Tanulók tevékenységének nyomon követése e-learning keretrendszerben

avagy nem minden Black Mirror, ami sötétnek látszik

(Log-analyses in e-learning frameworks – (or dark phenomena are not always Black Mirrors)

Kivonat: A tanulmány egy sorozat második részeként kísérletet tesz arra, hogy valós, de anonimizált adatvagyon alapján egy konkrét oktatási rendszer paramétert érintő döntési helyzet kapcsán bemutassa, miként lehet egy szubjektív/intuitív emberi döntés előtt, ennek előkészítéseként és/vagy utólag a döntéshozatal folyamatát kibernetikus rendszerként támogatni.

Kulcsszavak: adatvagyon-gazdálkodás, log-adatok, elemzésmódszertan, döntéstámogatás, hermeneutika

Abstract: The article - as part II in an article series - based on anonymized data assets about educational phenomena tries to demonstrate the complex handling of a real (intuitive/subjective) situation in frame of a cybernetic process chain (before, during and/or after the decision making as such).

Keywords: data asset management, log-data, analytical methodology, decision support, hermeneutics

[[1]](#footnote-1)

Tartalomjegyzék

[1. Bevezetés 3](#_Toc505332841)

[2. Szakirodalmi áttekintés 5](#_Toc505332842)

[3. Az adatvagyon jellemzői általában 7](#_Toc505332843)

[3.1. Elsődleges adat-ellenőrzések és ezek eredményei 7](#_Toc505332844)

[3.2. A feldolgozásra alkalmasnak ítélt adatvagyon karakterisztikái 8](#_Toc505332845)

[4. A probléma 8](#_Toc505332846)

[5. Modellek 12](#_Toc505332847)

[5.1. Idősoros elemzések 12](#_Toc505332848)

[Az elemi szál 13](#_Toc505332849)

[Konzisztencia-alakzatok 17](#_Toc505332850)

[5.2. Genetikai potenciál elemzések 17](#_Toc505332851)

[Döntés előtt 18](#_Toc505332852)

[Döntés után 21](#_Toc505332853)

[5.3. Döntési kényszer-index 23](#_Toc505332854)

[A nyers és a sorszámozott tanulási minta 23](#_Toc505332855)

[Modellek 24](#_Toc505332856)

[Eredmények 24](#_Toc505332857)

[5.4. Potenciáltól való eltérések modellje 28](#_Toc505332858)

[6. Összegzés 30](#_Toc505332859)

[6.1. Fejlesztési terv/vázlat 31](#_Toc505332860)

[6.2. Célok és várható eredmények 31](#_Toc505332861)

[A fejlesztés szükségessége 32](#_Toc505332862)

[Lépések 32](#_Toc505332863)

[7. Melléklet 34](#_Toc505332864)

[8. Irodalomjegyzék 35](#_Toc505332865)

# Bevezetés

*A bevezetésben a tanulmány alapvetései előtt* ***elsődlegesen*** *az alcímet illik tisztázni:* A Black Mirror az információs társadalom már ma is felsejlő jövőképeivel való szembenézés támogatása a film eszközeivel. **Szembesülni** kevésbé a boldog kifejletekkel, mint inkább a **potenciális árnyoldalakkal** szokás és érdemes. A társadalmak és/vagy az egyének személyiségfejlődése kapcsán a látszólag fekete/fenyegető jelenségekkel való megküzdés az a kihívás, mely előtt elfutni sem nem érdemes, sem nem lehetséges.

*A tanulmány lényege a felkészítés az autodidakta és/vagy az (elvárt/tűrt) automatizált döntéstámogatásra:* A **tanulmány** célja, hogy egy anonim adatvagyon elemzésével párhuzamosan (hosszabb távra tekintve **sorozat** **formájában**) **rá** lehessen **mutatni** arra **milyen döntési pontok** mentén is **értelmezhetők** a tanulás, a tanítás, a tanuló, az oktató, a tananyag, ill. bármely más oktatási erőforrás? A sorozat első része a 2017-es EDEN konferencián került bemutatásra [13]. Ebben a részben a fejlesztés stratégia tervét ismerhette meg az Olvasó. Az EDEN-konferencia keretében tartott prezentáció már átlépett az esettanulmányok bevonásának szintjére [14]. Ezt az irányt folytatja a 2. rész. A 3. részben a dashboard-ok és az OLAP-szolgáltatások elméleti és gyakorlati aspektusai kerülnek fókuszba. A projekt kapcsán a belső peer-review eljárások keretében fokozatosan kerülnek kijelölésre a cikksorozat potenciális elemei, melyek feladata az intuitív szakértői felvetések mögötti bizonyító erejű elemzése minél inkább automatizált lehetőségének kidolgozása és bemutatása.

Az e-learning-alapú oktatás keretében a **tanulási/tanítási folyamat**ról ennek **reprodukálhatóságát** **szem előtt tartó naplózással** tranzakciós szintű adatok állíthatók elő. De még blended learning keretrendszerben is követhető a tanár-diák interakciók virtuális térben zajló vetülete. A reprodukálható tanár/diák-rendszerhasználat alapozza meg azt az adatteret, **mely nem kell, hogy előre definiált elemzési célok mentén csak az ehhez szervesen illeszkedő részleteket gyűjtése, hiszen a reprodukálható tanulási/tanítási folyamatok adatsoraiban minden benne van. Még azon célok támogatására alkalmas adatok is, melyek az adatgyűjtés elrendelése kapcsán fel sem merültek tételesen**. Az elemzési célok pedig egymást katalizálják: adott kérdésre adható legvalószínűbb válasz megismerése új kérdéseket generál általában véve.

Mottó: az tanul hatékonyan egy adott jelenségről, aki előtte már megtanult tanulni!

*Naplózni eddig is naplóztak az informatikai szakemberek:* A tanulási/tanítási folyamat e-learning rendszeren belüli **nyomon követés**e kapcsán abból a **spontán** ön**fejlődő** állapotból kell kiindulni, ahol az e-learning keretrendszerek **alapvetően** **műszaki, önellenőrzési, üzemelés-biztonsági, jogosultságkezelési, stb., ill. ezzel párhuzamosan általános rendszerhasználat-elemzési okokból képeztek log-állományokat**, vagyis a kezdetekben még jórészt nem a tanulás/tanítás-módszertani optimalizálás, az általános döntéstámogatás alapján kerültek a naplózott tranzakciók kiválasztásra. A spontán adatvagyon-gyűjtés **ellenpontja a tanulási/tanítási magatartás teljes reprodukálhatósága**, ahol adott tanuló/tanár belépését követően minden képernyőpixel, egérmozgás, billentyűleütés, sőt akár a felhasználó szemkamera általi képernyő értelmezése is bármikor az elemző számára pontosan úgy követhető, **mintha az elemző mindvégig a tanuló/tanár háta mögül figyelte volna a történéseket valós időben** és/vagy utólag bármikor/bárhányszor.

A naplózás melletti mindennapokra az első példák egyike az átlag-munkavállaló életében a menetlevél. Hasonlóképpen szinte mindenki számára ismert log-megjelenési forma az egyes szerverek látogatottságának mérése, az egyes tartalmak letöltésének mennyisége, vagy éppen a like-ok száma.

*A jogi szabályozás nem öncél!* A tanulmány itt és most csak érintőleg foglalkozik azokkal a jogi kérdés-komplexumokkal, melyek az ilyen jellegű adatkezelések kapcsán automatikusan és szinte mindenkiben felmerülnek. Amennyiben a tanulás/tanítás munkahelyi tevékenységként kerül előírásra (vö. továbbképzések), akkor a munkáltató a pszicho-szociális kockázatok kezelését előíró törvény kapcsán [15] eleve köteles a munkavállaló potenciális érő kitettségeket folyamatosan értelmezni tudni. Másrészt természetesen a munkavállalóknak (tanulóknak/tanároknak) joga van bizonyos mértékben pl. a személyes adatok védelméhez. Az alapjogokra visszavezetődő antagonizmusok sorát lehet vélelmezni azon konkrét döntési probléma kapcsán: mi szerint hol ér véget a személyes élettér jogos védelme és hol lép be a közösségi érdek (vö. stresszes munkavállaló okozta munkahelyi baleset)? S itt kell utalni arra is, hogy magának **az egyénnek is vannak antagonista érdekei**: vagyis a személyes adatok védelme mellett a tanulási környezet, mint szolgáltatás minőségének magas szintje, ezen minőség fejlődése, a rendszer minél inkább testre szabott (vö. kockázat-minimalizáló, erőforrás-optimalizáló) működése mind-mind olyan elvárások, melyek pl. fogyasztóvédelmi aspektusból azonnal fontosak lennének, ha az egyén a rendszerrel pl. előfizetés keretében lépne üzleti/jogi kapcsolatba. Operatív jogi kérdés lesz hosszabb távon minden potenciális érintett (oktatók, tanulók, minőségbiztosítók, tananyagfejlesztők, stb.) számára olyan szerződések prezentálása az e-learning rendszerbe való első belépés előtt, ennek kapcsán, ahol az adatgyűjtés mikéntje és az adatfeldolgozás céljai az érintettek tudomására kerülnek.

A jog és a valóság szerves kapcsolatát leíró egyik legaktuálisabb példa a XXI. században annak a kérdésnek az elemzése, miszerint: egy önvezető autó kit veszélyeztessen elsőként, ha mindenképpen elkerülhetetlen az emberi sérülés: a saját utasait, az idegen a járókelőket, az egyéb járművek utasait?

*Valódi és gondolatkísérlet formájában lehet eljutni az általánosítás képességéhez:* **Jelen tanulmány egy kísérleti adatgyűjtés valós tapasztalatait értelmezi általában véve és vezeti le** azokat **az** **általános érvényű elemzés-módszertani részleteket**, melyekkel az Olvasók, mint az LLL (life long learning) kapcsán előbb-utóbb, ill. kisebb-nagyobb rendszerességgel érintetté válók (tanulók és/vagy tanárok és/vagy elemzők, stb.) hasznos, ha minél hamarabb, minél többször, minél több szempontból szembesülnek.

A gondolatkísérletek egyik magyar nyelven született, megkerülhetetlen irodalmi alkotása a Szörnyeteg a hajtóműben című scifi – Szélesi Sándor többszörösen díjazott szerző tollából…

\*\*\*

A tervezett cikksorozat ezen részében egy **konkrét döntési helyzet** és az e körüli elméleti kérdések kerülnek bemutatásra. A döntési helyzet lényege, hogy véletlenszerűen éppen egy blended learning képzésen belül, egy e-szemináriumi foglalkozás-sorozat keretében kiadott feladatokra 2500 karakterben limitált hallgatói válaszok hosszának dinamikus figyelésével egy robotizált LMS (learning management system) **mikor és milyen irányba, milyen mértékben kellene, hogy módosítsa (módosítani javasolja) a 2500 karakteres felső határt a szöveghossz tekintetében?** Más megfogalmazásban: **felismerhetők-e a rendszerhasználati adatok alapján feszültségek a rendszerben, ill. hogyan néz ki adott felhasználás mellett egy rendszerparamétert érintően a döntési kényszer-index időbeli alakulása,** ahol döntési kényszer-index alatt adott paraméter által generált rendszerfeszültség idősoros alakulását kell érteni.

Az esettanulmány bemutatja a szubjektív/intuitív döntéshozatalt magát, ill. **egy idősoros és egy szakértői rendszer-elvű szimuláció** (egy ún. genetikai potenciálszámítás) gondolatmenetét, eredményeit a **döntéshozatal előtt**, majd ellenőrző jelleggel a **döntés után**i rendszer-történések feldolgozása mellett is – a döntés helyességének utólagos ellenőrzését lehetővé téve (ismét csak idősorosan és szimulatív módon).

A karakterlimit változtatása ebben a cikkben nem teljes komplexitásában kerül értelmezésre. A log-alapú rendszerértelmezés jelenleg még nem (lehet) képes stratégiai paraméterekkel dolgozni. Ilyen **stratégiai paraméter** pl. az, hogy egy adott szakterület munkavállalóinak kommunikációs megoldásait milyen kockázatokkal vélünk terheltnek (vö. jogi szaknyelv értelmezhetőségi zavarai, terjengőssége – amivel szemben pl. a haiku-világ lehetne a rendszerelméleti ellenpont – s vélelmezhetően optimum jellegű megoldás lenne az ideális). A log-alapú megközelítés keretei között (vö. Föld-bolygó történései alapján) pedig sajnos tetszőleges általánosságokról (vö. csillagászatról) nem lehet érdemben beszélni…

# Szakirodalmi áttekintés

Az oktatásban használatos adatvagyonok feltérképezése a nemzetközi szakirodalomban sem nyúlik vissza hosszú múltra. A webes adatbányászat az adatbányászat tágabb témakörének részeként a kétezres évek fordulóján jelent meg először a nemzetközi színtéren, bár Etzioni már 1996-ban felvetette a World Wide Weben fellelhető adatok felhasználásának lehetőségeit és problémáit. Zaïane és Lou (2001) vizsgálta a webes adatbányászat lehetőségeit az online környezetek javítása érdekében. Pahl (2004) tanulmánya már kifejezetten az e-learning tananyagokkal kapcsolatos interakciókat értékelte. Nachimas és Herskovitz (2007) az online tanulói viselkedésmintázatok vizuális megjelenésének lehetőségeit járta körbe. Ai és Laffey (2007) tanulmányukban arra a következtetésre jutottak, hogy a webes adatbányászat segítségével következtetni lehet a hallgatói sikerességre, így hasznos eszköznek tartják a tanulási eredmények javítására is.

Az újabb szakirodalomban Educational Data Mining (EDM) és Learning Analytics (LA) kulcsszavakkal tárgyalják az oktatási adatbányászat és tanuláselemzések témakörét.

Romero és Ventura 2010-es tanulmánya nyújtja az egyik legteljesebb áttekintést az adatbányászatról általánosságban, illetve az online oktatási környezetekben végzett webes adatbányászati eredményekról. Meghatározásuk szerint (Romero és Ventura, 2010) az oktatási adatbányászat olyan számítógépes módszerek kifejlesztése, kutatása és alkalmazása, amelyek segítségével olyan mintákat lehet azonosítani oktatással kapcsolatos nagy méretű adatvagyonokban, amelyek más módszerekkel való azonosítása nehéz, vagy lehetetlen lenne, az adatok óriási mennyiségéből fakadóan. A témával foglalkozó publikációk száma az utóbbi években ugrásszerűen megnőtt, érzékeltetve ezzel is a téma iránti tudományos érdeklődést.

A szerzők az oktatási adatbányászat alábbi leglényegesebb céljait határozták meg:

1. Adatelemzés és vizualizáció
2. Visszajelzés az oktatóknak
3. Ajánlások a hallgatóknak
4. Hallgatói teljesítmény előrejelzése
5. Hallgatói viselkedés modellezése
6. Nemkívánatos hallgatói viselkedés azonosítása
7. Hallgatók csoportosítása tanulási jellemzőik alapján
8. Szociális hálózatok elemzése
9. Koncepciótérképek kialakítása
10. Tartalomfejlesztés
11. Oktatástervezés

A tudományterület fejlettségét a szerzők „kamaszkorúnak” definiálják érzékeltetve, hogy már nem a kezdeti stádiumban van, de még sok fejlődési lehetőség előtt áll.

Scheuer és McLaren áttekintő tanulmányukban felhívják a figyelmet a felhasználók, hallgatók fiziológiai jellemzőinek észlelésére és vizsgálatára is (arckifejezés, szemmozgás, izzadás, testhelyzet), amelyekből származtatott adatok szintén bevonhatók az elemzésbe. A tudományág legfontosabb kérdéseinek egyrészt azt tekintik, hogy az addig „fekete doboznak” tekintett oktatási folyamatba betekintést nyerhetünk, ezáltal jobb rendszereket fejleszthetünk ki az adatelemzés segítségével. Másrészt az ezen eredmények alapján kidolgozott és alkalmazott újszerű technológiák, mint a játékos tanulás, mobil tanulás, mindenhol jelenlevő (ubiquitous) tanulás, kifinomultabb, személyre szabott oktatási módszerek alkalmazását teszik lehetővé.

A továbbiakban a „kemény” adatok (kattintások, oldalakon töltött idő, videó-tananyag hossza, stb.) elemzésén túl a hallgatók és oktatók szóhasználatának vizsgálatával a természetes nyelvi feldolgozás (natural language processing, NLP) segítéségével lehet következtetéseket levonni, predikciókat tenni a felhasználói viselkedés jövőbeli kimenetelével kapcsolatban. Ahogy Crossley és szerzőtársai (2015) kimutatták, 67.8%-ban meg lehet jósolni a MOOC résztvevők szóhasználatából, hogy le fognak-e morzsolódni, vagy sem.

A kutatás másik iránya nem közvetlenül az online felületről kinyert aktivitásokból építkezik, hanem a hallgatói értékeléseket vizsgálja. Az oktatói munka hallgatói véleményezéséből származó adatok statisztikai módszerekkel való vizsgálatával is értékes következtetésekre lehet jutni (Jiang et al, 2015).

A magyar nyelvű szakirodalomban Balogh és Horváth 2010-es tanulmánykötete mutatja be legátfogóbban a hazai kezdeményezéseket. Tóth (2010) egy Moodle kurzus használatát értékeli a hallgatói aktivitások elemzésével, melynek alapján különböző hallgatói típusokat (kitartó, feladó, kóstolgató) különböztet meg tanulási stílusuk alapján. Az eredmények alapján oktatásszervezési, tutorálással kapcsolatos ajánlásokat is tesz.

Nyéki (2010) a Coedu keretrendszer használatát elemezve tesz hasonló következtetéseket a rendszerhasználati szokásokkal, és a hallgatók motivációjával kapcsolatban.

Jókai és kollégái (Jókai et al, 2010) kérdőívek, a Moodle beépített statisztikái és adatbányászati eszközök segítségével három fő tanulási mintát azonosítottak a tananyagok, tesztek és gyakorlatok használatának különböző arányának elkülönítésével.

Petákné Balogh (2014) disszertációjában az ILIAS LMS használatának statisztikai elemzéséből vont le következtetéseket az online aktivitás mértéke és a hallgatók vizsgaeredményeinek összevetésével. Megállapította, hogy az ILIAS-ban töltött idő szignifikánsan javítja a hallgatók által teljesített kreditek arányát, a nappali képzési formában még jobban is, két és félszer nagyobb mértékben, mint a távoktatásosoknál.

Az elemzési igényekkel párhuzamosan az elemzési módszertanok fejlődése is megfigyelhető. Az ebben a cikkben alkalmazott módszertan a hasonlóságelemzés lesz, melyről a nemzetközi összehasonlító elemzések nagyfokú rugalmasságot állapítanak meg [12]

Látható tehát, hogy a tudományterület „feljövőben van”, egyre nagyobb a tudományos érdeklődés az online tanulási folyamat során keletkezett adatok gyűjtésére, elemzésére, következtetések levonására. Az áttekintett nemzetközi és hazai szakirodalom alapján megállapíthatjuk, hogy a kutatások csapásirányai egyfelé mutatnak, hasonló kutatási kérdések merülnek fel a témával foglalkozó cikkekben.

