000 USD) | 0.999 | 0.515 | 47.3% | 0.996 | 0.391 | 53.3% |
| Eltelt napok  száma az első finanszírozás és a felvásárlás között | 1.000 | 0.306 | 58.1% | 0.993 | 0.720 | 39.8% |
| Eltelt napok száma az utolsó finanszírozás és felvásárlás között | 0.996 | 3.325 | 6.9% | 0.990 | 0.982 | 32.4% |
| Eltelt napok száma az alapítás és az első finanszírozás között | 0.999 | 0.765 | 38.2% | 0.993 | 0.663 | 41.7% |
| Eltelt napok száma az alapítás és utolsó finanszírozás között | 1.000 | 0.049 | 82.5% | 0.995 | 0.464 | 49.7% |

A diszkriminancia-elemzés eredményeképp a 8. ábra szolgáltatja az „előrejelzéseket” a csoportba tartozáshoz. Ahogy az ábrán kivehető és a csoportátlagok szignifikanciájánál is sejthető volt, az előrejelzés egy-egy csoportba való tartozásához az klasszifikációs pontossága közel 50%. A teljes adathalmaz esetén 145 szervezet felvásárlása történt egy másik országból, a független változók ismeretében 63% pontossággal (91 rekord esetén) történt pontos csoportosítás a kategóriára vonatkozóan. Az USA kizárásával még bizonytalanabb a helyzet, tehát a független változóink ismeretében az előrejelzés voltaképp közelít a fej-vagy-írás valószínűségéhez egy érmefeldobás esetén. A táblázat többi sora az előbbiekhez hasonlóan értelmezhető. Kijelenthetjük, hogy a függő változóink ismeretében csak minimálisan tudunk becslést adni, hogy egy adott szervezetet belföldi vagy külföldi cég vásárolt-e fel.

**8. ábra: Becslés csoportbatartozáshoz**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Felvásárlás (ország azonosság) | | Becslés | | Összesen |
| -1 | 1 |
| Teljes adathalmaz | | | | |
| db | -1 | 91 | 54 | 145 |
| 1 | 274 | 333 | 607 |
| % | -1 | 63% | 37% | 100% |
| 1 | 45% | 55% | 100% |
| USA kizárásával | | | | |
| db | -1 | 47 | 34 | 81 |
| 1 | 10 | 11 | 21 |
| % | -1 | 58% | 42% | 100% |
| 1 | 48% | 52% | 100% |

**9. ábra: Becslés csoportbatartozáshoz – Kontingencia mátrix (Teljes Adathalmaz)**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jó = Találatok száma / Összes objektum = (91 + 333) / 752

Rossz = Hibák száma / Összes objektum = (54 + 274) / 752

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| Teljes adathalmaz (Kontingencia Mátrix) | | |
| 56.38% | 43.62% | 100.00% |

**10. ábra: Becslés csoportbatartozáshoz – Kontingencia mátrix (USA kizárásával)**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jó = Találatok száma / Összes objektum = (47 + 11) / 102

Rossz = Hibák száma / Összes objektum = (34 + 10) / 102

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| USA kizárásával (Kontingencia Mátrix) | | |
| 56.86% | 43.14% | 100.00% |

