

Kodolányi János Egyetem

SZAKDOLGOZAT

**MURMANN MARCELL
ÜZEMMÉRNÖK-INFORMATIKUS
ALAPKÉPZÉSI SZAK**

**Budapest
2025**

**Kodolányi János Egyetem
Újmédia és Kreatívipari Kar**

**A Big Data szerepe, alkalmazási környezete, valamint
a Magyar Honvédség vonatkozásában történő felhasználása,
azon belül is a toborzás témakörben**

*The role of Big Data, its application environment, and
its use in relation to the Hungarian Armed Forces,
particularly in the area of recruitment*

Konzulens: Dr. Pitlik László

**Készítette: Murmann Marcell
Üzemméternök-informatikus szak**

Budapest

2025

Tartalomjegyzék

1. FEJEZET: BEVEZETÉS	6
1.1 CÉLOK	6
1.2 MOTIVÁCIÓ	7
1.3. CÉLCSOPORT.....	7
1.4. HASZNOSSÁG	8
1.5. A DOLGOZAT SZERKEZETÉRŐL.....	8
2. FEJEZET: SZAKIRODALMI HÁTTÉR.....	9
2.1 BIG DATA TÖRTÉNETI ÁTTEKINTÉSE, FOGALMA ÉS JELLEMZŐI	9
2.1.1. <i>Big Data</i> korszak kialakulása	9
2.1.2 <i>Big Data</i> fogalma.....	9
2.1.3 <i>Big Data</i> alkalmazási környezete, elemzések és módszerek	11
2.2 KATONAI ALKALMAZÁSOK	16
2.3 A HR ÉS BIG DATA KAPCSOLATA	17
2.4 TOBORZÁS ÉS MEGTARTÁS A MAGYAR HONVÉDSÉGBEN	19
2.4.1 <i>Jelenlegi helyzet bemutatása</i>	19
2.4.2 <i>Toborzás</i>	21
2.5 A HASONLÓSÁGELEMZÉSRŐL ÁLTALÁBAN	23
2.6. A TANTÁRGYAK ÉS A SZAKDOLGOZAT KAPCSOLATA.....	24
2.6.1 <i>Adatbázisok</i>	24
2.6.2 <i>Adatszerkezetek és algoritmusok</i>	24
2.6.3 <i>Az elektronika fizikai alapjai</i>	24
2.6.4 <i>Elektronikus áramkörök</i>	24
2.6.5 <i>Emberi viselkedés és kommunikáció</i>	24
2.6.6 <i>Európai civilizáció és identitás</i>	24
2.6.7 <i>Felhasználói interfészek és vizualizáció</i>	24
2.6.8 <i>Hálózatok és számítógép architektúrák</i>	25
2.6.9 <i>Informatikai védelem és biztonság</i>	25
2.6.10 <i>Innovatív információs és kommunikációs technológiák</i>	25
2.6.11 <i>IT biztonsági fejlesztések minőség és projektmenedzsmentje</i>	25
2.6.12 <i>A jog szerepe a modern társadalmakban</i>	25
2.6.13 <i>Komplex társadalomtudományi ismeretek</i>	25
2.6.14 <i>Kultúra, sport, munkahelyi jólét</i>	26
2.6.15 <i>Matematikai alapok</i>	26
2.6.16 <i>Mentori óra</i>	26
2.6.17 <i>Mesterséges intelligencia az IT biztonság területén</i>	26
2.6.18 <i>Operációs rendszerek</i>	26

2.6.19 Programozás	26
2.6.20 Programozási alapelvek és módszertanok	26
2.6.21 Rendszermodellezés	27
2.6.22 Rendszertervezés.....	27
2.6.23 Szoftverarchitektúrák	27
2.6.24 Szoftvertesztelés.....	27
2.6.25 Szoftverüzemeltetés.....	27
2.6.26 Tudásmenedzsment az IT biztonság területén.....	27
2.6.27 Vállalati gazdaságtan	27
2.6.28 Vezetési és vállalkozási ismeretek	27
3. FEJEZET: SAJÁT FEJLESZTÉS BEMUTATÁSA A SZERZŐDÉSHOSSZABBÍTÁS-ELEMZÉSHEZ	28
3.1. ALAPADATOK	28
3.1.1. Az adatvagyon koncepciója és szükségessége.....	28
3.1.2. Adatgyűjtési fázis	29
3.1.3. Az adattisztítás és validálás fázisa.....	30
3.2. SZÁMÍTÁSOK MAGYARÁZATA	30
3.2.1. Korrelációs mátrixok	31
3.2.2. Korrelációs eltérések elemzése	32
3.2.3. A hibák normalizálása és tűrések értelmezése.....	34
3.2.4. A tűrések jelentősége	35
3.2.5. Solver.....	35
3.2.6. 20 soros kísérleti panel	36
3.2.7. COCO Y0-modell alkalmazása	37
3.2.8. Számítás összegzése	39
3.5 ADATVÉDELEM ÉS GDPR	40
3.6 AUTOMATIZÁLÁSI LÉPÉSEK	43
3.6.1. Automatizált lépések sorrendje	45
3.6.2. Integrációs lehetőség.....	47
3.6.3. Felhasználási terület.....	47
3.7 TESZTELÉS.....	47
3.8 IT-BIZTONSÁGI KÉRDÉSEK.....	48
3.8.1. Fenyegetésmodell	48
3.8.2. Technikai kontroll	49
3.8.3. Titkosítás.....	49
3.8.4. Eszköz- és platformbiztonság	49
3.8.5. Fájl- és verziókezelés	49
3.8.6. Hálózat és adatmozgás	49