Az idősor-elemzés látszólag kiforrott szakterülete a matematikai-statisztikának [16], mégis ki kell emelni, hogy alapvető kérdések nem kerültek tisztázásra évtizedek alatt: pl. melyik adott adatsor esetén a leginkább adekvát függvény-típus egy trendvonal illesztését elvárva?

Szakértői rendszerek: A cikkírása során a szakértői rendszerek is bevonásra kerülnek a modellezési folyamatba. A szakértői rendszerek [17] itt és most legfontosabb tulajdonsága, hogy a bennük tárolt kombinatorikai tudás keletkezhet induktív módon is, vagyis a szakértői rendszer képes még soha nem tapasztalt paraméter-együttállások következményeire vonatkozóan becsléseket adni (vö. genetikai potenciálbecslés).

Szimulációk: „A szimuláció olyan vizsgálat, amikor egy rendszer, folyamat fizikai vagy számítógépes modelljén tanulmányozzák a rendszer várható, illetve valódi viselkedését” [18]. A szimuláció mögött sokféle modell lehetséges, pl. szakértői rendszer is. Így az egyes kulcsfogalmak értelmezése egymástól nem függetleníthető érdemben.

Konzisztencia: A modellezés során az egyes modell-variánsokról tetszőlegesen sokféle mutató határozható meg, s ezek mindegyike kapcsán lehetséges arra vonatkozó vélelemet megfogalmazni, vajon a mutató értéke mikor utal jobb modellre. A modellezés során tehát a legjobb modellt sok mutató ideális kombinációjaként kell értelmezni, ami nem más, mint a modellezés konzisztenciájának maximalizálása [pl. 19].

# Az adatvagyon jellemzői általában

*Az adattemető olyan adatbázis, melybe zömmel csak befelé áramlanak a jelek, s szinte senki, semmikor nem kíván ezen jelek alapján információkat előállítani, új tudást generálni:*

Az adatvagyon fogalmát akkor érdemes használni, **ha** **érdemi adatvagyon-gazdálkodás folyik** az tanulási/tanítási folyamatokat naplózó adatbázisokhoz kapcsolódóan. Az adatvagyon-gazdálkodás lényege, hogy **minden gyűjtött adatról a lehető legpontosabban és részletesebben le van írva ennek jelentése, mértékegysége, értelmezési tartománya (vö. metaadat)**. Ennek hiányában, vagy ennek hozzáférhetetlensége esetén a minimális értelmezési kockázatokat fel kell tárni, ami stratégiailag annyit jelent, hogy ki kell tudni alakítani az adatvagyon azon szűrletét, mely elemzésre érdemesnek tűnik. **Operatívan** ez azt várja el, hogy **feltárásra kerüljenek** egy fajta **plauzibilitás és/vagy konzisztencia-tesztek formájában** **az egyes nyers jelenségek alapstatisztikái** (maximum, minimum, átlag, medián, szórás, stb.), **ill. az egyes jelenségek egymással való összevetéseiből származtatott mutatók** esetén hasonló karakterisztikák (pl. korrelációk).

Az adatvagyon-gazdálkodás minimumára jó példa a TEIR meta-keresője (<https://www.teir.hu/rqdist/main?rq_app=meta&rq_proc=meta>), ahonnan a kulcsszó megadása után az ezzel kapcsolatos adatvagyonok legfontosabb részletei azonnal elérhetők.

## Elsődleges adat-ellenőrzések és ezek eredményei

Az egyik leghatékonyabb ellenőrzés a nyers (naplózott) jelenségek **szélsőértékeinek**, (különös tekintettel maximális értékeinek, ill. az ezek mögött megbújó eloszlások) feltárása. Konkrét példaként érdemes kiemelni a múltból a videó-blokkokkal támogatott e-learning tartalmak esetén pl. az előre/hátra-tekerések számát, mint nyers jelenséget. Amennyiben egy adott hosszúságú videón belül az előre/hátra-tekerések maximális száma több nagyságrenddel meghaladja a videó-hossz másodpercekben kifejezett értékét, akkor ez egy olyan plauzibilitási határérték-átlépés, mely kapcsán az adott jelenségről rendelkezésre álló adatok értelmezhetősége megkérdőjelezhető, mert fizikailag nem hajtható végre emberi kéz által ennyi manipuláció adott idő alatt. A konkrét példa szálát követve ilyen esetben a nyersadatok gyűjtésének algoritmikus részét kell ellenőrizni, ahol vélelmezhetően feltárható lesz, hogy az előre/hátra-tekerések száma (mely már eleve nem is feltétlenül egy klasszikus nyers adat, hanem nyers adatok kumulált értéke) olyan módon került gyűjtésre, ahol a kumulációt végző adattároló nem minden elvárt ponton került kinullázásra, így olyan halmozódás lépett fel, mely a racionális tartományból kitolta látszólag egyes konkrét felhasználók esetében a maximum értékét.

Egy másik klasszikus ellenőrzési problémakör az **időpecsétek** léte és ezek láncolata, valamint az időpecsétek esetleges hiányának kezelése. Ha például egy e-tanulási folyamatba való belépés időpecsétje általában létezik, már akkor is felmerül a kérdés, lehet-e egyáltalán, s ha igen, miáltal lehetséges az, hogy ez az adat ne létezzen? Az e-tanulási folyamatból való kilépés időpecsétje a belépéssel ellentétben már olyan adat, ahol az adatok jelentős része nem konkrét log-adat, hiszen pl. az internetkapcsolat megszakadása, a futtató számítógép lefagyása, kék-halála, stb. miatt a naplózási folyamat megszakadhat valóságosan is anélkül, hogy ezt bármilyen módon a szerveroldali adatbázisok észlelni lennének képesek.

## A feldolgozásra alkalmasnak ítélt adatvagyon karakterisztikái

A feldolgozásra való alkalmasság célfüggő. Amennyiben az elemzési cél a tanulási folyamat és a tanulási eredményesség közötti kapcsolat, akkor elsődlegesen olyan tanulók adatsoraira van csak szükség a tanulási mintázatok közül, ahol a tanulás eredményességéről is van adat.

Amennyiben többtáblás adatgyűjtés történik a redundáns tartalmak és a tárolandó adatmennyiség minimalizálása miatt, akkor kínosan kell ügyelni arra, hogy minden tranzakció tartalmazza az összes olyan indexet, melyek alapján az egyes adatbázis-táblák összekapcsolása félreérthetetlenül lehetséges.

Ha a tanulási sikeresség monoton/konstans értékké válik (pl. értékelési politikák eredményeként), akkor a sikert pl. a konstans értékelés (vö. megfelelt) megszerzéséhez szükséges idővel lehet helyettesíteni…

# A probléma

A karakterlimit megváltoztatásának problémája kapcsán **két vizsgálatsort** végeztünk.

* Az első esetben 12453 rekordnyi megjegyzést elemeztünk 2016.11.10-2017.02.24 közötti időszakban. A második adathalmaz – az első kiegészítéseként – 2017.07.06-ig tartott és 17228 rekordnyi megjegyzést dolgoztunk fel benne.

A vizsgált e-szemináriumokban többféle írásos megnyilvánulás található. Tekintsük át ezt a folyamatot! A hallgató kap egy kérdést, amelyre be kell küldenie egy megoldást. Erre az oktató választ ír. Ehhez a válaszhoz a többi hallgató megjegyzéseket fűz hozzá a feladaton belüli aktivitás keretein belül. A vizsgálaton belül a folyamatnak csak a második és harmadik lépését elemeztük: vö.

(1. HALLGATÓI MEGOLDÁS)→ 2. OKTATÓI VÁLASZ → 3. HALLGATÓI MEGJEGYZÉSEK

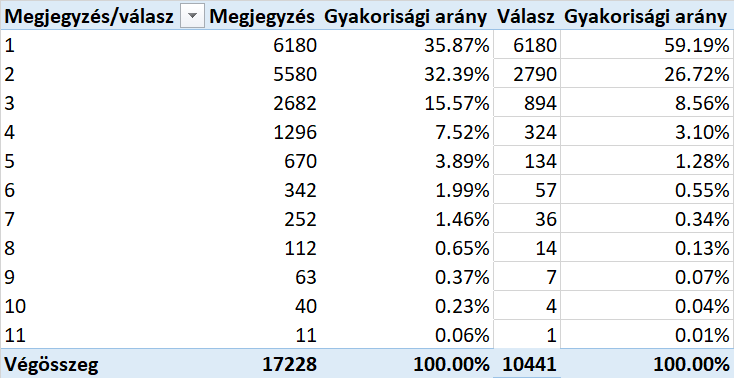
A **megjegyzés-orientált nézetben** csak azok a válaszokat elemeztük, melyekhez hallgatók megjegyzést fűztek. Vagyis a háttérben létezik egy teljes körű válasz-adatbázis is, mely azon válaszokat is tartalmazza, melyekre egyáltalán nem érkeztek hallgatói megjegyzések. Egyes válaszok több megjegyzést is életre hív(hat)tak. [[2]](#footnote-2)

Az alábbi statisztikák mutatják be a 2017.07.06-i állapot szerint az összes válasz elemzéshez való kapcsolódását:

|  |  |  |
| --- | --- | --- |
| Jelenség | Érték | Mértékegység |
| Az összes válasz-rekord száma a teljes vizsgálatban: | 23152 | db |
| Az összes megjegyzés-rekord száma a teljes vizsgálatban (ill. 2017.02.24-ig): | 17228 (8722) | db |
| Megjegyzést generáló válaszok száma a teljes vizsgálatban (ill. 2017.02.24-ig): | 10441 (7192) | db |
| Megjegyzést nem generáló válaszok száma a teljes vizsgálatban: | 6787 | db |
| Megjegyzést generáló válaszok átlagos súlyozatlan hossza (2017.02.24-ig) | 2205 | karakter |
| Megjegyzést generáló válaszok előfordulási gyakorisággal súlyozott hossza (2017.02.24-ig) | 2207 | karakter |
| Az összes válasz átlagos hossza (2017.02.24-ig) | 2201 | karakter |

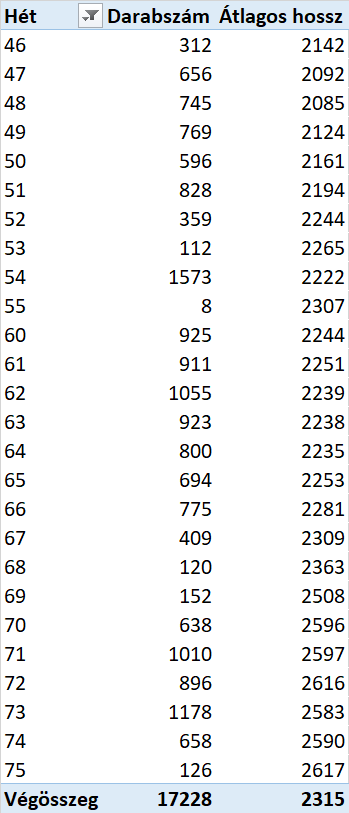
Az átlagos karakterhosszak különböző részhalmazokban látható alakulása alapján az a vélelem, miszerint a hosszabb szövegeket kevésbé olvassák el az érintettek, így ezek kevésbé válnak majd további interakciók alapjává, eleve kizárható volt. Ezen vélelem esetleges stabilitása azt jelentette volna, hogy a karakterszám-emelés még akkor is aggályos, ha a rendszerfeszültségek (l. alább) léteznek, ill. növekvők, mert a karakterszám-emelés kapcsán inkonzisztens állapotot jelentene, ha párhuzamos célváltozók (pl. interakció-katalizáló erő) romlást mutatnának. Más megfogalmazásban: ha antagonizmusok lettek volna tetten érhetők a célváltozók tekintetében, akkor eleve a paraméter optimumát kellett volna keresni…

1. ábra: Az egyes válaszokhoz kapcsolódó megjegyzésszámok megoszlása (forrás: saját számítások)



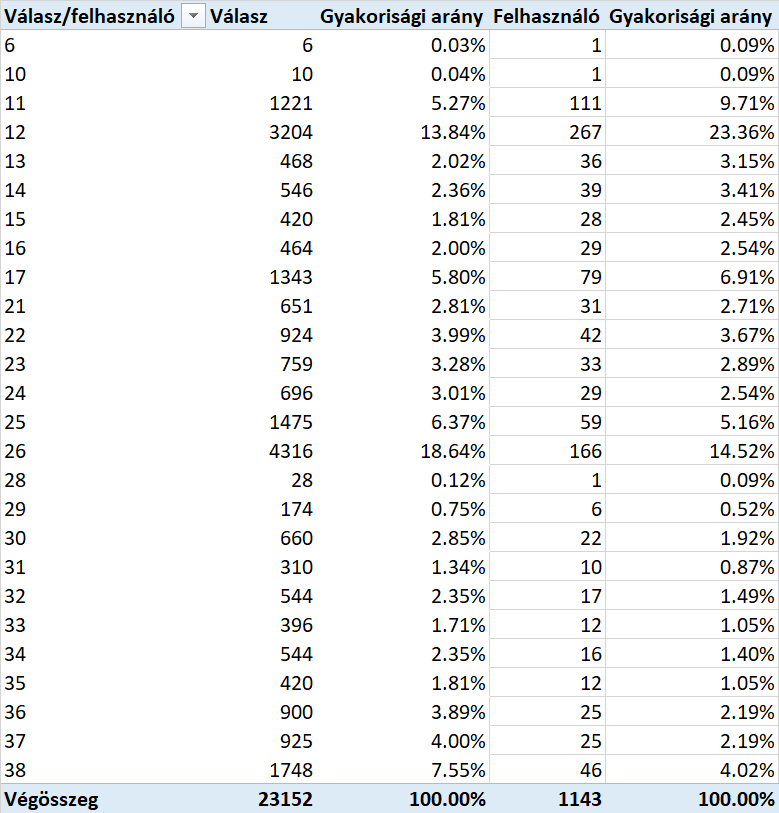
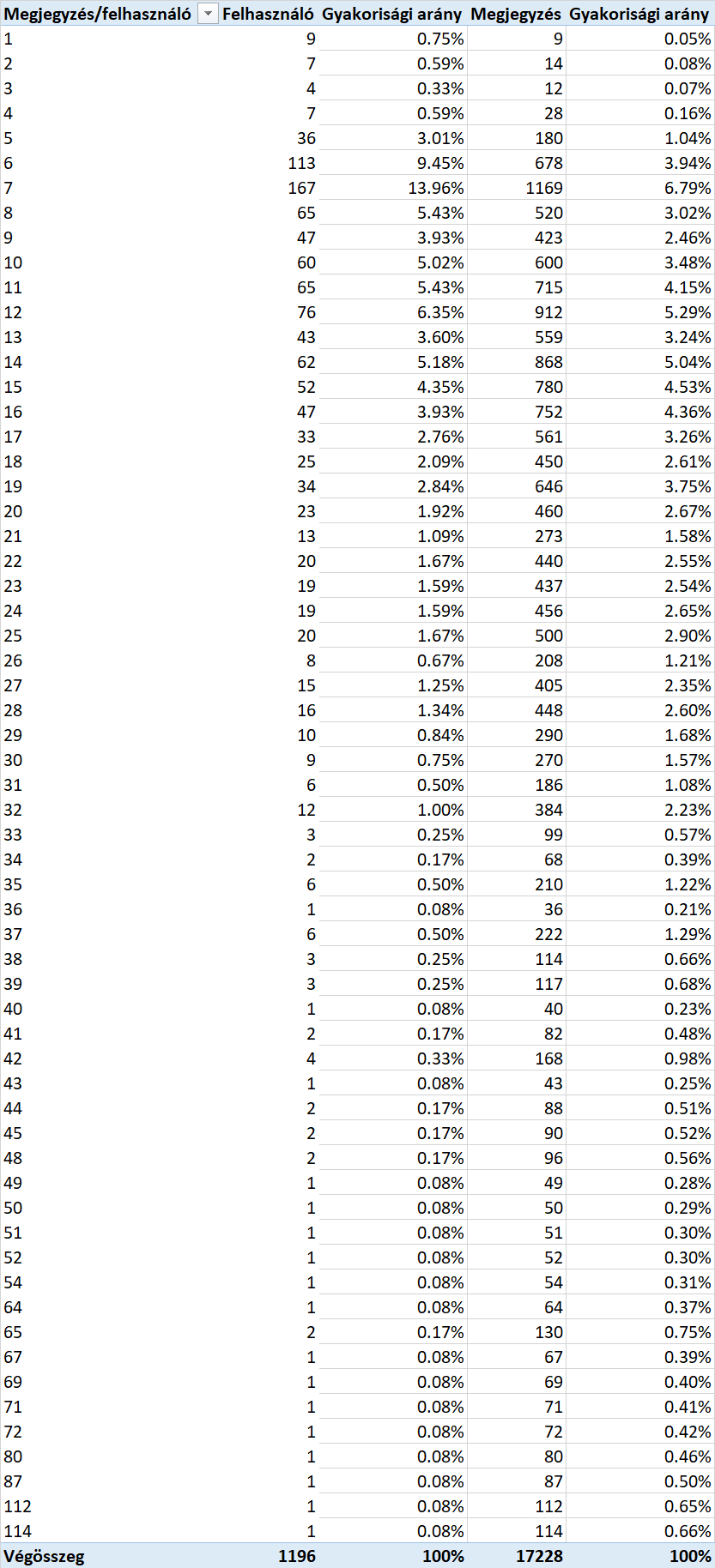
Vagyis 11 megjegyzés volt a legtöbb, ami 1 darab válaszhoz kapcsolódott (vö. 1. ábra).

1. ábra: A megjegyzést generálni képes válaszok darabszáma és az érintett válaszok átlagos hossza hetente (forrás: saját számítások)



A 46-52. hét 2016-ot jelenti, az 53. héttől 2017-ről beszélünk (vö. 2. ábra).

A rendszerben 1143 felhasználó írt válaszokat és 1196-an adtak le megjegyzéseket. Ennek kapcsán felmerülhet az Olvasóban a kérdés: vajon mi lehet a magyarázata annak, hogy többen írnak megjegyzést, mint választ? A potenciális válasz: tanárok nem írhatnak választ, csak megjegyzést, vagyis a megjegyzést írók száma maximum a tanárok számával lehet nagyobb a választ írók számánál (feltételezve, hogy akadhat olyan diák, aki bármi okból, de nem ír megjegyzést egyetlen egyet sem).

1. ábra: Válaszok (csak diákok) és megjegyzések (tanárok és diákok együtt) – (forrás: saját számítások)

A legtöbb válasz 38, a legtöbb megjegyzés 114. A tipikus válaszszámok: 26, ill. 12. A tipikus megjegyzésszám: 7 darab (vö. 3. ábra).