5.1.2. A felvásárlás földrajzi sajátossága (kontinens)

A második vizsgált függő változó (hasonlóan az előzőhöz) a startup szervezet és az azt felvásárló cég földrajzi egyezőségét, illetve nem egyezőségét írja le, azonban ez esetben a kontinens tükrében. A 11. ábra szemlélteti, hogy az egyes független változóknak, milyen hatása van a földrajzi felvásárlásra (ugyanarról a kontinensről történik tranzakció, vagy sem). A Wilks’ Lambda alapján, ezen változók közül is csak az elsőnek van szignifikáns hatása a csoportokra. A Box’s M mutató értéke mindkét esetben közel 0.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Felvásárlás (kontinens azonosság) | **Teljes adathalmaz** | | | **USA kizárásával** | | |
|  | Wilks' Lambda | F | Szig. | Wilks' Lambda | F | Szig. |
| Finanszírozási körök (db) | 0.992 | 5.678 | **1.7%** | 0.984 | 1.612 | 20.7% |
| Egy főre jutó finanszírozás értéke (1000 USD) | 1.000 | 0.110 | 74% | 0.984 | 1.651 | 20.2% |
| Eltelt napok  száma az első finanszírozás és a felvásárlás között | 1.000 | 0.138 | 71% | 0.988 | 1.227 | 27.1% |
| Eltelt napok száma az utolsó finanszírozás és felvásárlás között | 0.997 | 1.987 | 15.9% | 0.996 | 0.370 | 54.4% |
| Eltelt napok száma az alapítás és az első finanszírozás között | 0.999 | 0.527 | 46.8% | 1.000 | 0.001 | 97.5% |
| Eltelt napok száma az alapítás és utolsó finanszírozás között | 1.000 | 0.025 | 87.4% | 0.984 | 1.614 | 20.7% |

**11. ábra: Csoportátlagok szignifikanciája**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jelmagyarázat: A Szig. oszlopban leolvasható számok az F próba szignifikanciáját ábrázolják, a vastaggal szedett számok esetében a kategóriaátlagok közötti eltérés szignifikáns (Szig. < 5%)

A diszkriminanciaelemzés eredményeképp a 12. ábra szolgáltatja az előrejelzéseket a csoportbatartozáshoz. Az USA kizárásával az előző elemzéshez képest valamivel javult a helyzet, a független változók a csoportbatartozás 71%-át pontosan képesek megbecsülni, kontinensazonosság esetén.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Felvásárlás (kontinens azonosság) | | Becslés | | Összesen |
| -1 | 1 |
| Teljes adathalmaz | | | | |
| db | -1 | 69 | 36 | 105 |
| 1 | 297 | 350 | 647 |
| % | -1 | 66% | 34% | 100% |
| 1 | 46% | 54% | 100% |
| USA kizárásával | | | | |
| db | -1 | 22 | 25 | 47 |
| 1 | 16 | 39 | 55 |
| % | -1 | 47% | 53% | 100% |
| 1 | 29% | 71% | 100% |

**13. ábra: Becslés csoportbatartozáshoz – Kontingencia mátrix (Teljes Adathalmaz)**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jó = Találatok száma / Összes objektum = (69 + 350) / 752

Rossz = Hibák száma / Összes objektum = (36 + 297) / 752

**12. ábra: Becslés csoportbatartozáshoz**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| Teljes adathalmaz (Kontingencia Mátrix) | | |
| 55.72% | 44.28% | 100.00% |

**14. ábra: Becslés csoportbatartozáshoz – Kontingencia mátrix (USA kizárásával)**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jó = Találatok száma / Összes objektum = (22 + 39) / 102

Rossz = Hibák száma / Összes objektum = (25 + 16) / 102

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| USA kizárásával (Kontingencia Mátrix) | | |
| 59.80% | 40.20% | 100.00% |

A pénznemre, vallásra és nyelvazonosságra vonatkozó számítások, hasonlóan értelmezhetőek, azok táblázata, mely a csoportba tartozás becsléséről szól, rendre a cikk 3., 4., és 5. mellékletében olvasható.

A diszkriminancia-elemzés elvégeztével és az eredmények értelmezésével az a konklúzió vonható le, hogy a becslés pontossága közel sem nevezhető elfogadhatónak, a tanulás után az újraklaszterezés legjobb eredményét a vallás-azonosság során értük el az USA kizárásával, ami 72.55% (5. számú melléklet). Olybá tűnik, hogy olyan módszerek alkalmazására lesz szükség a probléma megoldásához, amely képes a megadott mintához precízebben a klaszterek közötti elválasztó függvényt approximálni, ezzel az egyes klaszterekhez nagyobb pontossággal az adott objektumokat hozzárendelni.