3.8.7. Összegzés	49
3.9. MESTERSÉGES INTELLIGENCIA ASPEKTUSOK	49
4. FEJEZET: VITA.....	51
5. FEJEZET: KÖVETKEZTETÉSEK	53
6. FEJEZET: ÖSSZEFOGLALÁS	54
7. FEJEZET: JÖVŐKÉP	55
8. FEJEZET: MELLÉKLETEK	56
8.1 IRODALOM- ÉS FORRÁSJEGYZÉK	56
8.2 ÁBRAJEGYZÉK.....	59
8.3 RÖVIDÍTÉSEK JEGYZÉK	60
8.4. RELEVÁNS LLM-KONVERZÁCIÓK TELJES SZÖVEGE	61
8.4.1. Számítással kapcsolatos kérdés és válasz	61
8.4.2. Automatizálással kapcsolatos kérdés és válasz.....	61
8.4.3. COCO-modell más HR témával kapcsolatos kérdés és válasz.....	62
8.4.4. Feketedoboz jellegű MI-moddellel kapcsolatos kérdés és válasz.....	63

1. fejezet: Bevezetés

Dolgozatomban a választott témát, a Big Data definícióját és szerepét (vö. 2.1.2 fejezet) alkalmazási környezetét (vö. 2.1.3 fejezet), valamint a Big Data korszak kialakulásának folyamatát (vö. 2.1.1 fejezet), működését és megjelenésének helyeit (vö. 2.1.3 fejezet) is érintem, illetve kitérek a katonai alkalmazására (vö. 2.2. fejezet). Ezenfelül a Big Data és az adatvédelem kapcsán a fontosabb fogalmakat bemutatom (vö. 3.5. fejezet) és definiálom (vö. 3.5. fejezet). A kutatási kérdéseim között szerepel a Big Data alkalmazásának kérdése a humánerőforrás-menedzsmentben (vö. 2.3. fejezet), illetve a szerződeshosszabbítási hajlandóságot befolyásoló tényezők hatásvizsgálatának lehetősége (vö. 2.5.2. fejezet). A hajlandóság meghatározásának alapjául egy Microsoft Excel-ben történő számítási sorozatokat vettem (vö. 3. fejezet). A számítási sorozatok egyik fő kérdése, hogy a valós adatok rendelkezésre állása nélkül hogyan is lehet valósághoz közeli értékeket eredményül kapni, amihez csak a statisztikai mutatók adottak. Ez előfordulhat úgy, hogy egy korábbi adatvagyon megsemmisül, de egyes még fennmaradó források tartalmazzák az ezekből készült leíró statisztikákat. Ezen számítási modell szerint igyekszem megállapítani, hogy mit is jelent ez egy Magyar Honvédség szerződéses állományába tartozó katona esetében (vö. 6 fejezet).

1.1 Célok

Kutatási célom, hogy a Bevezetésben említettek szerint a kizárólag statisztikai mutatók alapján a 3.2. fejezetben bemutatott számítások szerint valósághoz közeli értékek legyenek megállapíthatók.

Ez kiemelten fontos, hiszen nem tudhatjuk előre, hogy az adott cég, vagy jelen esetben a Magyar Honvédség megszeretné-e osztani a katonák adatait tartalmazó adatbázist az elemzést végző céggel/személlyel. Elengedhetetlen része a fejlődésnek – véleményem szerint ez a Magyar Honvédség számára is kitűzött cél -, hogy a toborzás mellett a jelenlegi állomány sikeres megtartása is előtérben legyen.