*Itt kell megemlíteni, hogy egy következő tanulmányban bemutatásra kerülnek azok a szövegbányászati megoldások, melyek alapján az alapvetően tanári megjegyzések sablonossága indexértékké konvertálható…*

# Modellek

A tény-alapú döntéstámogatás lényege, hogy adott döntési helyzetben a rendelkezésre álló adatok alapján tisztán algoritmusokra alapozva lehessen döntést generálni. Jelen esetben az egyébiránt intuitívan kezelt döntési helyzet nem más, mint az, hogy

* szükséges-e a felhasználók által kért/javasolt karakterlimitet a feladatokra írható válaszhossz tekintetében megemelni?
* s ha szükséges, akkor mennyivel érdemes módosítani a limitet?

## Idősoros elemzések

Az idősoros elemzések lényege, hogy a mindenkor rendelkezésre álló idősoros adatok (heti válaszhossz-átlagok) alapján meg kell állapítani pl. az alábbi **karakterisztikákat** (K1, K2, K3, K4):

1. mi az idősor (lineáris) trendje, vagyis mennyi a heti átlagos karakter-felhasználási változás?
2. mennyi a lineáris trendhez tartozó R2 érték?
3. milyen alakú (maximummal vagy minimummal rendelkező-e) a másodfokú trend, vagyis mekkora és milyen előjelű a négyzetes tag koefficiense?
4. mennyi a másodfokú egyenlettel jellemezhető trendhez tartozó R2 érték?

A fenti 4 kérdéshez/karakterisztikához kapcsolódó **értelmező szabályok**, ahol a szabályok felállítása által lehet megfelelni annak az elvárásnak, miszerint tudás az, ami forráskódba átírható, minden más emberi képesség (pl. az intuíció) művészet. **Az idősor-elemzés karakterisztikáinak értelmezése**:

* **minél** nagyobb pozitív szám a lineáris trend értéke [K1], **annál** *nagyobb a rendszerben lévő feszültség* a limit előbb-utóbb való elérését, meghaladási kényszerét illetően
* **minél** kisebb a távolság a lineáris trend által az utolsó ismert időpontra vonatkozó becsült értéke [K1] és a limit között, **annál** *nagyobb a rendszerben lévő feszültség* a limit előbb-utóbb való elérését, meghaladási kényszerét illetően
* **minél** nagyobb a másodfokú modell négyzetes tagjának koefficiense [K3], **annál** *nagyobb a rendszerben lévő feszültség* a limit előbb-utóbb való elérését, meghaladási kényszerét illetően (ahol a negatív koefficiens jelenléte alapvetően telítési tünetek előjele)
* **minél** nagyobb a másodfokú modell R2 értéke [K4] a lineáris modell R2 értékéhez [K2] képest, **annál** *nagyobb a rendszerben lévő feszültség* a limit előbb-utóbb való elérését, meghaladási kényszerét illetően, amennyiben az R2 értékek közötti különbség mértékéhez a jobbnak vélt modell legnagyobb hatványkitevőre emelt koefficiensének előjelét rendeljük…

Természetesen igaz, hogy a fenti karakterisztikák és értelmező szabályok halmaza **tetszőlegesen** **továbbbővíthető**. Amennyiben már a karakterisztikák feltárását is automatizálni kívánja egy rendszer, úgy tetszőleges nyers vagy származtatott jelenségek közötti korrelációkkal korlátlan mennyiségben lehet látszólagos látszat-korrelációs kapcsolatokra, irányokra szert tenni. Mivel vélelmezhetően minden, mindennel összefügg, így a robotizált attribútum-választás kockázata legfeljebb annyi, hogy nem az emberi gondolkodásmódok számára leginkább racionális (pl. legrövidebb hatásláncú) kapcsolatok kerülnek kiválasztásra. Éles döntési helyzetekben a karakterisztikák és szabályok intuitív kialakítása a mindenkori döntéshozók és elemzők **felelőssége**.

Ha több tényezős szabályrendszer áll egy döntés alapjaként rendelkezésre, akkor felmerül a „**lehet-e minden előálló konstelláció másként egyforma?**” hipotézis vizsgálatának kényszere. A „minden-másként-egyforma” antidiszkriminatív elvet pl. hasonlóságelemzéssel lehet matematikai értelemben egyetlen egy döntési potenciál-skálára transzformálni, ahol a döntési potenciál értéke minél nagyobb, annál inkább racionális a limithatár módosításának (a használt attribútumok irányától függően: emelésének) megfontolása.

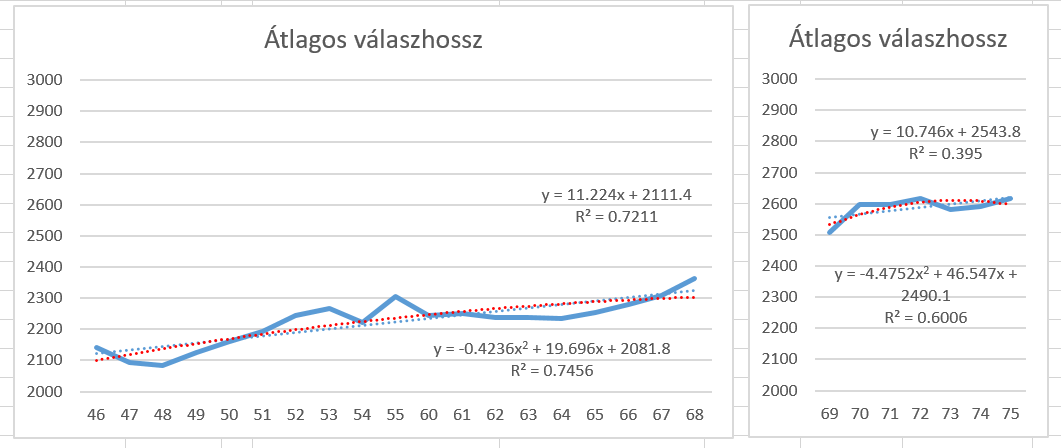
Amennyiben a **potenciálérték** minden konstellációra azonos (azaz minden naptári hét másként egyforma lenne), akkor nincs értelmezhető döntési helyzet. Mivel a konstellációk hétről hétre állnak elő (s az idősoros elemzések eleve elvárják egy minimális elemszámú tapasztalat meglétét), így a döntési potenciál-index önmagában is idősoros jelenséggé válik, ahol az index trendjére vonatkozóan is további hermeneutikák alkothatók (feltételezve, hogy a hasonlóságelemzéshez szükséges minimális elemszámú konstelláció nem folytonosan azonos potenciál-értéket mutat). Például minden olyan esetben, amikor az új konstelláció értéke a legnagyobb potenciált testesíti meg, az idősoros vizsgálatokat kiegészítő szimulációs modellek kialakítása is szükséges, ill. ha a potenciál-index konstellációról konstellációra nő, akkor minden egyes aktuális konstelláció kapcsán további szimulációs modellek építése szükséges a több elemű (konzisztencia-vezérelt) döntéstámogatás megalapozása érdekében. ***Már itt felvethető, hogy egy olyan döntési kényszer-index, mely idősoros alapadatokból kerül levezetésre és önmagában is idősoros adatokat eredményez, önmaga inputjává képes válni annak vizsgálatát lehetővé téve, hogy a döntési kényszer-index lefutásából következik-e döntési kényszer. S ez az önmaga inputjává váló modellezés a konzisztencia-maximalizálás speciális eszköze.***

A karakterisztikák és az értelmező szabályok úgy a **teljes**, mint minden racionálisan képzett (vö. jelen esetben: megjegyzés-generálásra képes válasz-) **részhalmazra** is érvényesek. A **konzisztencia-vizsgálat** alapszabálya: minél több szempontból (pl. minél több részhalmazra) vonatkozóan azonos a döntési konklúzió, annál erősebb (legitimebb/validabb/robosztusabb) egy döntés – feltételezve, hogy az egyes konzisztencia-alakzatokra nem válik igazzá a „minden-másként-egyforma” elv…

### Az elemi szál

Az alábbiakban a fenti robotizálásra alkalmas komplex értelmezési térből elsőként egyetlen egy pillanatfelvételre vonatkozó értelmezési erőteret ragadunk ki, s ezt tekintjük az elemzési sorozatok alapelemének:

Karakterisztikák a megjegyzés generálására alkalmas részhalmazokra vonatkozóan: hatáskésleltetéssel – vagyis a 2500-as korábbi limit alatti és feletti részhalmazok szerinti bontásban, ahol a hatáskésleltetés a 2500-as határ átlépésének tényleges döntéshez (67. hét végéhez) képesti bekövetkezését jelenti, ami itt és most 1 hét (vö. 67. vs. 68. hét). A bontás értelemszerűen a döntés előtti (balra) és utáni (jobbra látható) állapotokra utal (vö. 4. ábra).





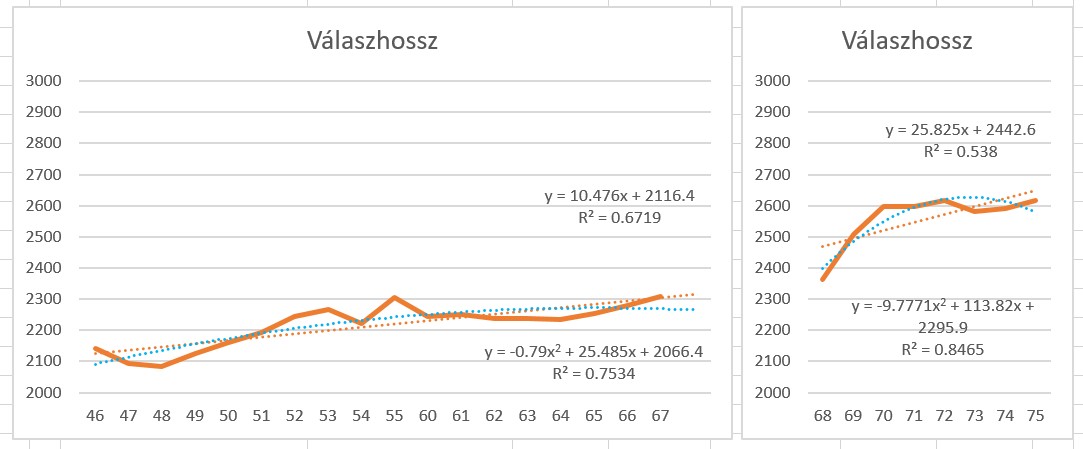
1. ábra: Átlagos válaszhosszak idősora a döntés előtt és után (forrás: saját számítások)

Az átlagos válaszhossz rendszeren belüli viselkedése (mielőtt döntés született volna a karakter-limit megemeléséről - 46-68. hét) folyamatos növekedést mutat, kb. 11.224 karakter/hét emelkedési rátával, ami önmagában is valószínűsíti a 2500-as határ belátható időn belüli elérését / egyre több érintettben a látens feszültség növekedését.

A limitemelés után (69-75. hét) az emelkedési szint csökken, egy határozott azonnali korrekció után már csak átlagosan 10.746 karakter/hét az emelkedés volumene.

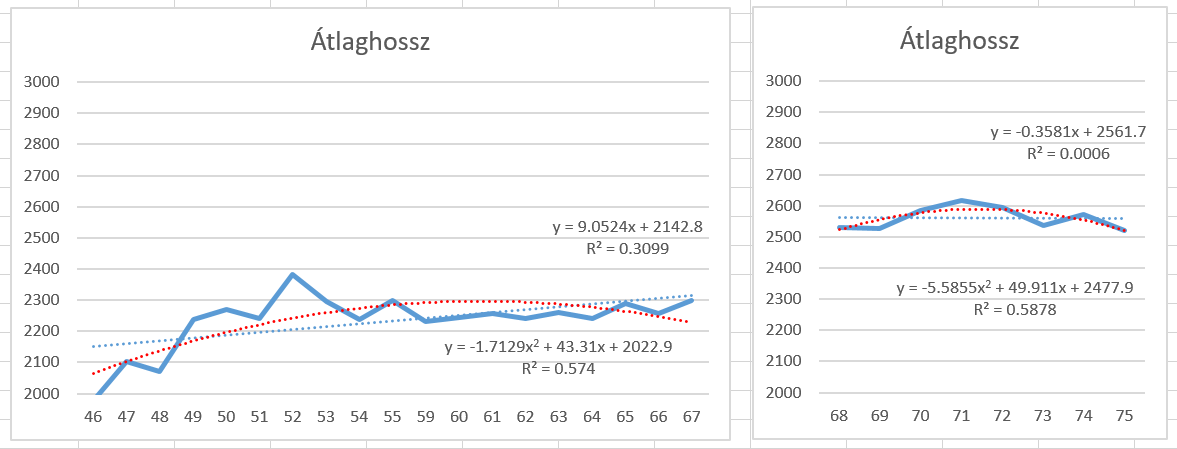
A trendvonal csökkenő emelkedése a döntés után a rendszerből helyesen, racionálisan kivett feszültségcsökkenés nyoma.

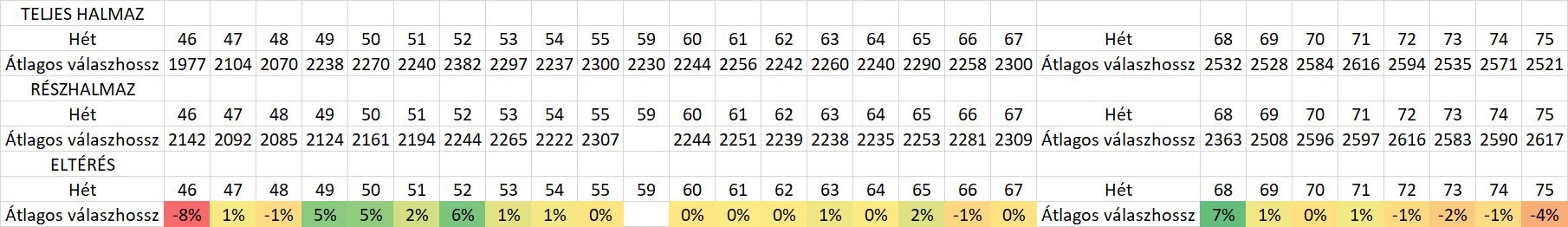
1. ábra: Karakterisztikák a megjegyzés generálására alkalmas részhalmazokra vonatkozóan (hatáskésleltetés nélkül a döntés időpontjához kötődően, azaz 46-67, ill. 68-75-ös tagolásban szemben a késleltetés esetén alkalmazott 46-68, ill. 69-75-ös tagolással) – (forrás: saját számítások)



Mint látható (5. ábra), a késleltetési hatás figyelembe vétele alig hat az ábrapárok bal oldali rétegében (11.224 vs. 10.476), míg a jobb oldalon a 10.746-os emelkedés a 25.825-ös értékkel áll szemben.

1. ábra: Karakterisztikák a teljes válaszhalmazokra vonatkozóan döntés előtti és döntés utáni bontásban (forrás: saját számítások)

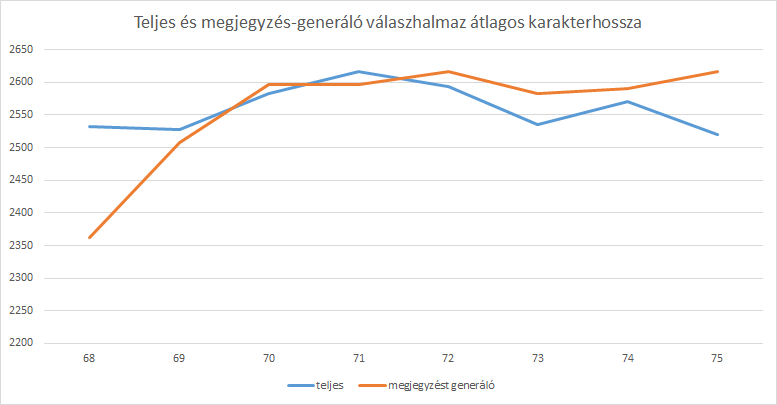




A teljes válaszhalmaz és a megjegyzés-generálásra alkalmas részhalmaz között a **legjelentősebb** (abszolút) különbség (ellentétes előjellel: vö. -8% vs. +7%) a vizsgálat kezdetén és a limitmódosítás kapcsán alakult ki, vagyis egy új paraméterekkel induló rendszer kezdetén, ami az egy hetes hatáskésleltetéssel is összefüggésben állhat.

A limitmódosítás előtt (+22 vs. -10) a teljes halmazra jellemző átlagos hossz szemmel láthatóan (is) nagyobb, míg a limit-módosítás utáni helyzetben a teljes halmaz már csak jelképesen hat a karakterszám növelésére. Következésképpen a limitemelés után (+9 vs. -8) csökkent a rendszerben a feszültség.

A megjegyzés-generáló válaszok a késleltetést követően jobban kihasználják a limitemelés kínálta lehetőséget (vö. 7. ábra):



1. ábra: Teljes és megjegyzés-generáló válaszhalmaz átlagos karakterhossza (forrás: saját számítások)

Amennyiben az összes válasz hetenkénti átlagos hosszát vizsgáljuk, akkor a döntés előtti és utáni időszakok esetén világosabban látható a lineáris trendvonalak meredekségeinek különbsége (vö. 9.0524 > -0.3581), s még világosabbá teszi a telítési tüneteket a másodfokú függvények alakja és illeszkedési pontossága (R2).

A telítési tünetek döntés előtti jelenléte felveti az elhamarkodott döntés kockázatát is (vö. komplexebb, filozofikusabb célok – pl. a közszolgálati kommunikáció rövidítési, minőségibbé tétele, nyelvi sajátosságainak köznyelvhez közelítése).

A megjegyzések alapjává váló és arra közvetlenül nem ható válaszok karakterisztikáinak ilyen jelentős eltérése további (szövegbányászati) vizsgálatok lehetőségét veti fel, ahol arra a kérdéskörre lehet válaszokat keresni, miszerint:

* mely statisztikai, ill. jelentéstartalmi rétegekben különbözik a teljes válaszhalmaz és a megjegyzéseket katalizáló válaszhalmaz,
* azaz mi lehet az oka annak, hogy
  + a megjegyzések alapjává váló válaszok esetén a limit-módosítása hatása késve jelenik meg?
  + a telítési tünetek nem ugyanolyan arányban érhetők tetten? (vö. pl. R2=0.7211~0.7456, ill. R2=0.3099 < 0.574)

Összefoglaló jelleggel vélelmezhető az átlagos válaszhosszak idősor-variánsai alapján, hogy

* a karakterszám-limit emelése racionális volt,
* de kivárás esetén a kényszerekhez való idomulás jeleit már mutatta a rendszer,
* s a szubjektív döntést az érintettek szóbeli javaslatai is elősegíthették.