**5.2. Neurális hálóval történő elemzés**

A mesterséges neurális hálók ötlete hosszú évtizedekre (vö. 1949) nyúlik vissza. (Raschka, 2015) Az alapötlet a modell mögött az emberi agy tanulási folyamatának lemásolása, majd ezt átültetve a gépbe, egy olyan mesterséges intelligencia létrehozása, mely az emberi biológiai folyamatot képes utánozni. A mesterséges neurális háló input mezői neuronokból tevődik össze, melyeket egy oszlopvektorban tárolhatunk. A neuronok ezután egy következő réteghez kapcsolódnak, ahol a dendritek súlyozzák a továbbított adatot, összegzik azt, majd a következő rétegben lévő neuronnak ez szolgál inputként. Itt történik egy aktiváció, mely a biológiában is előfordul. Az aktivációs függvény, mely esetünkben egy logisztikus szigmoid függvény transzformálja a neuron input adatát output adattá.

Megjegyzendő, hogy neurális hálók esetén is figyelmesen kell eljárnunk, mivel számos konfigurációs beállítás lehetséges, pl. az alábbiak:

* Tanulási ráta: azt mutatja meg, hogy a hibafüggvényen milyen mértékben történjen a hiba visszaszámítása pl. egy gradiens algoritmus alapján.
* Rétegek száma: azt mutatja meg, hogy a neurális háló hány rétegből áll.
* Neuronok szám: azt mutatja meg, hogy a belső rétegek hány neuronból álljanak.
* Tanulási körök száma: azt mutatja meg, hogy a hálónak hányszor történt a tréning adat megmutatásra.
* Aktivációs függvény: azt mutatja meg, hogy a milyen függvény került beállításra, mely az inputokat aktiválja. Leggyakoribb eset a logisztikus szigmoid függvény (főleg klasszifikációs problémákhoz), melynek értékkészlete 0-1.

Neurális háló modellel elkészítésre került az ország-azonosságra vonatkozó elemzés (kizárólag azon adatokra, melyben USA nem szerepel), és azt tapasztaltuk, hogy közel 30%-kal jobb eredményt lehet elérni, mint diszkriminancia-elemzéssel, melyet a 15. ábra szemléltet. Minden olyan felvásárlási tranzakciót, ahol nem áll fenn az ország azonosság, helyesen becsült. Az alábbi konfigurációval elkészített hálót alkalmaztuk a számításhoz:

* Tanulási ráta: 0.1
* Rejtett neuronok: 30
* Tanulási körök száma: 3
* Input-változók száma: Megegyezik a diszkriminancia-elemzésnél felhasznált változókkal, melyek egyszerre kerültek felhasználásra a következményváltozó becsléséhez

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Felvásárlás (ország azonosság) | | Becslés | | Összesen |
| -1 | 1 |
| USA kizárásával | | | | |
| db | -1 | 81 | 0 | 81 |
| 1 | 18 | 3 | 21 |
| % | -1 | 100% | 0% | 100% |
| 1 | 86% | 14% | 100% |

**15. ábra: Becslés csoportba tartozáshoz - Neurális háló modell**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| USA kizárásával (Kontingencia Mátrix) | | |
| 82.35% | 17.65% | 100.00% |

**16. ábra: Becslés csoportbatartozáshoz – Kontingencia mátrix (USA kizárásával)**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jó = Találatok száma / Összes objektum = (81 + 3) / 102

Rossz = Hibák száma / Összes objektum = (0+ 18) / 102

A neurális háló approximációja azt mutatja, hogy a vizsgált adathalmaz objektumait 82.35%-os precizitással sikerült megfelelően klaszterezni, mely a diszkriminancia-elemzés legjobb részeredményét így is közel 10%-kal megveri. Érdemes azonban kiemelni, hogy a neurális háló eredményei egyoldalúnak tekinthető, mivel látszólag azt az eredményt szolgáltatja, hogy minden objektumot először a különbözőséget feltételező csoportba helyezett, majd voltak (3 db) olyan objektumok is, melyek a másikba kerültek, mivel bizonyos eltérő mintát hordoztak magukban.