A Magyar Honvédség személyügyi rendszere hosszú évek alatt épült ki, így keletkeztek a HR vonatkozású adatvagyonok, amelyek zárt katonai adatbázisban találhatók, így csak az illetékes szerveknek van betekintési jogosultságuk. Ez az adatbázis minden a számításhoz szükséges adattal rendelkezik, ám ez teljesen elzárt a külső elemzői környezettől. Az adatok az Infotv. 27. § (2) bekezdés a) pontja szerint, valamint a honvédelemről és a Magyar Honvédségről szóló

2021. évi CXL. törvény 15. § (1) bekezdése alapján nem nyilvánosak, ezért még kutatási, elemzési célra sem kiadhatók. A Magyar Honvédség által felkért szereplő a számítások elvégzésére egy egyfajta honvédségen kívüli civil elemzői kör, akik részére külső szereplő révén a korábban említett jogszabályok szerint nincs hozzáférési jogosultságuk a teljes adatok megismerésére, tehát a megkapott statisztikai mutatókkal kell dolgozni. Az adatvagyon rekonstruálása kiemelt fontosságú, hiszen akkor tekinthető érdeminek bármilyen elemzés vagy modell, ha a statisztikai mutatók szerint felépített adatvagyon minél jobban megközelíti a valós értékeket, mivel így az előrejelzése nem csupán elméleti, hanem a gyakorlatban is használható közelítései a tényleges helyzetnek. Amennyiben a rekonstruált adatvagyon kellően tükrözi a valóságot, akkor a Magyar Honvédség felé a következő eredményeket lehet visszacsatolni. Első körben megállapítható mely jellemzők/mintázatok esetében mondható sikeresnek a toborzás. Ezenfelül az eredményül kapott értékek szerint előre jelezhető a leszerelési szándék, valamint előre kalkulálható a létszámmozgás, amit felhasználva célzott megtartási lépések tervezhetők. A Magyar Honvédség számára sürgetően megoldandó probléma a létszámstabilitási kérdés, a toborzás hatékonysága, a megtartási ráta javítása, valamint a költséghatékonyság és mindemellett nem kompromittálódnak a valós adatok. A felsorolt problémák kérdéseire a modellből kapott válasszal a Magyar Honvédség jelentős előnyre tehet szert és ez megmutatkozna a haderőfejlesztési folyamatokban is.

1.2 Motiváció

Úgy gondolom, hogy jelenleg, mint minden munkáltató jelentős munkaerőhiánnyal szenved, ami valószínű, hogy egy közepes vállalatnál nem jelent annyira komoly hátrányt, mint egy hadsereg esetében. Ahhoz, hogy a Magyar Honvédség az alaprendeltetési feladatait megfelelő idő alatt, hatékonyan végre tudja hajtani, nem csupán technikai újításokra van szükség, hanem elsősorban egy stabil, kielégítő számú professzionális állományra, akik szakmailag kiemelten végzik a munkájukat és ennek fenntartásához kellő motiváltsággal rendelkeznek. A kellő nagyságú állomány eléréséhez szükség van a szerződéses katonákra is, hogy szerződésük hosszabbításával továbbra is erősítsék a teljes állományt.

1.3. Célcsoport

Célcsoport ebben a témakörben egyértelműen a Magyar Honvédség személyügyi feladatokat ellátó állománya, valamint az általuk felterjesztett adatok alapján döntést hozó vezetők. A személyügy minden bizonnyal éves jelentésekben foglalkozik a katonák be- és kiáramlásával, statisztikában megjelenítve a pontos számokat.

1.4. Hasznosság

Az általam bemutatott modellben bár nem valós adatokkal dolgoztam, hiszen a katonákra vonatkozó személyes adatok adatvesztés miatt hiányoznak, ám ezen számításokkal valós adatokkal is meghatározható, hogy egyes katonáktól lehet-e várni szerződés hosszabbítást. Tehát a számítási modellt úgy építettem fel, mintha nem lennének elérhetők a cél tábla alapját adó valós értékek. Az korábban említett statisztikához kiegészítve alkalmazható lenne a hosszabbítások számának előrejelzése, hogy mikor is várható egy nagyobb létszámcsökkenés a szerződéses állományt érintően. Ezt látva a vezetőknek világossá válhat, hogy mely tényezők befolyásolják jelentősen a katonák döntését a leszerelést vagy új szerződés aláírását illetően. Ehhez mérten tudják alakítani, szervezni és fejleszteni a „gyenge pontokat”, ami segítséget nyújthat a kitűzött létszám eléréséhez. A megoldás kialakításához számításba kell venni az erőforrás- és költségigényt. Első körben ehhez szükség van hardverekre, vagyis munkaállomásra, külső SSD-re a biztonsági mentések elvégzéséhez, ami számszerűsítve megközelítőleg 500.000 Ft-ot jelent. Szoftveres vonalon mivel Microsoft Excel-ben történik a számítás, a Microsoft 365 vállalati verzióra van szükség, ami éves szinten körülbelül 40.000 Ft/fő. Az operációs rendszer ára a munkaállomás értékébe beszámításra került. Mérnöki munkaidő a teljes kialakítás során 2.000.000 Ft/év körüli összeg, az éves karbantartás és az esetleges segítségnyújtás 1.000.000 Ft/év. A célcsoport, ami jelenleg a Magyar Honvédség, de más jellegű felhasználás esetén bármelyik vállalat, reálisan az első évben 3.000.000 Ft-ot fizet, amit akár az éves toborzási keres 5-10%-át is adhatja, valamint a második évtől fenntartási szinten 1.000.000 Ft-ot/év. Utóbbi összeg átcsoportosítható az informatikai eszközök beszerzésére szánt keretből, hiszen mivel nem maga végzi a számításokat, így nincs szükség plusz személyek foglalkoztatására, valamint az ezzel járó műszaki beszerzés teljesítésére.