Ha valaki (pl. egy döntés-előkészítő robot) ugyanezen vizsgálati lépéseket már az 59-60. hét táján (ill. ideális esetben minden újabb válasz beérkezése után elvégezte volna) egy kibernetikus rendszerben, s ahogy történt is, nem változik meg a limit értéke a következő 2 hónapban, akkor a limitet helyben hagyó döntés a fenti 3 grafikon alapján vizuális kontroll mellett is racionális lett volna – hiszen a telítési tüneteknek jeleit a limit-stabilizáció melletti érvként lehetett volna felfogni. A karakterszám-limit emelése után masszív érvnek látszik ennek lépcsőzetes ugrást jelentő hatása, de a megjegyzésekre (vagyis a blended learning keretében zajló e-szeminárium által katalizálni tervezett interakciókra) valóban ható válaszok esetén látható késleltetés továbbra is felveti a limit-tartás esetleges racionalitását. Ahogy az egyedi konstellációk értékelését figyelve megállapítható, egyedi konstellációkról épp úgy nem lehet/illik/érdemes értékítéletet mondani és/vagy elvárni, mint adott termék ár/teljesítmény-arányáról sok versenyzős közbeszerzés nélkül…

A fenti sok szempontú értelmezés üzenete: a robotizálás irányába ható döntéshozatali mechanizmusok esetén nem mindenkor engedhető meg a döntéssel befolyásolni engedett keretfeltételek gyakori és ellentétes irányú finomhangolása az érintettek stabilitás iránti igényét alapul véve (vö. jegybanki alapkamat-pályák taktikai és stratégiai alakítása). Azonban és amennyiben egy kibernetikus rendszerben az érintett felhasználók szocializálása megtörténik arra, hogy a vezérlések egyensúlykeresők (vö. ABS az autókban), akkor ennek tűrése nem jelenthet érdemi kockázatot. Hibrid-elvű vezérlések esetén az adaptáció minimális számú lépéssel való közelítése is racionális döntési stratégiai lehet.

Summa summarum: a robotok, hasonlóképpen, mint az emberek képesek a célstruktúrák komplexitásának megfelelő válaszreakciók generálására! Dashboard-jellegű döntéstámogatás kiérlelt elemzési lépéssorok és az ezekre felkészített döntéshozók esetén racionálisak. Addig, amíg a döntéshozó élni kíván elemzési szabadságával (pl. OLAP-támogatás szinte korlátlan kombinatorikai térben), addig minden dashboard-kísérlet sokkal inkább az OLAP-funkcionalitás ergonómiai/önkényes korlátozásaként hathat. (Erről szól a cikksorozat 3. eleme…)

### Konzisztencia-alakzatok

Az elemi szál a maga statikus voltában két irányba is továbbszőhet:

* a **genetikai potenciál** elemzése felé, ahol a cél a rendszer abszolút outputjának statikus feltárása,
* a **döntési kényszer-index** felé, ahol a cél az alapelemként definiált lépések hétről hétre való megismétléséből származó eredmények értelmezése.

Mivel ezek a kiegészítések egyben önálló modellezési gondolatok, önálló alfejezetként kerülnek bemutatásra ezen fejezeten belül:

## Genetikai potenciál elemzések

Minden rendszer esetén, így a tanulási folyamatok kapcsán is létezik a **termelési függvény** fogalma, vagyis az inputok és outputok vélelmezhető összefüggésének jelensége. Egy rendszer (pl. idősoros és/vagy térbeli) állapotvariánsait megfigyelve bárki számára látható, hogy adott inputok együttállását valamilyen outputok követnek. Jelen esetben: a naptári hetek mentén az oktatási rendszer quasi tetszőleges paraméterei (pl. felhasználói létszám, tárgyak száma, felhasználók nemének aránya, stb.) mellett pl. az átlagos megjegyzés-hossz, vagy éppen az átlagos válaszhossz valamilyen logikát követve (esetlegesen úm. véletlenszerűen) változik. A termelési függvények feladata annak leírása: milyen input-variánsokhoz melyik a legvalószínűbb output-szint. Vagyis egy oktatási/tanulási jelenségkörben azt lehet a genetikai potenciál interdiszciplináris áthallásain keresztül vizsgálni, milyen hosszú szövegek keletkezhetnek maximum, ha a szöveg hosszúságát növelő tényezők „ideális” együttállása bekövetkezik? Egy ilyen becslés egyrészt azért fontos, mert ezzel lehet érzékeltetni milyen messze van egy adott rendszerparaméter (vö. 2500 karakteres limit) a racionálisan vélelmezhető rendszerhatárértéktől – s egyben a modellek minőségbiztosítását is szolgálja (lévén egy olyan modell, mely az eddigi maximális output-értéket sem képes leképezni, kockázatok forrása).

A genetikai potenciál egy rendszer esetében az a maximális output-érték, mely az ún. ideális input-kombináció(k) esetén elvárható. Az operációkutatási szakkönyvek például beszélnek arról a jelenségről, vajon mennyi lehet egy adott növény genetikai potenciálja, ha ennek minden környezeti igényét (tápanyag, víz, energia, stb.) maximálisan kielégítjük? Ez utóbbi példa világosan rámutat arra, hogy a genetikai potenciálbecslés elsődlegesen akkor lehetséges, ha az egyes inputok és az éppen vizsgált output közötti kapcsolatnak van alapvetően valamilyen iránya: pl. minél több tápanyagot, vizet, energiát (pl. napfényt) kap egy növény, annál nagyobb terméseredményre lesz képes. A példa ereje ismét csak alkalmas arra, hogy egyben a genetikai potenciál értelmezésének kockázatát is fel lehessen vetni: általában véve igaznak tűnik, hogy a túl sok semmiből sem lehet ideális (pl. árvíz, aszály, forróság, méter magas műtrágya-réteg a termőtalajon = Karthágó). „Szerencsére” a valós rendszer-állapotvariánsok ritkán érintik ezen szélsőséges állapotokat, így a valós konstellációk eredőjeként a rendszer maximális kimenetét (genetikai potenciálját) illetően racionális becslések vezethetők le – pl. hasonlóságelemzéssel, mely olyan lépcsős függvényeket vezet le LP (pl. lineáris programozási) keretek között, melyek tanulási potenciálja (vagyis az ismert tények alapján az inputok alapján becsült és a ténylegesen ismert outputértékek közötti pl. korreláció értéke) magas és ez a tanulási sikeresség nem feltétlenül törik meg akkor sem, ha ismeretlen állapotvariánsok következményeire vonatkozó output-becslések kerülnek utólag szembeállításra a tényleges bekövetkezésekkel.

A genetikai potenciál becslésére a lépcsős függvények azért adnak triviálisan lehetőséget, mert a rendszer bementi oldalán ismert maximális értékek (első helyet leíró rangsorszámok) minden input-változó esetén való együttállása a kombinatorikai tér szerves részeként értelmeződik külön modellezési feladatok felvállalása nélkül is. A genetikai potenciál és a hasonlóságelemzések kapcsolata esetében ki kell emelni még a **konszolidált genetikai potenciál** fogalmát, mely azt jelenti, hogy a potenciálszámításba csak olyan input-szint (csereérték) tartozhat bele egy-egy változó kapcsán, mely esetében az egyedi értékelés veszélye minimális, azaz mely több mint egyszer fordul elő, (ill. szervesen következik a több lépcső értelmezéséből).

A fenti elméleti alapokra támaszkodva a genetikai potenciálbecslés egy tanulási folyamat esetén is értelmezhető és elvégezhető. A várható eredmények kapcsán elmondható általánosságban véve előre, hogy

* egy genetikai potenciál lehet magasabb, mint a tanulási mintában eddig látott legmagasabb output érték – vagyis a megtanult rendszer vélelmezni engedi, hogy egyes input-konstellációk még további outputnövekedést okozhatnak:
  1. a genetikai potenciál lehet úm. racionálisan nagyobb az ismert rendszer-maximumhoz képest,
  2. s lehet irracionális (pl. nagyságrendekkel eltérő, ill. más egyéb, a tanulási mintából nem, de az általános rendszerkarakterisztikákból, törvényszerűségekből ismert határokat túllépő),
* a genetikai potenciál lehet quasi azonos a már ismert rendszer-maximummal:
  1. ebben az esetben az ismert rendszer-maximum lehet valóban az a konstelláció, ahol minden inputváltozó ideális értéket vesz fel,
  2. s lehet ettől alacsonyabb input-erőterű – vagyis alapvetően és vélelmezhetően alulbecsült,
* s végül lehet egy becsült genetikai potenciál kisebb, mint az ismert rendszer-maximum, ami a tanulási folyamat:
  1. technikai és/vagy
  2. rendszerszintű (pl. az inputok és az outputok között véletlenszerűséget sugalló) zavarai utalhat.

Az alábbiakban következzenek az intuitív döntés nyomán előállt limitemelés előtti és utáni elemzések paraméterei és eredményei:

### Döntés előtt

Az embereket érintő komplex rendszerekben klasszikus ok-okozati összefüggésekről már teoretikusan sem lehet feltétlenül beszélni (vö. Pygmalion-effektusok lehetősége). Ebből kifolyólag inputváltozóként egy rendszer értelmezésekor minden változó egyformán racionális, ill. ezek bármilyen véletlenszerű kivonta is éppen annyira, mint a teljes (quasi végtelen) halmaz, vagy bármely más részhalmaz. A változók száma pl. azért végtelen elvileg, mert bármely természetes (triviálisan mérhető) változóhalmaz elemei között (pl. felhasználók száma, tantárgyak száma) származtatással tetszőlegesen sokféle új változó alkotható (pl. egy tantárgyra jutó felhasználószám). Nyers tehát egy attribútum, ha quasi csak le kell kérdezni a log-ok adatbázisából. Míg származtatásról akkor beszélünk, ha érdemi transzformációk (pl. méret-függetlenítés) kerülnek végrehajtásra nyers/származtatott adatokon).

A válaszok potenciális hosszának becsléséhez az alábbi változók kerültek quasi véletlenszerűen kiválasztásra (**zárójelben** az attribútumok jellemzésével – ahol a méret-függetlenség azt jelenti, hogy maga a mutatószám alkalmas kisebb és nagyobb jelenségek relativizáló leírására: pl. Kína és Luxemburg esetén is lehet beszélni népsűrűségről az országok méretei közötti nagyságrendi különbségek ellenére):

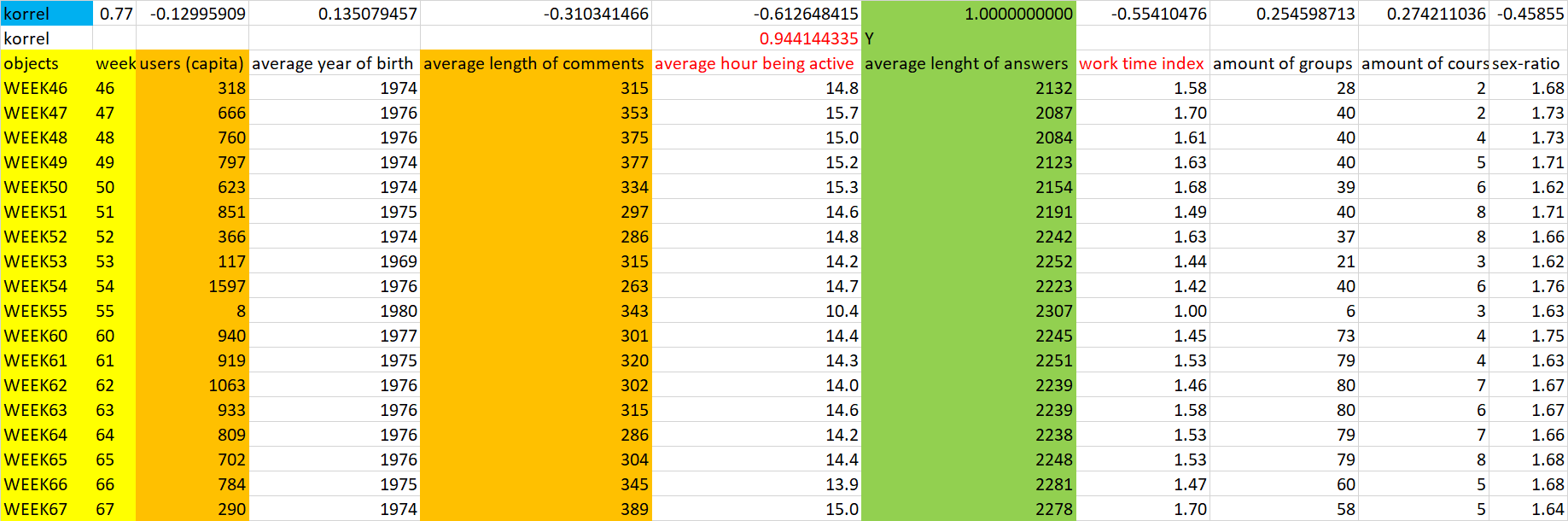
* **(1)** felhasználók száma az adott héten az e-learning rendszerben az adott héten megjegyzést író felhasználókra vonatkozóan (nem méret-független, quasi nyers attribútum)
* **(0)** a felhasználók átlagos életkora az adott héten megjegyzést író felhasználókra vonatkozóan (méret-független, származtatott attribútum)
* **(1)** a napon belüli átlagos óraszám, amikor a válaszok által generált megjegyzések születnek (vö. a válaszok vélelmezett olvasásának átlagos óraszáma 0-24 között) az adott héten megjegyzést író felhasználókra vonatkozóan (méret-független, származtatott attribútum)
* **(1)** átlagos munkaidő-valószínűségi index (ahol a munkaidőn belüliség és a munkaidőn kívüliség állapotát leíró számértékek átlagát képezzük – „1” = munkaidő, „2” = szabadidő) az adott héten megjegyzést író felhasználókra vonatkozóan (méret-független, származtatott attribútum)
* **(0)** az adott héten írt megjegyzések által érintett, aktív tanuló csoportok száma (méret-függő, nyers attribútum)
* **(0)** az adott héten írt megjegyzések által érintett, aktív tantárgyak száma (méret-függő, nyers attribútum)
* **(1)** a nemek aránya az adott héten megjegyzést író tanulók esetében (ahol a nemeket leíró számkódok átlagát képezzük – „1” = férfi, „2” = nő) (méret-független, származtatott attribútum)

A felsorolt elemek előtt **zárójelben** álló kódok jelentése:

* **(1)**: fordított arányosság, azaz minél kisebb az input (X), annál nagyobb az output (Y)
* **(0)**: egyenes arányosság, azaz minél nagyobb az input (X), annál nagyobb az output (Y), ahol

az output értelemszerűen a megjegyzések generálására képes válaszok átlagos hossza (Y – méret-független, származtatott attribútum).

A kódok tehát egyben az Y = f(Xi) kapcsolatrendszer generális irányát adják meg. Az irány elsődleges jellegét a mindenkori X(i) és az Y közötti korreláció előjele adja.



1. ábra: Tanulási minta nyers-adatai (forrás: saját számítások)

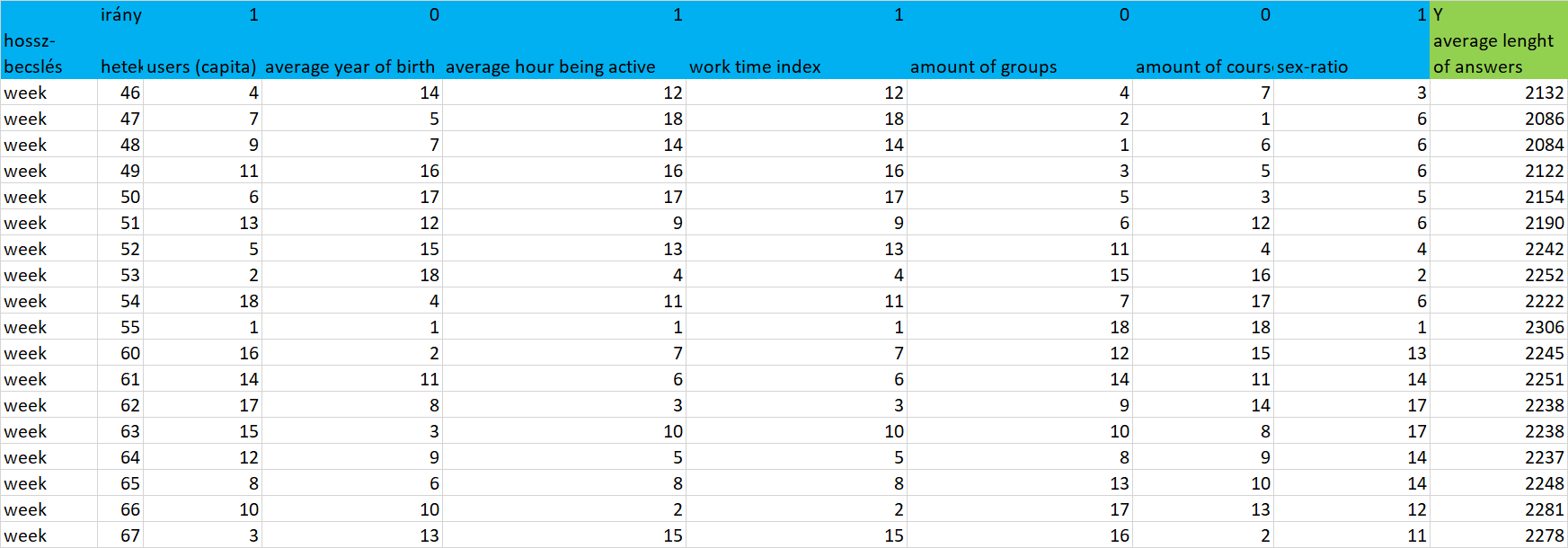
Egyes input-változók (vö. 8. ábra) között tetszőleges kapcsolat lehet (vö. korreláció = 0.944 az átlagos aktivitási óraszám és a munkaidőhöz való kötődés között, ahol értelemszerűen a kisebb átlagos óraszám fejezeti ki inkább a munkaidőt, ill. a munkaidőt jelentő „1”-es kód kisebb, mint a szabadidőt jelentő „2”-es kód).

Érdekességképpen kiemelendő, hogy az objektumok (vagyis a naptári hetek) önmagukban is 0.77-es korrelációban állnak a következmény-változóval (Y), ami a korreláció, mint karakterisztika formájában tükrözi vissza a trendvonalak alakját és paramétereit.