**5.3. Hasonlóságelemzés**

A hasonlóságelemzés elmélete a nyolcvanas évekre nyúlik vissza, amikor is Dobó Andor hasonlóságok hasonlósága tétele alapján elkészült az első szoftveres megoldás (Joker-program). Az online hasonlóságelemzések lehetősége 2006 óta adott (vö. MY-X FREE: <http://miau.gau.hu/miau/196/My-X%20Team_A5%20fuzet_HU_jav.pdf>)

A hasonlóságelemzés célja a klasszifikációs, termelési függvénytípusú Y=f(Xi) inputmintázatok esetén a következményváltozó értékeinek lépcsős függvénnyel történő közelítése, ahol a lépcsős függvény egyrészt már komplexebb, mint egy regresszió, hiszen az input-változók együtthatója a változó rangsorának függvénye. Másrészt a lépcsős függvény egy fajta egyszerű neurális hálóként is értelmezhető, ahol a lépcső maga a hálózat.

A hasonlóságelemzés kapcsán elvárás, hogy a találati arányt abból a természetes kiindulási pontból lehessen továbbjavítani, ami minden modellszámítás nélkül is rendelkezésre áll, vagyis egy monoton összefüggésből, ahol a becslés alapértéke a következményváltozó-érték, mely gyakrabban fordul elő a másiknál. A modellezés érdemi mozgástere ezen monoton becslés feljavítása a 100 %-os találati irány mentén – lehetőség szerint úgy, hogy többet lehessen nyerni a monoton stratégiától való eltéréssel a ritkább következmények esetében, mint amennyi ennek találati arányra gyakorolt vesztesége a gyakori következményeknél.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Felvásárlás (ország azonosság) | | Becslés | | Összesen |
| -1 | 1 |
| USA kizárásával | | | | |
| db | -1 | 81 | 0 | 81 |
| 1 | 19 | 2 | 21 |
| % | -1 | 100% | 0% | 100% |
| 1 | 90% | 10% | 100% |

**17. ábra: Becslés csoportba tartozáshoz - Hasonlóságelemzés**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Azok az eredmények tehát, ahol egy összefüggésrendszer alig képes elmozdulni a monoton becslések nyújtotta találati arány szintjéről, lényegében még instabilak, kevéssé értékesek. Az elmozdulás hiányának egyik triviális oka a magyarázó tényezők inadekvát jellegében keresendő, míg a másik erőtér a megfelelő matematikai apparátus meg nem találását jelenti. A neurális hálók és a lépcsős függvények tehát csak jelképes elmozdulást tudtak kikényszeríteni a felvásárlás ország-azonossága jelenségkör kapcsán – ami összevetve a korábbi fejezet alacsony kontingenciáival az inputtényezők alacsony klasszifikáló erejét erősítik meg.

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| USA kizárásával (Kontingencia Mátrix) | | |
| 81.37% | 18.62% | 100.00% |

**18. ábra: Becslés csoportbatartozáshoz – Kontingencia mátrix (USA kizárásával)**

**Forrás:** <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

Jó = Találatok száma / Összes objektum = (81 + 2) / 102

Rossz = Hibák száma / Összes objektum = (0+ 19) / 102

1. **Összefoglalás**

Konklúzióként levonhatjuk, hogy a kísérletben a Mesterséges Intelligencia használt eszközei (neurális háló és hasonlóságelemzés) képes volt a jobb eredmény elérésére, azaz az adatokban rejlő rejtett mintát látszólag magasabb precizitással a felszínre tudták hozni, míg a diszkriminancia-elemzés sugallt eredménye, hogy az adatokban kevesebb az összefüggés, a függvényszerű megfeleltetés. Érdemes lehet mindezek után levonni a következtetést, hogy mely módszerek tekinthetőek valóban alkalmasnak további kutatás-elemzés céljából.