1.5. A dolgozat szerkezetéről

A szakdolgozat először bemutatja a Big Data történeti áttekintését, valamint kitér a fogalmi jellemzőire. Ezt követően az egyes már működő katonai alkalmazások ismertetésére kerül sor, valamint a Big Data és HR kapcsolatára. A témához kapcsolódóan kitér a Magyar Honvédség jelenlegi toborzási helyzetére, ezután a már ismert fejlesztendő területre válaszul megalkotott saját fejlesztés bemutatása következik. A részletes számítási magyarázat után az adatvagyon szükségességére és GDPR-ral kapcsolatos tudnivalókra hívja fel a figyelmet. A szakdolgozat zárását a kidolgozott, kutatott téma és a fejlesztési anyagok összefoglalása képezi.

2. fejezet: Szakirodalmi háttér

2.1 Big Data történeti áttekintése, fogalma és jellemzői

2.1.1. Big Data korszak kialakulása

Mint Belényesi Pál írja, „*A Big Data – a nagy mennyiségű strukturálatlan adat megjelenése az utóbbi évek felgyorsult technikai fejlődésének eredménye.*” A digitális világ elengedhetetlen elemei a digitális piacok és az ezeken működő, egyre növekvő felhasználói bázissal rendelkező digitális platformok, vagyis digitális felületek. Ezek a platformok egyre szélesebb körben terjednek és egyre több ember számára nyújtanak hozzáférést különféle szolgáltatásokhoz. (Belényesi, 2015, 127.o)

A Big Data kifejezés először a 2000-es évek elején jelent meg, a csillagászat és a genomika területén, mára szinte minden területen fellelhető, így magát a kifejezést nehéz definiálni, mivel különböző területeken más-más megfogalmazás a legmegfelelőbb rá. (Mayer-Schönberger - Cukier, 2014) „*Az adatokat ma már nem tekintjük statikus vagy gyorsan avuló dolgoknak, amelyek hasznossága csak addig tart, míg el nem érjük azt a célt, amiért összegyűjtöttük őket – például landolt a repülőgépünk, vagy a Google esetében befejezték egy keresés eredményének feldolgozását. Jelenleg inkább az üzleti világ nyersanyagának tekinthetők, alapvető bejövő információnak, amelynek felhasználásával megteremthető a gazdasági érték egy új formája. Megfelelő gondolkodásmóddal az adatok valójában újrahasznosíthatók, hogy újítások és új szolgáltatások forrásaivá válhassanak. Titkokat árulhatnak el azoknak, akik kellő alázattal, tettekésséggel és megfelelő eszközökkel figyelnek mondanivalójukra.*” (Mayer-Schönberger - Cukier, 2014, 13.o) Egy adathalmaznak három kritériumnak szükséges megfelelnie, hogy Big Data lehessen, ezek a következők: szerkezetében változatos adatok, nagyon nagy adatmennyiség, a gyors feldolgozás szükségessége. „*Amennyiben a fenti kritériumok közül bármelyiknek nem felel meg az adathalmaz, (pl.: homogén szerkezetű adathalmazról van szó), akkor Small Data-val van dolgunk (még ha a mennyiség alapján furcsán is hat az elnevezés).*” (Computerworld, 2017)

2.1.2 Big Data fogalma

Ha jellemezni kellene valahogyan a fogalmát, a Big Data nem más, mint nagy adatmennyiségekből számolt előrejelzések, valószínűségek készítése matematika alkalmazásával, az ehhez szükséges nagy adatmennyiségeket a társadalom generálja nap mint

nap. Egy, az IBM 2017-es kutatása szerint egy nap alatt 2,5 trillió bájt adatot generálunk és ez a szám egyre csak növekedik. Ezeket a hátrahagyott adatokat virtuális vagy digitális lábnyomnak is szokás hívni. (IBM, 2017)