Itt érdemes megjegyezni, hogy az adatok forrása lehet pl. egy OLAP-szolgáltatás, mely itt meghivatkozott formájában képes a hetek, mint objektumok mentén egyenként egy-egy mutatószámot (pl. felhasználószámot, átlagos megjegyzéshosszt lekérdezni: vö. <http://miau.gau.hu/eolap/db1/2_olap_m_din.php3>)

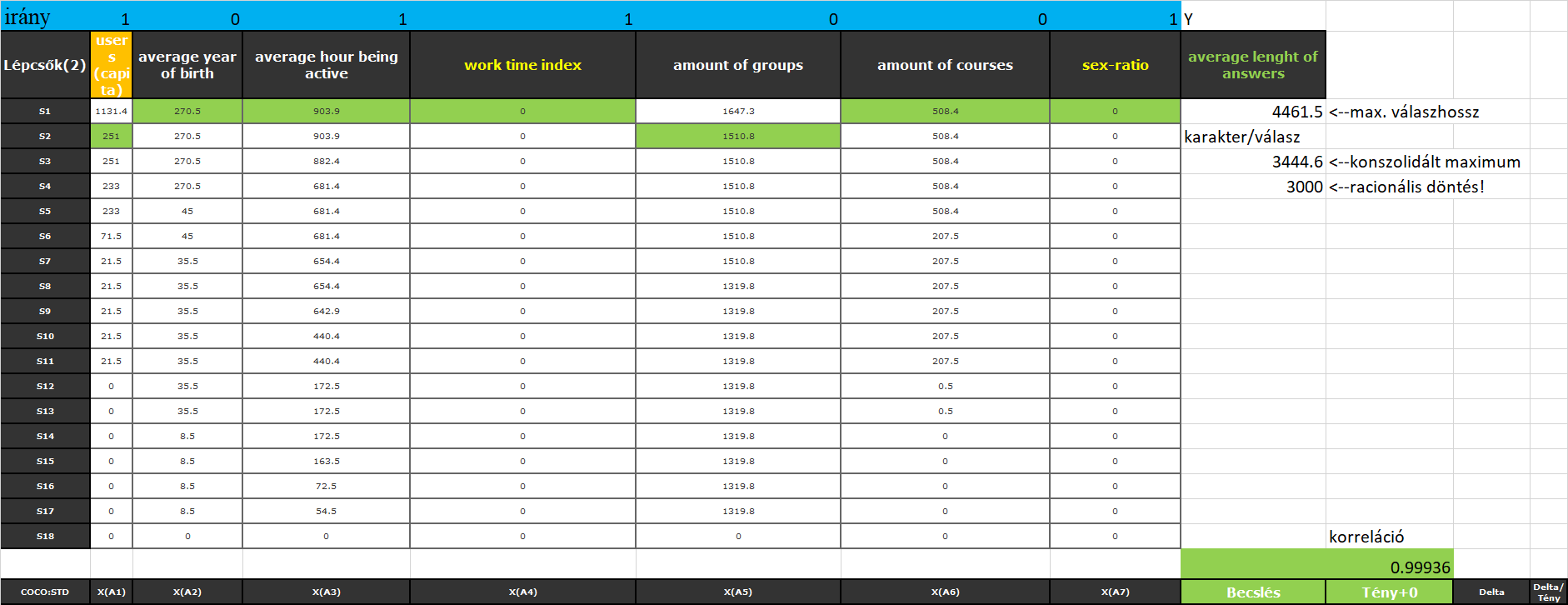
A termelési függvény becsléséhez pl. az átlagos megjegyzéshossz nem került felhasználásra, hogy a megjegyzésekhez való kötődés ne legyen még tovább hangsúlyozva. Minél több a rendelkezésre álló input (magyarázó) változók száma, annál inkább felmerül a kérdés: milyen kevés inputtal lehet milyen magas tény-becslés-korrelációt kikényszeríteni? Illetve a hasonlóan „jó” (de nem csak a korreláció által jellemzett) alternatív modellek közül melyiket illik a legjobbnak tekinteni?

1. ábra: Tanulási minta rangsor-nézete a hasonlóságelemzésben (forrás: saját számítások)

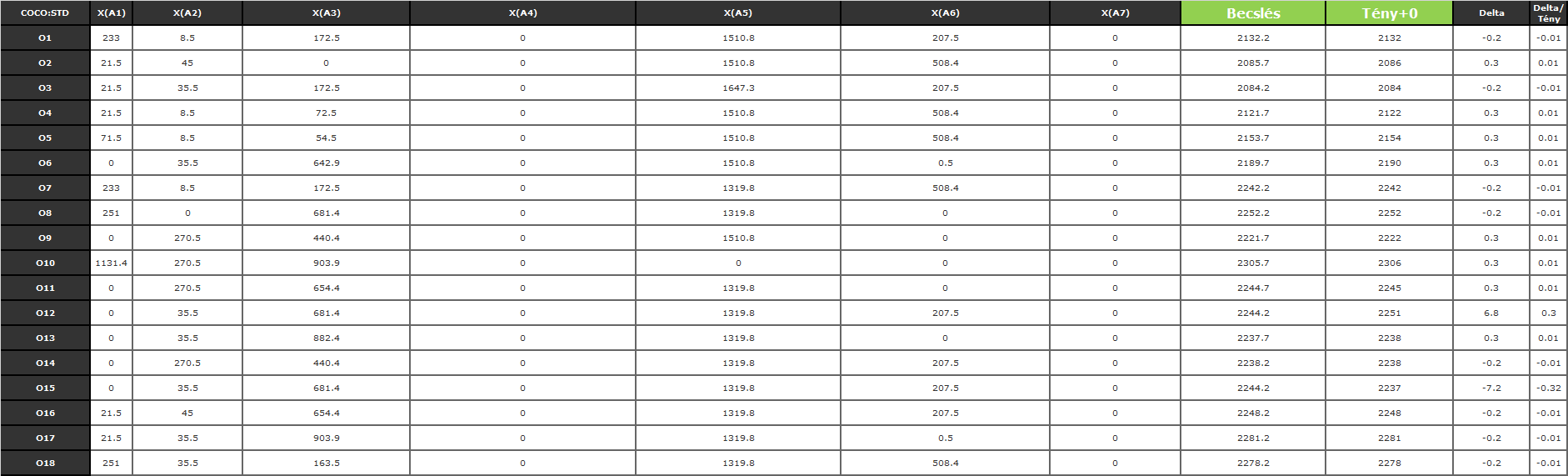


A nyers alapadatokból a hasonlóságelemzés érdekében inputváltozónként rangsorolni kell (vö. 9. ábra) az egyes objektumokat (jelen esetben a naptári heteket).

1. ábra: A hasonlóságelemzés eredménye egy lépcsős függvény (forrás: saját számítások)



A tanulási eredmény (vö. 10. ábra) magas korrelációs értéke a tényleges átlagos válaszhosszak és a becsült válaszhosszak között az alábbi tételes kapcsolatrendszerből vezethető le (vö. 11. ábra):



1. ábra: Lépcsős függvény (forrás: saját számítások)

A rendszer genetikai potenciálja konszolidálatlanul az első lépcsőszint (S1) összege lenne: 4461.5 karakter

A konszolidált genetikai potenciálba csak azok a lépcsők számíthatnak bele, melyek legalább kétszer előfordulnak (vö. zöld cellák az S1 és S2 sorokat illetően az eredménytáblában): 3444.6 karakter

Az ismert output-maximum: 2306 karakter, ami mindenképpen 2500 alatt kellett, hogy legyen, lévén a 2500-as limitszabályozás érvényben volt.

Summa summarum:

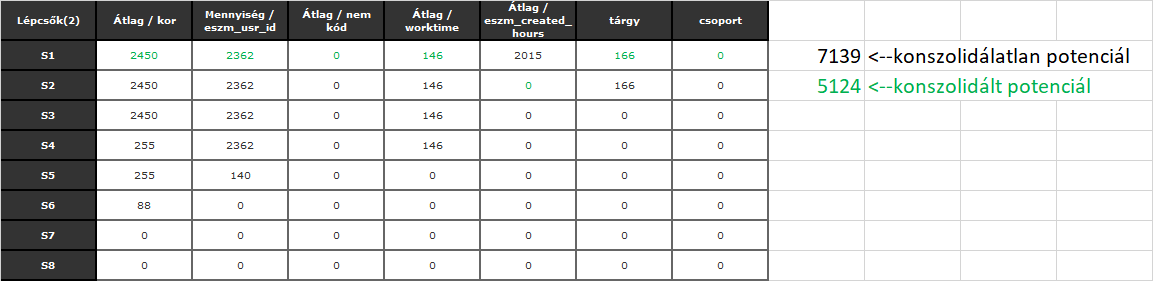
* A vizsgálható naptári hetek (O1, …, O18, azaz 46. vs. 67. hét) alatt megismert rendszer esetében a quasi véletlenszerű inputok és az átlagos válaszhossz között nagyon stabil (konszolidálatlan) modell volt építhető (korreláció = 0.999).
* A lépcsős függvény konszolidálása után a racionális output-maximum a rendszerben 3444.6, ami racionális mértékben, s egyben jelentősen meghaladja a szabályozási limit 2500-as értékét.
* Vagyis az inputoknak megfelelő irányok mentén való elmozdulással adott esetben a rendszerben szereplők késztetése a többet írni akarásra, vélelmezhető – ami az idősoros elemzésekben is vélelmezett trendek által is jelzett feszültséges forrása lehet.

A robot-szakértő a fenti egyedi esetben tehát a másodfokú egyenlettel leírt trendvonal minimális R2-értéktöbbletén túl lényegében potenciális feszültségeket lenne képes beazonosítani a rendszeren belül, vagyis egy limitemelést racionálisnak tartana.

**A limitemelés mértékét illetően: mindennemű szofisztikált modell nélkül is jól alkalmazható az intervallumfelező stratégia, ami 3444.6 és 2500 között jól kommunikálható kerekítéssel maga az új 3000-es karakterlimit!**

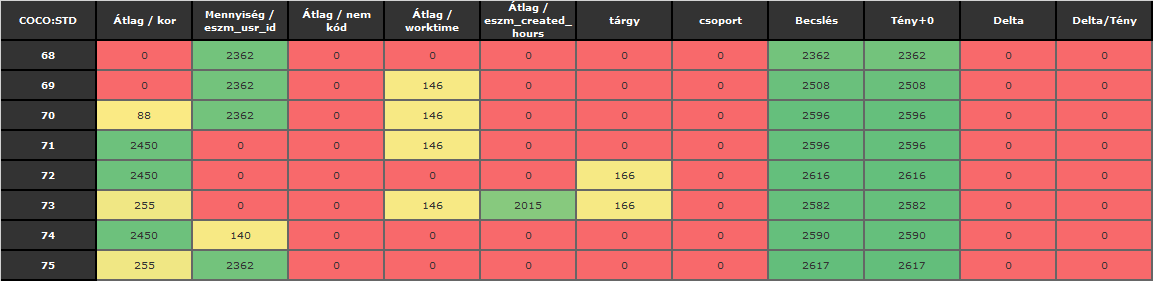
### Döntés után

A döntés utáni időszak (68-75. hét) genetikai potenciálja (vö. 12. ábra) kapcsán a korábbi 3444 karaktert kitevő konszolidált potenciál látszólag tovább nő, ugyanis az új időszak rel. kevés megfigyelése (objektuma) okán a modellben a tény-becslés eltérések mértéke nulla, vagyis a korreláció, ill. az R2 = 1.00



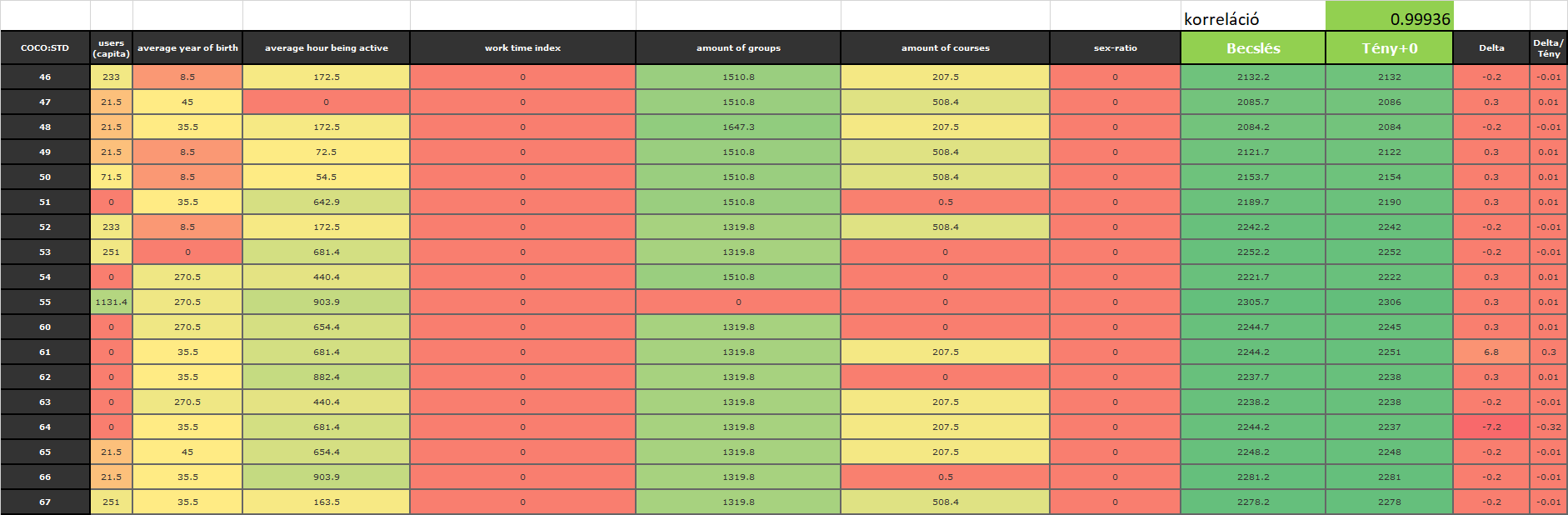
1. ábra: Genetikai potenciál becslése a döntés utáni időszakra (forrás: saját számítások)

Az alábbi ábrapár (13-14. ábra) a konszolidáció második rétegét vizualizálja, amennyiben a felső (13.) ábrán szemmel is látható, hogy a tény és a becslés soronként azonos és a zöld színárnyalat alapvetően soronként egyetlen egy komponensből reprodukálható:



1. ábra: A becslés belső szerkezete (forrás: saját számítások)

A döntés előtti időszakban a potenciálérték számítása mögötti lépcsős függvény „hő-térképe” (vö. 14. ábra) azt jelzi, hogy egy becslés több komponensből áll össze, tehát nincs szükség és mód másodlagos konszolidációra.



1. ábra: A döntés előtti modell belső szerkezete (forrás: saját számítások)

A másodlagos konszolidáció lényege, hogy nem ismer el magasabb genetikai potenciált, mint a rendszer ismert maximuma, ha a potenciál fogalma egyetlen egy rétegű marad minden ismert megfigyelés esetén. Jelen esetben tehát vélelmezhető, hogy a 3000 karakteres rendszerparaméter átlépésének kockázata quasi nulla.

## Döntési kényszer-index

A döntési kényszer-index az idősoros elemzések fentebb bemutatott egyedi rétegének általános érvényű alkalmazását jelenti. Amennyiben minden héten az idősoros elemzések elvégzésre kerülnek, akkor ezek a K1-K2-K3-K4 karakterisztikákon keresztül a kapcsolódó szabályok és döntési kényszerre ható irányok segítségével tanulási mintává konvertálható, ahol egy objektum-attribútum-mátrix (OAM) keretében az objektumok a naptári hetek, az attribútumok az irányítható szabályok, s a modell lényegi kérdése: a lehet-e minden naptári héthez azonos döntési kényszerértékét racionálisan vélelmezni kérdés…

A vizsgálatok alapjául a megjegyzéseket generálni képes válaszok részhalmaza került kiválasztásra abból kiindulva, hogy ezek azok a válaszok, mely érdemben nyomot hagynak az érintetteken, ahol a nyom pl. a rel. rövidnek tűnő kifejtettség frusztráló életérzése lehet.

Az idősoros vizsgálatokhoz ezek adatigényét és az évváltást egyszerre figyelembe véve a legrövidebb időszak a 46-52. hétre vonatkozó adatsor volt, míg a leghosszabbként értelemszerűen a 46-67-es intervallum adódott (a döntés előtt pillanatig vizsgálva a homogén rendszert).

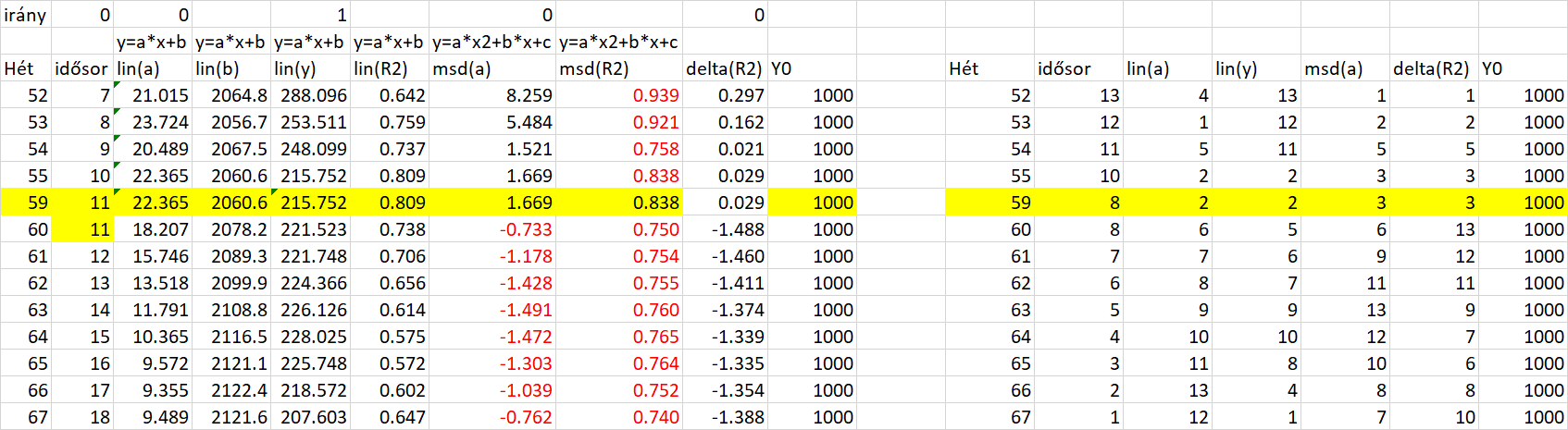
Az idősoros elemzések kapcsán ki kell térni arra a látszólagos részletkérdésre, miszerint a hetek egymásutániságában esetlegesen beálló anomáliák (vö. pl. az 59. hét adataiból nem került ki egyetlen egy megjegyzést generáló válasz sem), vagyis amikor a hetek nem egymás után következnek, miként kezelendők? Ezek az aránymódosulások ugyanis numerikusan kihatnak az idősor-elemzések eredményeire. Mivel a tényleges heti együttállások nem épülnek fogaskerékszerűen egymásra, vagyis két adattal rendelkező hét között eltelt idő mértéke nem tűnik érdemi információnak, így a feldolgozható heteknek szigorúan véve csak az egymást követő volta került felhasználásra, ezek tényleges távolsága nem.

A mellékletben a megjegyzést generáló válaszhalmaz mellett a teljes válaszhalmaz karakterisztikái is megfigyelhetők annak érdekében, hogy a fókuszok közötti hatás-különbségek érzékelhetővé váljanak.