A starup-ok szempontjából az elemzések arra mutatnak rá, hogy az adatvagyon relatíve nagy volumene ellenére érdemi klasszifikáló erőről az egyes mutatószámok alapján nem lehet beszélni…

1. **Irodalomjegyzék**
2. Harbert T. (2014): Competing with Two Guys in a Garage. Letöltve: [www.rewrite.ca.com](http://www.rewrite.ca.com), <http://rewrite.ca.com/us/articles/application-economy/competing-with-two-guys-in-a-garage.html> 2017.04.15
3. Magyar Startup Közösség (2016): Mi az a Startup? Letöltve: <http://magyar-startup-kozosseg.hu/> 2017.04.15
4. Sajtos L. – Mitev A. (2007): SPSS kutatási és adatelemzési kézikönyv. Budapest, Alinea Kiadó.
5. Sebastian Raschka (2015): Python Machine Learning, Packt Publishing, Birmingham.
6. Robehmed N. (2013): What Is A Startup? Letöltve: [www.forbes.com](http://www.forbes.com), [http://www.forbes.com/sites/natalierobehmed/2013/12/16/what-is-a-startup  
   /#1e23b2714c63](http://www.forbes.com/sites/natalierobehmed/2013/12/16/what-is-a-startup/#1e23b2714c63) 2017.04.15
7. Melléklet: Korrelációelemzés diszkriminancia-elemzéshez (saját szerkesztés)

**Forrás**: <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Finanszírozási körök (db) | Egy főre jutó finanszírozás értéke (1000 USD) | Eltelt napok száma az alapítás és felvásárlás között | Eltelt napok  száma az első finanszírozás és a felvásárlás között | Eltelt napok száma az utolsó finanszírozás és felvásárlás között | Eltelt napok száma az alapítás és az első finanszírozás között | Eltelt napok száma az alapítás és utolsó finanszírozás között |
| Teljes adathalmaz | | | | | | | |
| Finanszírozási körök (db) | 1.00 | 0.23 | 0.33 | 0.48 | -0.18 | -0.07 | 0.47 |
| Egy főre jutó finanszírozás értéke (1000 USD) | 0.23 | 1.00 | 0.07 | 0.08 | -0.08 | 0.01 | 0.12 |
| Eltelt napok száma az alapítás és felvásárlás között | 0.33 | 0.07 | 1.00 | 0.77 | 0.46 | 0.63 | 0.87 |
| Eltelt napok  száma az első finanszírozás és a felvásárlás között | 0.48 | 0.08 | 0.77 | 1.00 | 0.55 | -0.02 | 0.56 |
| Eltelt napok száma az utolsó finanszírozás és felvásárlás között | -0.18 | -0.08 | 0.46 | 0.55 | 1.00 | 0.05 | -0.04 |
| Eltelt napok száma az alapítás és az első finanszírozás között | -0.07 | 0.01 | 0.63 | -0.02 | 0.05 | 1.00 | 0.68 |
| Eltelt napok száma az alapítás és utolsó finanszírozás között | 0.47 | 0.12 | 0.87 | 0.56 | -0.04 | 0.68 | 1.00 |
| USA kizárásával | | | | | | | |
| Finanszírozási körök (db) | 1.00 | 0.46 | 0.40 | 0.51 | -0.20 | -0.01 | 0.56 |
| Egy főre jutó finanszírozás értéke (1000 USD) | 0.46 | 1.00 | 0.11 | 0.14 | -0.12 | -0.01 | 0.19 |
| Eltelt napok száma az alapítás és felvásárlás között | 0.40 | 0.11 | 1.00 | 0.78 | 0.42 | 0.59 | 0.84 |
| Eltelt napok  száma az első finanszírozás és a felvásárlás között | 0.51 | 0.14 | 0.78 | 1.00 | 0.53 | -0.04 | 0.53 |
| Eltelt napok száma az utolsó finanszírozás és felvásárlás között | -0.20 | -0.12 | 0.42 | 0.53 | 1.00 | -0.01 | -0.14 |
| Eltelt napok száma az alapítás és az első finanszírozás között | -0.01 | -0.01 | 0.59 | -0.04 | -0.01 | 1.00 | 0.65 |
| Eltelt napok száma az alapítás és utolsó finanszírozás között | 0.56 | 0.19 | 0.84 | 0.53 | -0.14 | 0.65 | 1.00 |