A fogalmat érdemes több oldalról megközelíteni, beleértve a közgazdasági, versenyjogi és pszichológiai szempontokat is. Mint minden új jelenségnél, a Big Data kapcsán is gyakran hangsúlyozzák a szabályozás elengedhetetlenségét, kiemelve gazdasági, társadalmi és szociális indokokat. Hasonlóan vélekednek a Big Data elsődleges haszonélvezői — adatelemzők, közösségi hálózatok és keresőmotorok tulajdonosai, valamint digitális platformokat működtetők — akik ugyan a terület szabályozatlanságát említik, de az adatbiztonság és adatkezelés jelenlegi módszereivel kapcsolatban számos kifogás fogalmazódik meg. (Belényesi, 2015)

Ez a szemlélet jól illeszkedik a szakirodalmi megközelítésekhez, ahol a Big Data-val kapcsolatban egyszerre ismerik el annak társadalmi előnyeit és az ebből fakadó szabályozási kihívásokat, különösen az adatbiztonság és versenyjog területén. *„A Big Data versenyjogi érintettségének megértéséhez azonban elengedhetetlen a digitális felületek egyezményes meghatározása.”* (Belényesi, 2015, 130.o)

Adatainkat négy paraméterrel jellemezhetjük:

- Meddig őrzik meg, tárolják az adatokat, tehát tárolási idő
- *„Milyen mértékben befolyásolja az adat későbbi előkerülése a nyomot hagyó viselkedését?”* (Szommer-Balogh-Racskó, 2014,1.old), vagyis érzékenység/bizalmasság
- Ki férhet hozzá az adott adathoz, valamint milyen adatbázisba kerül tárolásra, avagy hozzáférhetőség szerint
- A felhasználó ismereteit jelenti az adatok rögzítésével kapcsolatban, maga a tudatosság (Szommer-Balogh-Racskó, 2014)

Andrea De Mauro kollégáival egy 2016-ban megjelent írásukban több Big Data meghatározást elemeztek, melyben a következőket emelték ki:

- A Big Data három fontos jellemzője: mennyiség, gyorsaság, sokrétűség, az angol megfelelője szerint három „V” (volume, velocity, variety)
- Kiterjed a technológia és az analízis módszerekre, azaz mire van szükség, hogy a rendelkezésre álló adatokat használni is tudjuk.

- Vonatkozik arra is, hogy az adatnak értéke van, tehát az adatokból nyerhető következtetések, elemzések a gazdaság számára hozzáadott értéket termel (De Mauro, 2016)

A végső meghatározás végül így alakult:

A Big Data az az információs erőforrás, amire jellemző, hogy nagy mennyiségű, gyors, változatos, amit speciális technológiai és analízis módszerekkel tudunk gazdasági értékévé konvertálni. (De Mauro, 2016)

A Big Data ezen felül társadalmi hatásokkal is bír, mivel egyre több adat rögzül, amikre különböző elemző algoritmusok épülnek, és az állam vagy magáncégek utána felhasználják. Ez azt jelenti, hogy hatalmas társadalmi hasznot lehet elérni az adatok felhasználásával, de ezzel együtt számos veszélyt is rejt magában. „*A kormányzatok személyesadat-kezelésének korlátok közé szorítása az első generációs adatvédelmi törekvések legfontosabb célja volt.*” (Zödi, 2017, 20.o)

2.1.3 Big Data alkalmazási környezete, elemzések és módszerek

A Big Data korszak alakulása során az adatosítás növekedett, ami azt jelenti, hogy mindenről információt gyűjtünk és digitálisan rögzítünk. Ezek, a társadalom adatai, cselekvései rögzülnek a közösségi hálózatokon, online szolgáltatásokon és az Internet használatán keresztül. (Mayer - Schönberger - Cukier, 2014)

Ami még változott, az az adatok tárolás, mivel még 2000-ben az adatok ¾-e analóg, tehát papír alapú volt, 2007-re úgy változott, hogy az adatok csak 7%-ra nem volt digitálisan rögzítve. Ez végül 2%-on állt 2013-ban, vagyis a körülbelül 1200 exabájtnyi adat 98%-a digitális, mely mennyiség könyv alakban kinyomtatva 52-szer fedné le az Egyesült Államok területét. (Mayer - Schönberger - Cukier, 2014)

Rengeteg féle módon generálódhat adat rólunk, ilyen például a hűségpont gyűjtő kártyák, digitális tartalmakat közzététele az interneten, webhelyekre regisztrálás, de ide lehet sorolni a videómegtekintési szokásainkat is. A Big Data korszakhoz egyre közelebb kerültünk az online közösségi média megjelenésével, majd mindennapossá válásával. Adataink nagy részét mi magunk osztjuk meg a közösségi oldalakon, ahol a felhasználók gyakran óvatlanul osztanak meg különböző tartalmakat és nem gondolnak az esetleges visszaélésekre, hogy esetleg illetéktelenek felhasználhatják adataikat olyasmire, amihez nem járulnának hozzá, ha

tudomásuk lenne róla. Szintén veszélyt jelent a keresőmotorok megjelenése és fejlődése is. Egy másik módja adataink kiszolgáltatásának az a különböző alkalmazások használata. (Zödi, 2017)