A modellezés során a Simpson-paradoxonra (<https://hu.wikipedia.org/wiki/Simpson-paradoxon>) történő utalásként azt is figyelembe vettük, hogy az idő múlásával az idősoros elemzések paraméterei vélelmezhetően egyre több megfigyelésre támaszkodnak, ahol az idősor-elemzéshez felhasznált összes (kumulált) adatszám rangsora értelemszerűen megfelel az időmúlás rangsorának.

### A nyers és a sorszámozott tanulási minta

A nyers tanulási minta a melléklet alapján bárki által rekonstruálható (vö. 15. ábra):



1. ábra: Nyers és standardizált tanulási minta (forrás: saját számítások)

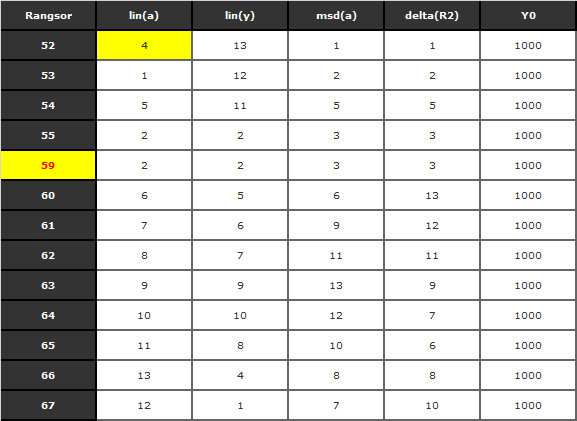
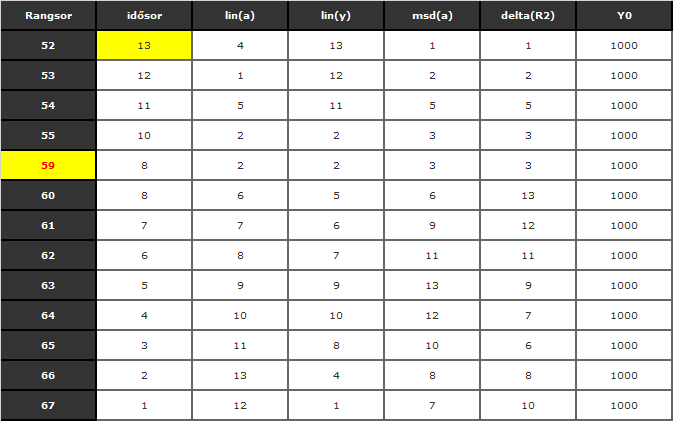
Az idősor elemszáma tehát megfelel a kumuláltan feldolgozott rekordok sorszámának. Az 59. hét nem produkált feldolgozható átlagos válaszhossz-adatot (vö. sárga jelzés). A másodfokú egyenlet R2 értékének és a lineáris egyenlet R2 értékének viszonya minden esetben a másodfokú egyenlet jobb illeszkedését mutatta, így az R2 értékek különbségéhez mindenkor a másodfokú egyenlet négyzetes tagjának koefficienséből levezetett előjel kapcsolódott.

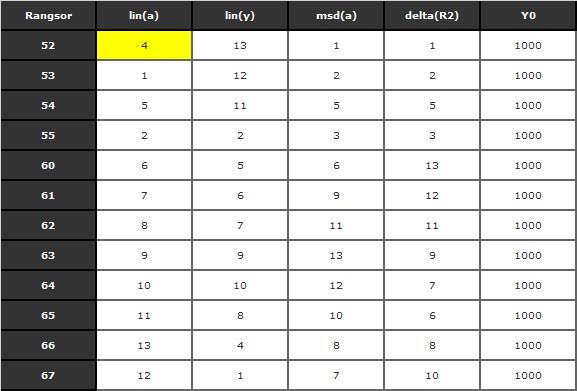
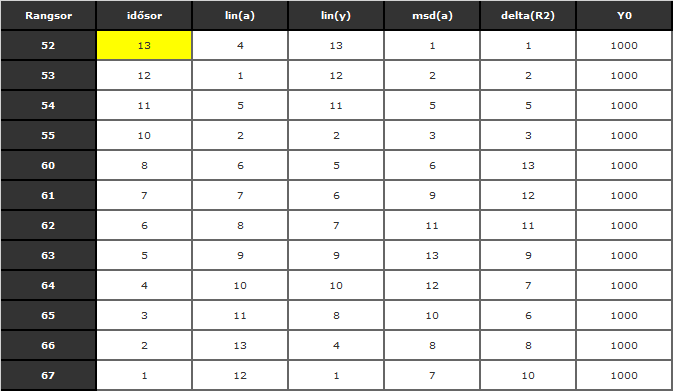
A sorszámozott tanulási minta az irányított attribútumok iránynak megfelelő sorszámvetületét tartalmazza, ahol a következmény-változó (Y0) egy konstans (1000) annak érdekében, hogy az egyes heteket antidiszkriminatív (és egyben objektivizáló, optimalizált) módon lehessen lépcsősfüggvényekből következő súlyokkal (rangsor-helyettesítési értékekkel ellátni online hasonlóságelemzés alapján (<http://miau.gau.hu/myx-free/coco/beker_y0.php>).

### Modellek

Összesen 4 (2\*2) párhuzamos modell készült annak érdekében (vö. 16. ábra), hogy ezek eltéréseit egy fajta érzékenység-vizsgálatként lehessen felfogni (vö. 59. hét hatása, ill. adatsorok mennyiségének hatása):

1. modell: 13 objektum (inkl. 59. hét), 5 attribútum (inkl. idősor hossza)
2. modell: 13 objektum (inkl. 59. hét), 4 attribútum (idősor-hossz nélkül)
3. modell: 12 objektum (59. hét nélkül), 5 attribútum (inkl. idősor hossza)
4. modell: 12 objektum (59. hét nélkül), 4 attribútum (idősor-hossz nélkül)

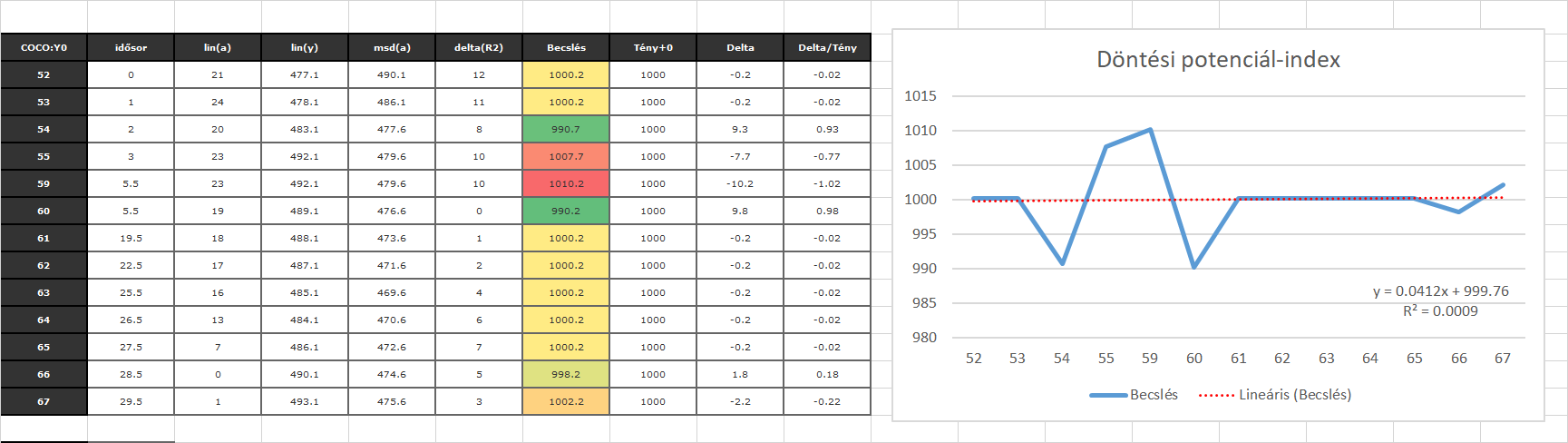




1. ábra: Modell-variánsok (forrás: saját számítások)

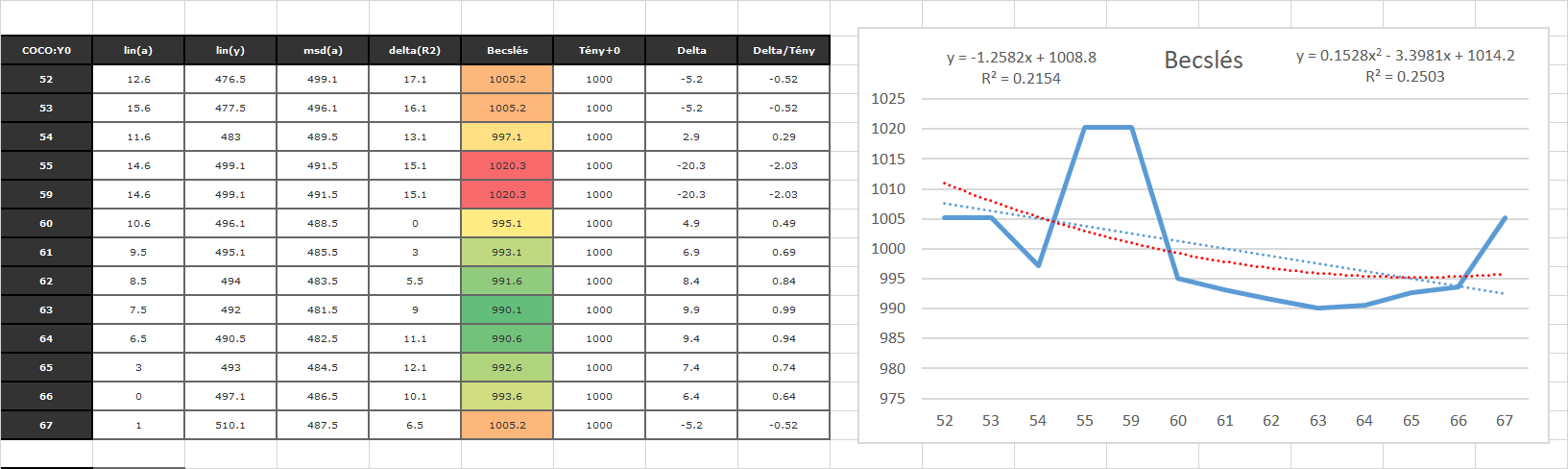
### Eredmények

Az 1. modell szerint: az 55-59. héten lett volna az első nagy döntési kényszerhelyzet (vö. 17. ábra). Majd egy második hullámban a valódi döntés előtt ismét adódni látszott egy újabb csúcs. A döntési kényszer-index lineáris trendjének meredeksége, ha gyengén is, de pozitív, azaz a rendszer a feszültség jelképesen növekvő volt. Az 55. héten rendelkezésre álló adatszám nagyságrendekkel kevesebb volt, mint az összes többi esetben.



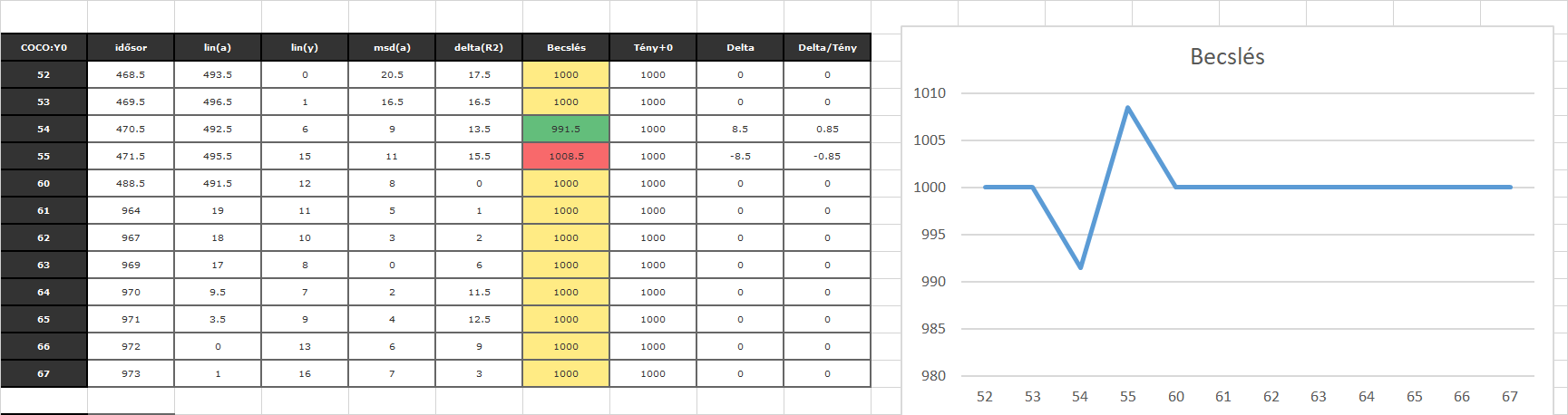
1. ábra: 1. modell (forrás: saját számítások)

A 2. modell értelmében (vö. 18. ábra): ahol is nem volt az adatmennyiségnek a többi attribútum eredőjével kölcsönható erőtere, a rendszerfeszültség általános alakulását kifejező lineáris trend csökkenő, de a másodfokú trendfüggvény magasabb R2 érték mellett masszívan a döntési kényszerhelyzet várható értékének növekedését sejtető. Az elsődleges döntési hullám ebben az esetben is az 55-59. hétre esett volna. Az elmaradó döntés után csökkenő feszültség azonban a 67. hétre ismét felerősödött.



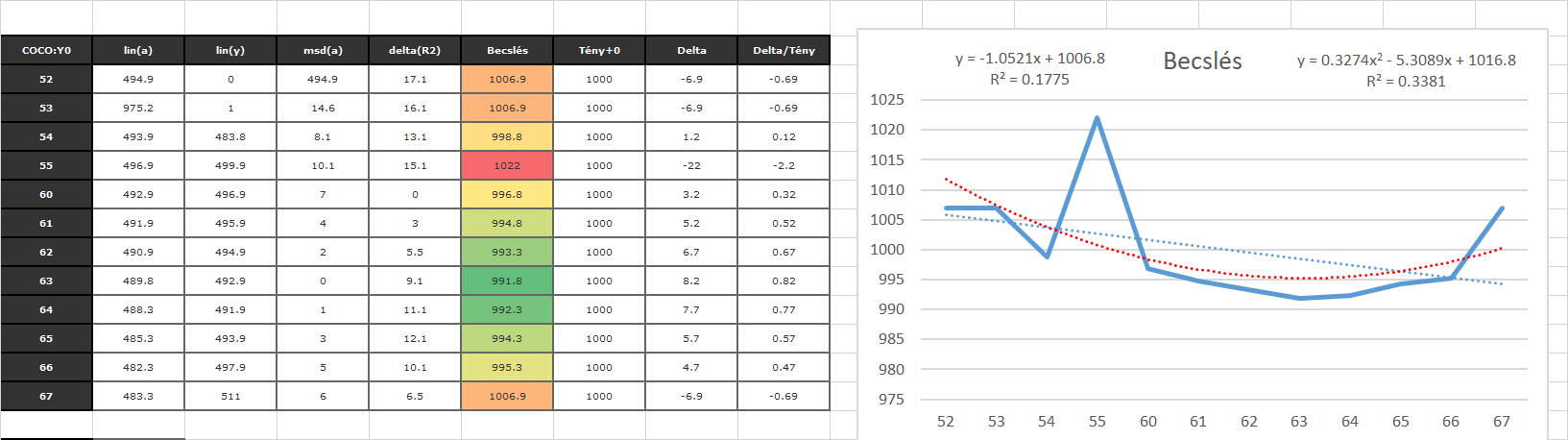
1. ábra: 2. modell (forrás: saját számítások)

A 3. modell alapján megállapítható (vö. 19. ábra): hogy a minden másként egyforma elvet már csak az 55. hét egyedisége sérti.



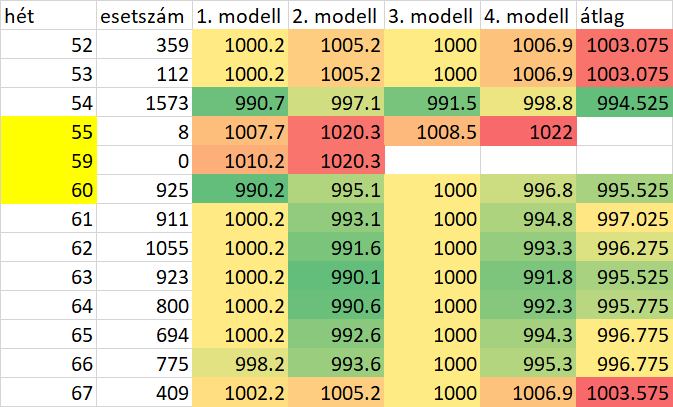
1. ábra: 3. modell (forrás: saját számítások)

A 4. modell (vö. 20. ábra), ahol sem az 59. hét zavaró hatása, sem az adatmennyiség kompenzáló hatása nem hagyja hátra bélyegeit, a legerősebb R2 érték a másodfokú modell esetén látható, vagyis a rendszeren belüli döntési kényszer értékének várható értéke vélelmezhetően meredeken nő…



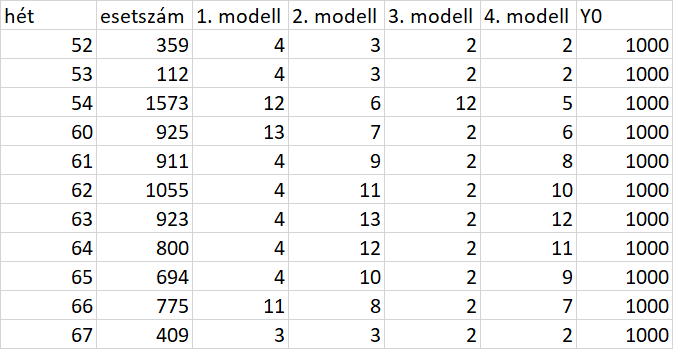
1. ábra: 4. modell (forrás: saját számítások)

Summa summarum (vö. 21. ábra): Amennyiben tehát az adathiányok által kockázatosnak ítélt 55. és/vagy 59. hét üzenetétől eltekintünk, s a 4 modell átlagát vizsgáljuk, akkor a 67. hétre a döntés szükségszerűsége magától értetődőnek tűnik.