1. Melléket: Diszkriminancia-elemzésből kizárt objektumokra vonatkozó Mahalabonis-távolság értékek

**Forrás**: <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |  |
| --- | --- | --- | --- |
| Szervezet megnevezése | Mahalanobis-távolság | Változók száma | Mahalanobis-távolság / Változók száma |
| Duck Creek Technologies | 25.63 | 6 | 4.27 |
| Wireless Ronin Technologies | 26.71 | 6 | 4.45 |
| Insurance Noodle | 32.24 | 6 | 5.37 |
| Mintera | 96.06 | 6 | 16.01 |
| Simpel Corporation | 49.69 | 6 | 8.28 |
| Desterra | 143.69 | 6 | 23.95 |
| Aristos Logic | 327.11 | 6 | 54.52 |
| Skire | 25.37 | 6 | 4.23 |
| Xigo | 29.57 | 6 | 4.93 |
| Vaultus | 39.81 | 6 | 6.64 |
| Endurance Specialty Holdings | 40.04 | 6 | 6.67 |
| Synoptek | 26.39 | 6 | 4.40 |
| GCA Services Group | 33.08 | 6 | 5.51 |
| O3b Networks | 39.22 | 6 | 6.54 |

1. Melléklet:

Becslés csoportbatartozáshoz: nyelv azonosság (diszkriminancia-elemzés)

**Forrás**: <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Felvásárlás (nyelv azonosság) | | Előrejelzés | | Összesen |
| -1 | 1 |
| Teljes adathalmaz | | | | |
| db | -1 | 44 | 43 | 87 |
| 1 | 257 | 408 | 665 |
| % | -1 | 51% | 49% | 100% |
| 1 | 39% | 61% | 100% |
| USA kizárásával | | | | |
| db | -1 | 27 | 14 | 41 |
| 1 | 32 | 29 | 61 |
| % | -1 | 66% | 34% | 100% |
| 1 | 52% | 48% | 100% |

Becslés csoportbatartozáshoz – Kontingencia mátrix (Teljes Adathalmaz)

Jó = Találatok száma / Összes objektum = (44 + 408) / 752

Rossz = Hibák száma / Összes objektum = (43 + 257) / 752

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| Teljes adathalmaz (Kontingencia Mátrix) | | |
| 60.11% | 39.89% | 100.00% |

Becslés csoportbatartozáshoz – Kontingencia mátrix (USA kizárásával)

Jó = Találatok száma / Összes objektum = (27 + 29) / 102

Rossz = Hibák száma / Összes objektum = (14 + 32) / 102

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| USA kizárásával (Kontingencia Mátrix) | | |
| 54.90% | 45.10% | 100.00% |

1. Melléket: Becslés csoportbatartozáshoz: pénznem azonosság (diszkriminancia-elemzés)

**Forrás**: <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Felvásárlás (pénznem azonosság) | | Előrejelzés | | Összesen |
| -1 | 1 |
| Teljes adathalmaz | | | | |
| db | -1 | 62 | 80 | 142 |
| 1 | 218 | 392 | 610 |
| % | -1 | 44% | 56% | 100% |
| 1 | 36% | 64% | 100% |
| USA kizárásával | | | | |
| db | -1 | 44 | 34 | 78 |
| 1 | 11 | 13 | 24 |
| % | -1 | 56% | 44% | 100% |
| 1 | 45% | 54% | 100% |

Becslés csoportbatartozáshoz – Kontingencia mátrix (Teljes Adathalmaz)

Jó = Találatok száma / Összes objektum = (62 + 392) / 752

Rossz = Hibák száma / Összes objektum = (80 + 218) / 752

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| Teljes adathalmaz (Kontingencia Mátrix) | | |
| 60.37% | 39.63% | 100.00% |

Becslés csoportbatartozáshoz – Kontingencia mátrix (USA kizárásával)