Az ingyenesnek tűnő szolgáltatásokért nem pénzzel, hanem személyes adatainkkal „fizetünk”, mivel a szolgáltatások használatához el kell fogadnunk az adatvédelmi szabályokat, amivel elfogadjuk, hogy adatainkhoz hozzáférhetnek és felhasználhatják, tehát mi magunk vagyunk a termék. Az adatvédelmi szabályzat megakadályozza, hogy felhasználják megadott adatainkat, ha nem fogadjuk el, de ebben az esetben a szolgáltatást is megtagadják tőlünk. A hátrahagyott adatokat az állam és a különböző magáncégek felhasználják, algoritmusokkal és statisztikai eszközökkel hasznos következtetésekre és előrejelzésekre jutnak. *„Az, hogy a kormányzatok és a magánszervezetek elkezdtek szinte parttalanul adatot gyűjteni és raktározni, valóban nem Big Data- specifikus jelenség - azelőtt is hajlamosak voltak rá -, másfelől jogi szempontból nem is különösebben izgalmas téma, hiszen hozzájárulás vagy törvényi felhatalmazás hiányában ez a gyakorlat egyértelműen jogellenes. Mégis, a Big Data - az adatok könnyű hozzáférhetősége révén, illetve amiatt, hogy az emberek önként adnak meg magukról egy sor érzékeny adatot, amelyeket olcsó és könnyű tárolni - ezt a régi problémát új dimenzióba helyezte.”*(Zödi, 2017, 22.o)

Lényegében a Big Data az előrejelzésekről, valószínűségek következtetéséről szól matematika alkalmazásával. A legtöbb felhasználó nincs tisztában vele, hogy mennyit is érnek a személyes adataik, pedig ezek az adatok ma a piac legértékesebb „árui”. Kérdéses az, hogy vajon a kapott szolgáltatások megérik-e azt, hogy az adatainkat kiszolgáltassuk különböző cégeknek. (Szommer-Balogh-Racskó, 2014) Az adatgyűjtés és az ebből nyert információk vitathatatlanul hasznosak, hiszen ezek a hatalmas adatmennyiségek segítenek bennünket a világ értelmezésében. Az adatok felhasználása lehetővé teszi, hogy olyan részleteket ismerjünk meg, amelyekhez kevesebb információ birtokában nem férhettünk hozzá. A Big Data hozzájárulhat különböző globális problémák megoldásához, segítheti a társadalmat egy felelősségteljesebb kormányzáshoz, valamint gazdasági fejlődéshez. (Mayer-Schönberger - Cukier, 2014) *„A Big Data korszaka azonban kihívást is jelent számunkra: jobban fel kell készülnünk a technológia alkalmazási módjaira, mert azok átalakítják majd az intézményeinket - és bennünket is.”* (Mayer-Schönberger - Cukier, 2014, 26.old.)

Különböző metaforákkal is próbálták bemutatni a Big Datát, például a mikroszkóphoz hasonlóan új dolgokat tesz láthatóvá, vagy egy rendelkezésre álló erőforrásra, leginkább az olajhoz hasonlítják, a digitális gazdaság erőforrása az adat. (Toonders, 2014)

„Azt, hogy „az adat az új olaj”, az internetes források szerint Clive Humby matematikus – akinek a nevé- hez fűződik például a Tesco Clubcard rendszerének kidolgozása – jelentette ki először.” (Zödi, 2017, 21.o) Mindkét metafora utal arra, hogy egyszerre jelentenek új lehetőséget és veszélyt, mivel ezek az eszközök jó dolgok, de felhasználhatóak rossz célokra is. Erre példa, hogy túl pontosan megjósolhatóvá válhatnak folyamatok, előrejelzések alapján az emberek könnyen manipulálhatóak lehetnek, amely lehet pozitív irányú (például környezettudatosság), azonban mi a garancia, hogy a vállalatok vagy az állam nem manipulálja az embereket? Ezen felül a diszkriminációt is megemlítem, például egy magasabb biztosítási díj, vagy a bűnelkövetés esélye miatti börtönbe zárás. A visszanevesíthetőség is komoly probléma, hiszen egy másik adatbázisra van szükség, amellyel összekapcsolhatják az anonimizált adatbázist. (Zödi, 2017)