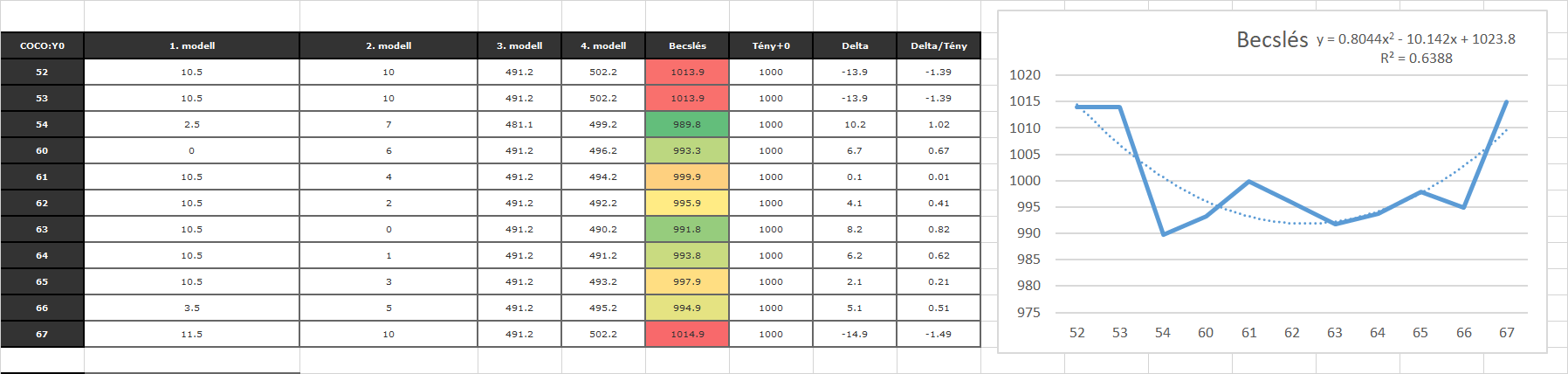
1. ábra: A 4 modell együttes értelmezése intuitív módon (forrás: saját számítások)

Természetesen ezen a ponton is érvényes a kijelentés: az átlag egy önkényes súlyozás, így a hasonlóságelemzésre alapozott anti-diszkriminatív modellezés itt is megkerülhetetlen (vö. 22. ábra):



1. ábra: A 4 modell együttes értelmezése optimalizálás keretében (forrás: saját számítások)

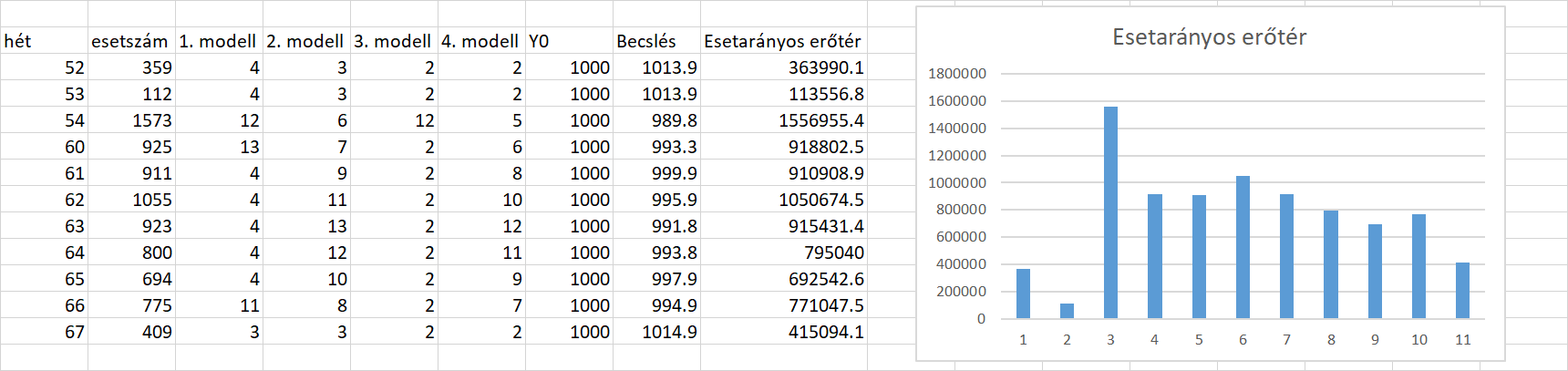
A sorszámozott, a kritikus heteket (55;59) elhagyó tanulási minta eredménye:



1. ábra: Az eredő modell (forrás: saját számítások)

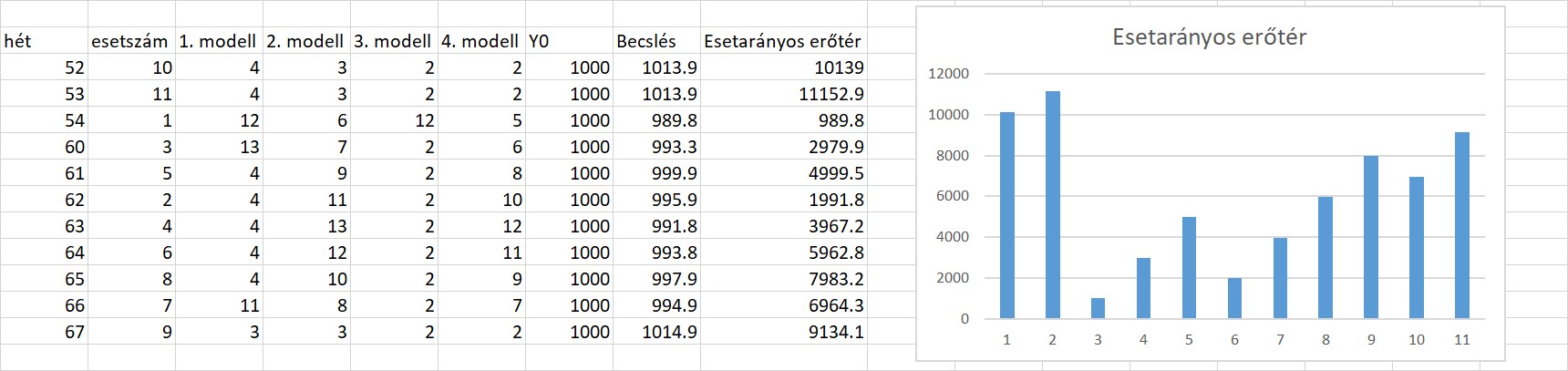
Mint látható (vö. 23. ábra), a 67. hétre, a döntés pillanatára a döntési kényszer felerősödése félreérthetetlen.

Amennyiben az esetszámokat a klasszikus formában vesszük figyelembe (vö. 24. ábra), akkor az erőtér alakja mintegy a visszájára fordul:



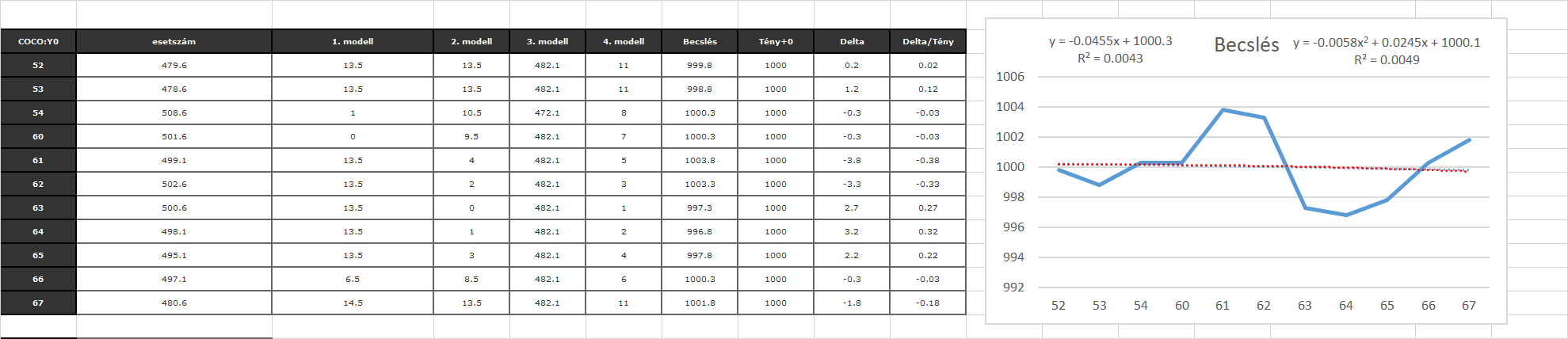
1. ábra: Esetszámok hatása (forrás: saját számítások)

Ha a klasszikus abszolút esetszámok helyett csak ezek avatárja, vagyis sorszáma kerül figyelembe vételre (vö. 25. ábra), akkor a döntési kényszerhelyzet index visszaáll a 4 modell anti-diszkriminatív alakjának megfelelő U-alakra:



1. ábra: Esetszámok hatása II. (forrás: saját számítások)

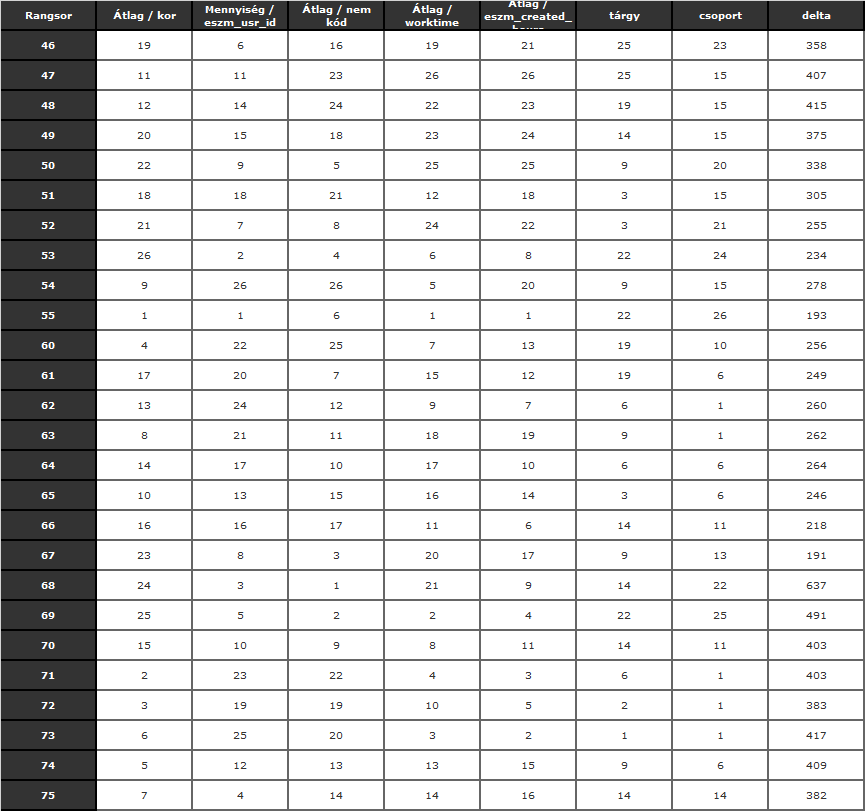
S végül, ha az esetszámok ismét csak egy anti-diszkrimanatív modellen (vö. 26. ábra) át nyernek értelmet, vagyis az esetszámok minél nagyobbak, annál inkább döntési kényszer irányába hatnak a többi tényezővel interakcióban (vö. hetek múlása, mint speciális erőtér), akkor a döntési kényszer-index értéke hullámzóvá válik, mely hullám a döntés előtt döntést legitimáló ívet ír le…



1. ábra: Esetek számának hatása III. (forrás: saját számítások)

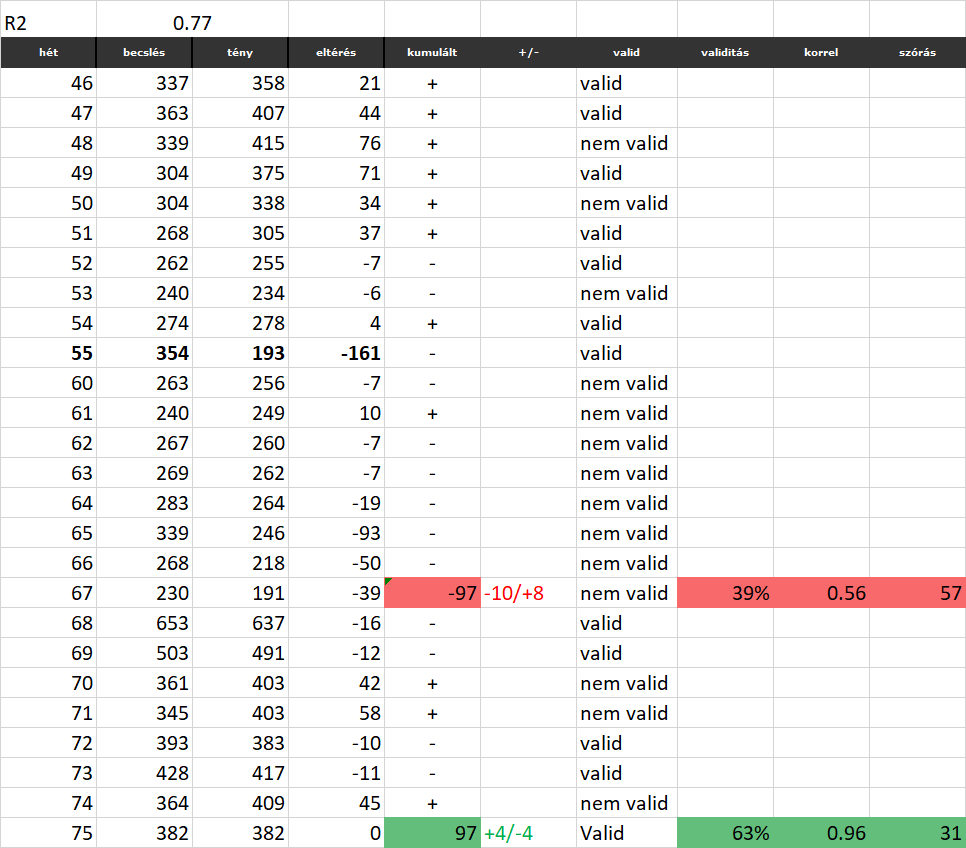
## Potenciáltól való eltérések modellje

Annak érdekében, hogy a rendelkezésre álló 46-75. naptári hetek adatai alapján egységes szemléletű, egy rendszert leíró modell jöjjön létre (vö. 27. ábra), a módosított karakterlimit-paraméter helyett függő változóként pl. a mindenkor érvényes karakterlimittől való távolságát érdemes vizsgálni a mindenkori heti átlagos válaszhossznak: vagyis az első 18 megfigyelés kapcsán 2500-hoz, majd a 68-75. hét kapcsán 3000 karakterhez kell viszonyítani (Y). Az így kialakuló tanulási minta egy termelési függvényt, vagyis egy standard hasonlóságelemzést vár el:

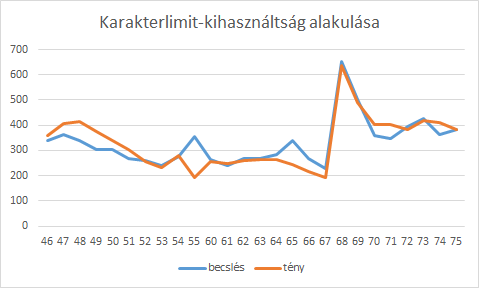


1. ábra: A potenciáltól való eltérés modellje (forrás: saját számítások)

A termelési függvény tehát a genetikai potenciálbecslés során már megismert inputok és a teljes időszakra érvényes irány-paraméterek mellett készült. Ennek eredménye (vö. 28. ábra):



1. ábra: Eredmények (forrás: saját számítások)

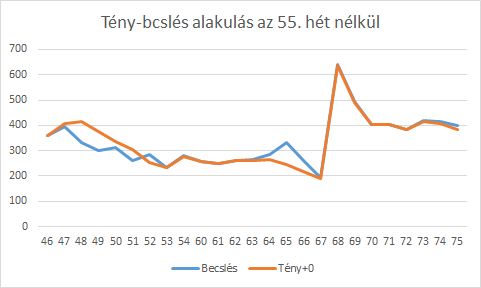


1. ábra: A karakterlimit kihasználtságának alakulása I. (forrás: saját számítások)

Ahogy az az eredmények (vö. 28-29. ábra) alapján világosan látható:

* a döntés előtti modell (rendszer-értelmezés) minden szempontból instabilabb,
  + kisebb az intervallum-specifikus modell-korreláció
  + nagyobb az invaliditás (vagyis az inverz és a direkt tanulási minta tükör-szimmetriájának sérülése: vö. a tagadás tagadásának igazságtartalma)
  + aránytalanabb a becslések és tények különbségének megoszlása a nulla becslési hiba körül
  + a döntés előtti becslési hibák szórása (az 55. hét hatása nélkül is) nagyobb, mint a döntés utáni időszak becslési hibáinak szórása
  + s legfőképpen a döntés előtti rendszer becslései átlagosan magasabb értéket adnak a tényekhez képest, mint a döntés után
* a teljes időszakot egyben látni kívánó modell R2 értéke rel. magas: 0.77
* az 55. hét alacsony felhasználó-száma miatt ennek a hétnek az értelmezhetősége a legkorlátozottabb – vö. <http://miau.gau.hu/eolap/db1/2_olap_m_din.php3> - chart1 és chart2

Az 55. hét torzító hatása nélküli modell (R2=0.89) esetén a döntés előtti é a döntés utáni időszak becslési pontossága (a tény-becslés értékpárok különbségének pl. a szórása) jelentősen eltér egymástól – ami a döntés előtti rendszer instabilitását jelzi a döntés utáni állapottal szemben, hiszen a döntés utáni időszakban alapvetően nincs is becslési hiba (vö. 30. ábra):



1. ábra A karakterlimit kihasználtságának alakulása II. (forrás: saját számítások)

# Összegzés

A fentiekben egy valós, de anonimizált adatvagyon alapján egy konkrét, stratégiai döntési helyzetet mutattunk be, miként lehet egy szubjektív, intuitív emberi döntés előtt, ennek előkészítéseként és/vagy utólag a döntéshozatal folyamatát kibernetikus rendszerként támogatni. A fentiek alapján kijelenthető hogy a támogatás ebben, vagy bármilyen más esetben csak akkor valósítható meg, ha a képzéshez kapcsolódó adatvagyon gazdálkodási rendszerünk kialakult.

Egy képzési adatvagyon gazdálkodási rendszer lépései sablonosak, melynek lépései végrehajtásának eredményeképp a döntések egy-egy eleme akár automatizálható. A döntés előkészítés nem áll meg ott, hogy a múltra/jelenre vonatkozó adatokat bemutatjuk, hanem egyfajta szenárió jóslást tudunk véghezvinni, s a döntéshozót ezáltal nagyobb mértékben támogatni.

Az első rész zárásaként az adatvagyon gazdálkodási rendszer kialakításának egy általunk is követett modelljét osztjuk meg.