Jó = Találatok száma / Összes objektum = (44 + 13) / 102

Rossz = Hibák száma / Összes objektum = (34 + 11) / 102

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| USA kizárásával (Kontingencia Mátrix) | | |
| 55.88% | 44.12% | 100.00% |

1. Melléket: Becslés csoportbatartozáshoz: vallás azonosság (diszkriminancia-elemzés)

**Forrás**: <https://data.crunchbase.com/v3/docs/excel-export>, saját szerkesztés

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Felvásárlás (vallás azonosság) | | Előrejelzés | | Összesen |
| -1 | 1 |
| Teljes adathalmaz | | | | |
| db | -1 | 17 | 16 | 33 |
| 1 | 269 | 450 | 719 |
| % | -1 | 52% | 48% | 100% |
| 1 | 37% | 63% | 100% |
| USA kizárásával | | | | |
| db | -1 | 9 | 4 | 13 |
| 1 | 24 | 65 | 89 |
| % | -1 | 69% | 31% | 100% |
| 1 | 27% | 73% | 100% |

Becslés csoportbatartozáshoz – Kontingencia mátrix (Teljes Adathalmaz)

Jó = Találatok száma / Összes objektum = (17 + 450) / 752

Rossz = Hibák száma / Összes objektum = (16 + 269) / 752

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| Teljes adathalmaz (Kontingencia Mátrix) | | |
| 62.10% | 37.90% | 100.00% |

Becslés csoportbatartozáshoz – Kontingencia mátrix (USA kizárásával)

Jó = Találatok száma / Összes objektum = (9 + 65) / 102

Rossz = Hibák száma / Összes objektum = (4 + 24) / 102

|  |  |  |
| --- | --- | --- |
| Becslés | | Összesen |
| Jó | Rossz |
| Teljes adathalmaz (Kontingencia Mátrix) | | |
| 72.55% | 27.45% | 100.00% |

1. A startup-okról rendelkezésre álló statisztikákban a startup szöveges leírása szövegbányászati megoldásokkal elvezethet egy olyan osztályozáshoz, ahol az egyes osztályok aránya számszerűen levezethető. [↑](#footnote-ref-1)
2. <https://www.crunchbase.com/#/home/index> [↑](#footnote-ref-2)
3. 2017. május [↑](#footnote-ref-3)
4. További ingyenes adatbázisok elérhetőek az alábbi linkekről:

   <https://www.f6s.com/f6s>

   <https://angel.co/>

   https://www.owler.com/

   https://startupxplore.com/startups/index

   http://www.startupranking.com/ [↑](#footnote-ref-4)
5. amely sajnos erősen hiányos, csaknem 90%-ban nincs kitöltve. Ennek oka vélhetően az üzleti titok fogalmához kapcsolódik, hiszen a feleknek nem feltétlenül érdeke a szerződés teljes nyilvánossága. [↑](#footnote-ref-5)
6. Észak- és Közép-Amerika a kutatásban azonos kontinensként, míg Dél-Amerika külön kontinensként szerepel. [↑](#footnote-ref-6)
7. Hivatalos nyelvek adatbázisaként az alábbi hivatkozás került felhasználásra: <https://www.infoplease.com/world>  
   /countries-world/languages-spoken-each-country-world [↑](#footnote-ref-7)
8. Adott ország hivatalos pénzneme az alábbi adatbázis alapján lett megállapítva: https://www.countries-ofthe-world.com/world-currencies.html [↑](#footnote-ref-8)
9. A hivatalos vallás az alábbi nyilvántartás segítségével került meghatározásra: <http://www.nationmaster.com>  
   /country-info/stats/Religion/Religions [↑](#footnote-ref-9)
10. <https://www.crunchbase.com/#/home/index> [↑](#footnote-ref-10)
11. O3b Networks Inc. nélküli adat. [↑](#footnote-ref-11)
12. A Vocalocity esetében történt (felhőszolgáltató), mely céget 2013-ban a Vonage vásárolt fel 130 millió dollár értékben. [↑](#footnote-ref-12)