Tekinthetünk a Big Datára úgy, mint egy paradigmaváltás része, amely érinti a statisztikai folyamatokat az adatgyűjtéstől egészen az adatok elemzéséig. Elmondható ezáltal, hogy befolyással bír a statisztikai hivatalok tevékenységére is, valamint átalakítja az elemzési módszereket. Ez szükségszerű, a hagyományos elemzési módszerek számos eleme egyáltalán nem vagy korlátozottan alkalmas a Big Data elemzésre. Adatmennyiségre való specifikációja és az elemzés pontosságának biztosítása, valamint a torzítatlanság elkerülése kiemelkedően fontos. Generálódik a fentebb említett okok miatt új elemzési módszerek, illetve ösztönzőleg hat a gép tanulás folyamatának statisztikai alkalmazásához. (Marton, 2023)

Különböző módszerek alkalmazhatóak a Big Data elemzése során, mely függ a vizsgálandó problémakör jellegétől. Ezen belül további kategóriákra oszthatók:

- a leíró elemzés, mely a rendelkezésre álló adatokból kinyerhető információkból próbálkozik következtetéseket levonni
- a prediktív elemzés, mely a potenciális jövőbeni lehetőségek feltárását teszi, legyen szó előrejelzésről vagy épp a kockázatok azonosításáról. Egyre népszerűbbé vált különböző területeken, hiszen az adatok egyre nagyobb számban történő rendelkezésre állásával egyre kifinomultabb előrejelzési módszerek és modellek jelentek meg

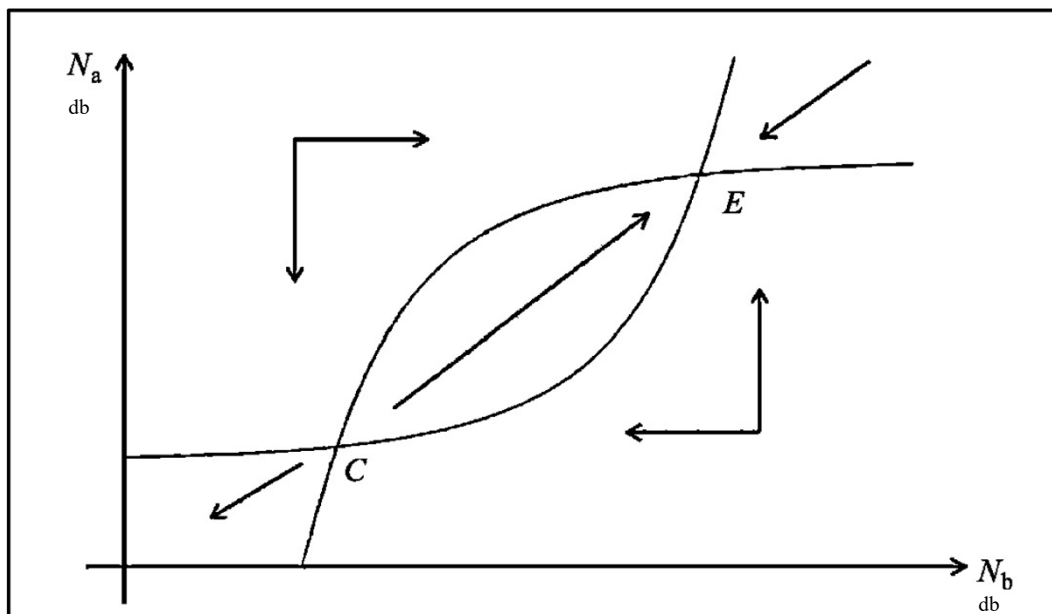
- diagnosztikai vizsgálat, ami azonosítja a bekövetkezett változások mögötti hátteret és azoknak okait. Leggyakoribb módszere a hipotézisek tesztelése, a korreláció-, és a regressziószámítás, ez utóbbi fontos szerepet játszik a prediktív elemzés előrejelzései során is
- a preskriptív analitika, ami a vizsgálandó problémakörre próbálja a megfelelő válaszlehetőséget adni, valamint ez az elemzés az alapja az adatvezérelt döntéshozatalnak. Ez az elemzés is gyakran támaszkodik a gépi tanulási folyamatok alkalmazására, azonban fontos kiemelni, hogy egyik módszer se alkalmas teljes mértékben helyettesíteni az emberi döntéshozatalt, de a preskriptív elemzés és az adatvezérelt döntéshozatal alkalmazása folyamatosan növekvő tendenciát mutat, megtalálható például az önvezető autóknál, az egészségügyben, a gyógyszeriparban és az olajiparban is. (Marton, 2023)

A regresszió- és korrelációszámítás, a hipotézisek tesztelése a leggyakoribb diagnosztikai vizsgálati módszerek. A regresszió-analízis, idősorelemzés és előrejelzés, osztályozási és klaszterelemzési módszerek, a faktoranalízis, korhoshoz elemzés, anomálielemzés, mély tanulási elemzések módszere pedig a konkrét és alkalmazható elemzési lehetőségek részről közelíthető módszerek. *„Ezzel további perspektívák nyílhatnak meg a teljes elemzési folyamat során kiindulva a leíró vizsgálatától egészen a preskriptív elemzésig, s az adatvezérelt döntéshozatalig”.* (Marton, 2023)