*Képzési rendszerek szakmai irányítását támogató döntéshozói információs rendszer kialakításának modellje:*

|  |  |
| --- | --- |
| Termék / szolgáltatás leírása | A képzési rendszerről gyűjtött adatok megjelenítésére, elemzésére, nemzetköziesítésére - ez alapján képzési folyamatok előjelzésére és döntéstámogatásra alkalmas felület döntéshozók számára. A szak- és felnőttképzés, valamint kapcsolódó alrendszerek által begyűjtött, tárolt információk, adatok elemzésével támogatja egy komplex Döntéshozói Információs Rendszer (DIR) létrehozását. |
| Tevékenység/Indikátorok | * Adatvagyont leíró statisztikák:   + objektumok száma   + attribútumok száma   + dashboard-ok száma   + OLAP-szolgáltatások száma   + ... * Elemzési erőforrások katalógusa:   + vizualizációs eljárások száma   + egyedi modelltípusok száma   + modell-sorozatok száma   + hermeneutikák száma   + szakértői rendszerek száma   + szimulációk száma   + … * Alkalmazás-orientált paraméterek:   + felhasználók száma   + felhasználások száma   + támogatott döntési helyzetek száma   + ismert döntési helyzetek száma   + … * ... |
| Fejlesztés időtartama | Folyamatos |
| Termék / szolgáltatás előkalkulációs költsége (bruttó) |  |

## Fejlesztési terv/vázlat

A képzési rendszerek dinamikus jellegüknél fogva folyamatos megfigyelést, információgyűjtést és adatelemzést igényelnek, amelyek révén biztosítható az adatvezérelt, tényalapú döntéstámogatás, rendszerfejlesztés és –működtetés.

Alapvető elvárás a képzési rendszerekkel kapcsolatban, hogy alkalmasak legyenek az érdemi kompetenciafejlesztésre, végső soron a tudásmenedzsment kultúra- és szemléletváltás előmozdítására.

A 2012-es párizsi UNESCO nyílt oktatási erőforrásokról szóló (OER) dekrétum aláírásával a Magyarország nemzetközi szinten is törekszik az oktatási erőforrások nemzetközi nyitására. Ennek első és megalapozó, minta jellegű lépése lehet Magyarország részéről a döntéshozói oktatási adatvagyon nemzetközi disszeminációja a TEIR rendszer mintájára.

A döntéshozói információs rendszer (DIR) a továbbképzési adatvagyon mennyiségétől, minőségétől, dinamikájától függően képes döntéstámogató funkciót betölteni. Az továbbképzési adatvagyon-gazdálkodás (mely egyben a minőségirányítás alapja és nem azonos önmagában a gyűjtéssel és tárolással) kell tehát, hogy minden erre épülő elemzési és adat-vizualizációs tevékenység is magas szintű lehessen.

## Célok és várható eredmények

A képzési rendszer digitalizációjából adódó **adatosítás** elengedhetetlen támogatást nyújt a teljes  rendszer minél hatékonyabb megalkotásának. Ennek révén válik elérhetővé a valós idejű és szekunder adatszolgáltatáson alapuló döntéshozói információs rendszer. A tanulás-tanítás támogatásához szükséges az automatizálás, és ez alapján a teljes fejlesztési ciklus tényalapú döntéshozatalra építése.

Az elsődleges cél, bármilyen a képzési rendszerrel kapcsolatos kérdés merül fel a döntéshozók részéről, azokra releváns választ tudjon kapni a rendelkezésre álló adatok alapján. Ennek eléréséhez a következő lépések szükségesek:

* A nulladik lépése minden online képzés során gyűjtött adatokkal kapcsolatos kérdés felvetésnek, az objektumok meghatározása, ill. a kiválasztott objektumokhoz minden rendelkezésre álló nyers attribútum  mátrixba (OAM-ba) rendezése.
* A nyers attribútumok alapján származtatott attribútumok definiálása - célirányosan.
* A következő továbbképzési adatvagyon gazdálkodási lépés az OLAP (On-Line Analitical Processing/online analitikai feldolgozás) lesz.
* Egy döntéshozói információs rendszert kívánunk létrehozni mely minden olyan OLAP/PIVOT-beállítást enged, melyhez értelmező szövegek kreálhatók - tárolandók. Így egy OLAP/PIVOT-paraméter-KATALÓGUS jön létre, melyet a döntéshozó a saját igényeinek megfelelően rendszerez, és/vagy információbróker/adatbányász-munkatársakon keresztül szólít meg.

***Ezek alapján, lehetőség nyílik meghatározni akár az operatív*** *- vö a tanulmányban szereplő példát -* ***, akár a stratégiai fejlesztési irányokat.***

*Összefoglalva a módszertan-fejlesztés megújításának lényegét, a kínálatorientált és inputvezérelt döntéshozás helyett olyan holisztikus (minden képzési komponensre kiterjedő), igényalapú, célorientált és tényvezérelt modellt képzelünk el, amely a releváns célokat az arra alkalmas eszközökkel kívánja elérni. Az elvárásokból kiindulva pedig a megvalósíthatóságot a sztenderdizálással támogatja mind az implementáció (prototípus-fejlesztés), mind a terméktámogatás (tesztelés), végül pedig a visszacsatolás (validálás) révén a tervezés fázisaiban.*

### A fejlesztés szükségessége

Célirányosan csak akkor lehet hatékonyan cselekedni, ha a kiindulási állapot és a célállapot távolságát, illetve bármely cselekvési alternatíva által okozott tervezett és majd tényleges - tervek és tények közötti - állapotváltozásokat mérni lehet. **A képzési rendszerek megkívánják az egyenszilárdságot, a transzparenciát, a hatásosságot és a minél magasabb hatékonyságot.**

### Lépések

Az adatvagyon gazdálkodási rendszerek 9 tevékenysége különíthető el melyek részben párhuzamosan is futnak, iteratív ciklusokban.

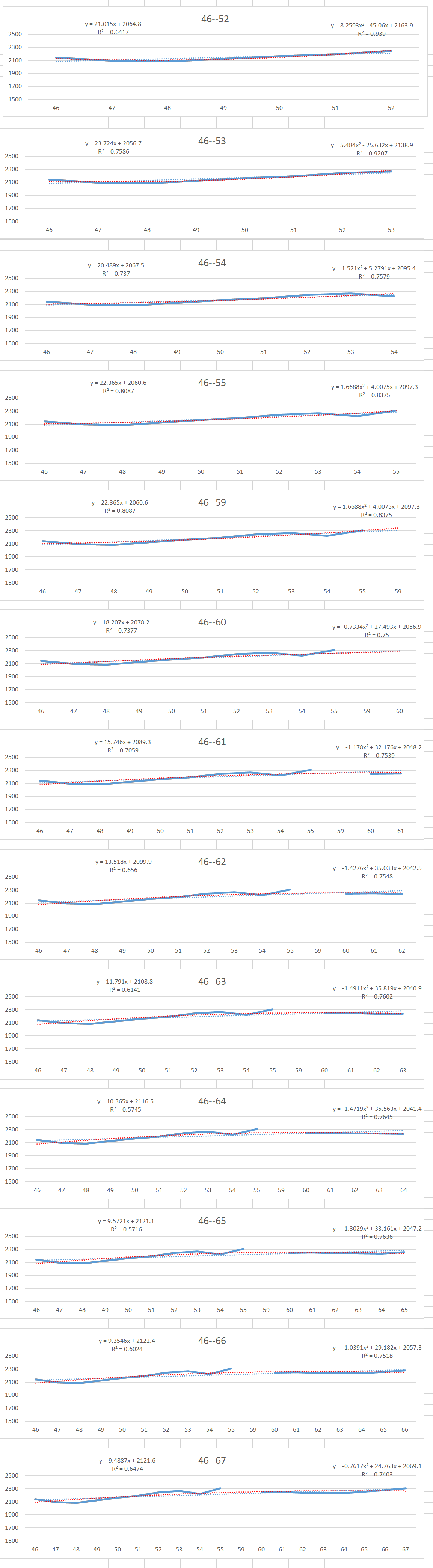
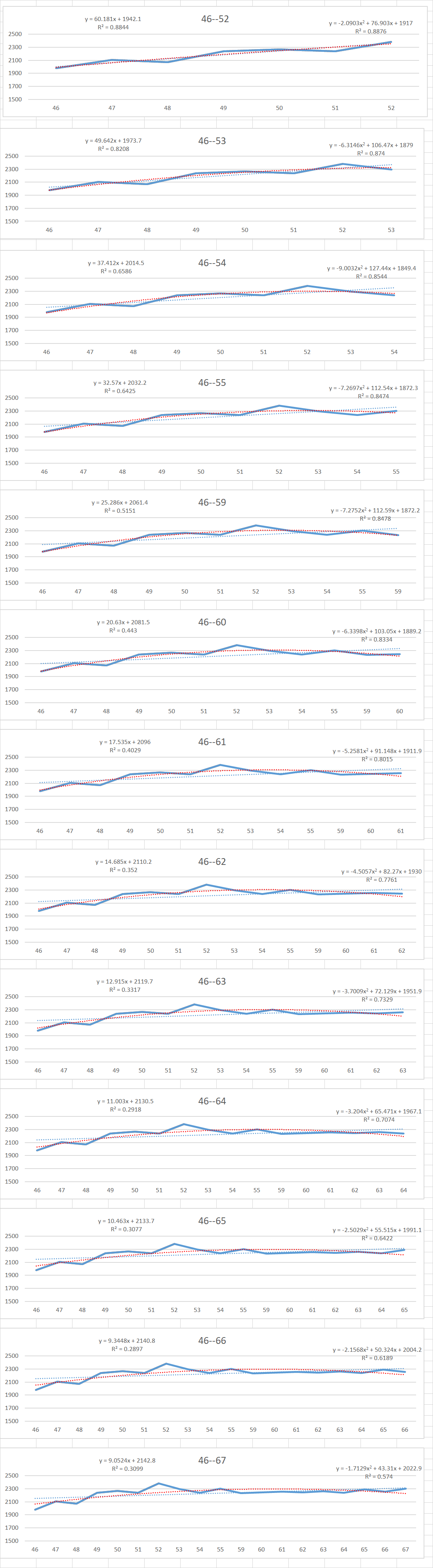
|  |  |  |
| --- | --- | --- |
|  | ***Tevékenységek*** | ***Eredmény*** |
| 1 | **Adatbázis feltárás** | Rendelkezésre áll az összes képzéshez kapcsolódó adatbázis és azok mezőinek technikai leírása. |
| 2 | **Mező konszolidálás** | Meghatározásra kerül mi mivel feleltethető meg, adat értelmezés – beleértve a definíciók megalkotását, konszolidációját - , adat ellenőrzés |
| 3 | **Elemzési célok meghatározása** | Tényalapú döntéshozás megalkotása  Minden létező objektumhoz minősítések létrehozása a legkülönbözőbb szempontokból.  Komplex fejlesztési módszertan komponenseinek meghatározása: a rendelkezésre álló továbbképzési adatvagyon értelmezése alapján kvantitatív kutatás meghatározása és indítása.  Kutatási fázisok: kutatási cél, kutatási stratégia, kutatási kérdések, operacionalizálás, mintavétel, adatfelvétel, adatelemzés, kódolás, eredmény prezentálás. |
| 4 | **Kvalitatív mérési módszerek megalkotása** | A kvalitatív kutatási eredmények alapján kvantitatív kutatás megvalósítása, - számszerű, statisztikai adatok generálása történik. |
| 5 | **Meta kereső** megalkotása | Ismert típushibákat kizáró robotizált eljárási rend kialakítása és működtetése az továbbképzési adatvagyon minősége érdekében. |
| 6 | **Adatgyűjtés**i módszerek meghatározása | A reprodukálhatóság érdekében további – „mit”,”hol”,”mivel” -lépések meghatározása. |
| 7 | **Vizualizáció** | Grafikai optimalizációs megoldás üzenet és célcsoport függően |
| 8 | **Automatizáció** | Az előző lépésekből legalább egy átadása az IT-nak. Pl: OAM-ok automatikus definiálása és feldolgozása. 1-7 iterációja  pl.: OLAP |
| 9 | **Szcenárió jóslás** (folyamatok előrejelzése) | A 8-es ponthoz prototípus szimulátor automatikus építése. |

**A fentiek eléréséhez szükséges humán erőforrás tervvázlat:**

Az adatvagyon gazdálkodási rendszerek kialakítása és üzemeltetése erőforrás igényes feladat. Ennek egy optimális humán erőforrás terv vázlatát mutatjuk be.

|  |  |
| --- | --- |
| **Product Manager** | Megérti a folyamatosan változó megrendelői igényeket és segít, hogy ezt a fejlesztő csapat is megértse, azzal együtt, hogy milyen sorrendet állítunk fel a feladatok között és miért |
| **Data Engineer** | Létrehozza és működteti a hardveres és szoftveres infrastruktúrát. (A szerverek bizonyára bérelhetőek, de valakinek kezelni kell rajtuk a szoftvereket, amiket használnunk.) Nagy tapasztalata van nagy mennyiségű adat tárolásában, strukturálásában és kezelésében. |
| **Junior Data Engineer** | Data Engineer munkáját támogató junior |
| **Data Analyst** | Együtt dolgozik a Data Engineer-rel, hogy a választott eszközök és adatstruktúrák a lehető legjobban támogassák a majdani elemzéseket. A megrendelői kérdések alapján egyedi elemzéseket és automatikus riportokat/dashboard-okat épít. Azért szükséges 2x fő, mert jellemzően így lesznek egyensúlyban a kapcsolódó szerepek munkaerejével; és mert több bevett minőségbiztosítási módszerhez is kell, hogy két elemző legyen. (pl.: Peer Prgoramming, Code Review) |
| **Junior Data Analyst** | Data Analyst-(ek) munkáját támogató személy(ek) |
| **Data Scientist** | Magas szintű statisztikai ismeretekkel rendelkezik, lehetőség szerint tud programozni, és ő tervezi meg és hajtja végre a legösszetettebb elemzési feladatokat, valamint támogatja a Data Analysteket a munkavégzésben, statisztikai tanácsokkal látja el őket. |
| **UX Designer** | Különböző módszerekkel validálja és pontosítja a Product Manager által azonosított ügyféligényeket, és az által tervezett interakciók működését. |
| **UX Researcher** | UX Designer munkáját támogató szociológiában jártas személy |
| **Graphical (UI) Designer** | Olyan grafikus toolkit-et tervez és tart karban, amit a Data Analyst-ek és a Data Engineer fel tudnak használni az általuk végzett fejlesztések során. |

# Melléklet

# Irodalomjegyzék

1. Ai, J., & Laffey, J. (2007). Web mining as a tool for understanding online learning. Journal of On-line Learning and Teaching, 3, 160-169.
2. Balogh, I. (2010): Adatbányászat alkalmazása az oktatásmenedzsmentben. In: Balogh Imre, Horváth Ádám (szerk.): Felhasználói viselkedés elemzése webes környezetekben (Web-analitikai módszerek alkalmazása viselkedéselemzésre), Székesfehérvár: DSGI, pp. 6-26
3. Etzioni, O. (1996). The World-Wide Web: quagmire or gold mine? Communications of the ACM, 39(11), 65-68. doi:10.1145/240455.240473
4. Jókai, E., Horváth J., Nagy, Zs. (2010): Tanulási szokások vizsgálata web bányászati technikákkal, Felhasználói viselkedés elemzése webes környezetben. , In: Balogh Imre, Horváth Ádám (szerk.): Felhasználói viselkedés elemzése webes környezetekben (Web-analitikai módszerek alkalmazása viselkedéselemzésre), Székesfehérvár: DSGI, pp. 73-99.
5. Nachmias, R., & Hershkovitz, A. (2007). Using Web mining for understanding the behavior of the online learner. International Workshop on Applying Data Mining in e-Learning (ADML'07). https://scholar.google.hu/scholar?q=Using+Web+mining+for+understanding+the+behavior+of+the+online+learner+R+Nachmias%2C+A+Hershkovitz&btnG=&hl=hu&as\_sdt=0%2C5
6. Osmar R. Zaïane, J. L. (2001). Towards Evaluating Learners' Behaviour in a Web-Based Distance Learning Environment. Paper presented at the In Proc. of IEEE International Conference on Advanced Learning Technologies (ICALT01.
7. Pahl, C. (2004). Data Mining Technology for the Evaluation of Learning Content Interaction - ProQuest. International Journal on ELearning, 3(4), 47/55.
8. Petákné Balogh, A. (2014): Az e-learning szerepe a felsőoktatási intézmények tudásmenedzsment rendszerében. Szent István Egyetem. Gödöllő.
9. Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.
10. Scheuer, O., & McLaren, B. M. (2012). Educational Data Mining. In Encyclopedia of the Sciences of Learning (pp. 1075-1079): Springer US.
11. Tóth, P. (2010): Virtuális kurzusbeli tanulási tevékenység vizsgálata web bányászati módszerekkel, In: Balogh Imre, Horváth Ádám (szerk.): Felhasználói viselkedés elemzése webes környezetekben (Web-analitikai módszerek alkalmazása viselkedéselemzésre), Székesfehérvár: DSGI, pp. 25-72.
12. TESS projekt, 2010: <http://www.tess-project.eu/deliverables/TESS_wp4_d41_Database_of_models_that_relate_species_and_incomes_to_land-use_15_Nov_2010_IST.pdf#page=41>
13. A. Balogh, L. Pitlik, F. Szani, M. Schnellbach, “Assessing diversity in learners background and performance” Diversity Matters! Proceedings of the European Distance and E-Learning Network 2017 Annual Conference Jönköping, 13-16 June, 2017 ISBN 978-615-5511-18-9 , pp. 235-241
14. A. Balogh, L. Pitlik, F. Szani, M. Schnellbach, “Assessing diversity in learners background and performance” Diversity Matters! Proceedings of the European Distance and E-Learning Network 2017 Annual Conference Jönköping, 13-16 June, 2017 HU-ISSN 1419-1652, MIAU, http://miau.gau.hu/miau/225/EDEN.pptx
15. <http://www.munkahelyistresszinfo.hu/a-munkahelyi-stressz-merese/a-pszichoszocialis-kockazat-jogi-szabalyozasa/>
16. <http://www.tankonyvtar.hu/hu/tartalom/tamop425/0027_MSTE7/ch01s02.html>
17. <https://miau.gau.hu/mediawiki/index.php/Szak%C3%A9rt%C5%91i_rendszer>
18. <http://janus.ttk.pte.hu/tamop/kaposvari_anyag/martinko_jozsef/ch07s04.html>
19. Ruff Ferenc, Pitlik László, Robotszem, HU-ISSN 14191652, No. 209. <http://miau.gau.hu/miau/209/szigma_folyt_final.doc>

1. Szerzők: Balogh Anikó, Pitlik László, Schnellbach Máté, Szani Ferenc

   Intézményi adatok: Apertus Nonprofit Kft. [↑](#footnote-ref-1)
2. A cikksorozat további részei dolgozzák fel majd fel azt a problémakört, vajon szövegbányászati és egyéb (szerzőt, tanulócsoportot, időszakot, tantárgyat, feladatot, oktatót, stb. leíró) jellemzők alapján milyen pontossággal lehet az megbecsülni, hogy egy-egy válasz mennyi (ill. milyen minőségű) megjegyzést képes katalizálni? [↑](#footnote-ref-2)