A hálózati gazdaságban közvetlen externáliáról (direct externalities) beszélünk, amikor egy felhasználó haszna abból származik, hogy más felhasználók is tagjai ugyanannak a hálózatnak. Teo tanulmánya Evans és Schmalensee 2010-es elemzésére hivatkozik, amely eredetileg a közösségi hálózatok kritikus tömegének elérésére összpontosított, és amely Teo szerint a virtuális fizetőeszközökre, így a Bitcoinra is alkalmazható. A Bitcoin esetében a résztvevők azok, akik bekapcsolódtak az úgynevezett „bitcoin bányászatba”, és ezáltal a hálózat aktív tagjaivá váltak (Evans et al., 2010).

Az idézett tanulmány példaként virtuális fizetőeszközökkel kapcsolatosan a különböző közösségi finanszírozó szolgáltatásokat, vagy az úgynevezett pénztárca alkalmazásokat gyártó cégeket említi meg. A közvetett externália utal a hasznosságra, ami olyan kiegészítő szolgáltatásokból ered, ami az adott hálózathoz kapcsolódó. Az indirekt externáliák által a kritikus tömeg problémája kétdimenzióssá válik. Ha egy grafikonon nézzük a kiegészítő

termékek számát (N_b) és a résztvevők számát (N_a), akkor az Evans és Schmalensee tanulmány alapján ez látható (Evans et al., 2010):



1. ábra Kritikus tömeg és az egyensúlyi helyzet a direkt és az indirekt externáliák függvényében (Evans et al., 2010)

A C-vel jelölt nyereg pont, ami a kritikus tömeg eléréséhez szükség van a kiegészítő termékek, szolgáltatások számának, illetve a résztvevők számának növekedésére, a helyileg stabil kiegyenlítődést E-vel jelöljük. (Evans et al., 2010)

Ahogy egy hálózat mérete növekszik, a benne található termék vagy szolgáltatás értéke is emelkedik. Emiatt a résztvevők hajlandóak időt, energiát vagy pénzt áldozni azért, hogy a hálózat tagjai maradjanak, illetve bekerüljenek oda. A hálózati gazdaságban lényeges szempont az is, milyen költséggel jár egyik hálózatból a másikba való átlépés. Ha a váltás egyszerű és olcsó, akkor nehéz bármiféle díjat érvényesíteni a hálózathoz való csatlakozásért. Ugyanakkor a hálózati hatás miatt általában csak akkor éri meg a váltás, ha a felhasználók egy jelentősebb része is követi ezt a lépést. (Ernie G. S. Teo, 2015)

Ezek alapján a Big Data platformok piacán a hálózati hatás kettős növekedésen alapul: az adatot szolgáltatók és az adatot felhasználók számának egyidejű növekedése szükséges a gazdaságosság és a fenntartható működés biztosításához. Ahhoz, hogy megérje a hálózatban maradás mindkét fél részéről, egyidejű növekedés szükséges, hogy elérjék az úgynevezett „nyeregpontot”. (Belényesi, 2015)

A digitális alapokra épülő piacok gyors növekedése új kérdéseket vet fel. Mivel a piac hatékony önszabályozása nem jellemző, ezekre a kérdésekre úgy tűnik, a piacfelügyeleti hatóságoknak

kell válaszolniuk. Például felmerülhet, hogy mit jelent az, ha vállalatoknak óriási mennyiségű felhasználói adatuk van (Big Data). Van-e ennek versenyjogi vonatkozása? Emellett az is fontos kérdés, hogy a digitális piacokon megadott felhasználói adatok miként tekinthetők: terméknek vagy nélkülözhetetlen eszköznek? *„Előzetesen annyi azonban megállapítható, hogy a versenyjogi szerepet tárgyaló definíciók, elemzések és vélemények egyelőre gyerekcipőben járnak. A megállapítások a szokásos akadémiai és/vagy ipari meghatározások kiforrottságát nélkülözik – ebben minden bizonnyal szerepe van a terület folyamatos átalakulásának is.”* (Belényesi, 2015)

2.2 Katonai alkalmazások

„Manapság annak lehetünk a tanúi, hogy a valós világ adatai korlátlan növekedésnek indultak, és ebből adódóan egyre nagyobb igény jelenik meg arra vonatkozóan az üzleti élet, a kormányzatok és a rendvédelmi szervek részéről is, hogy valós időben lehessen lekérdezéseket végrehajtani és információt, tudást kinyerni a meglévő adatbázisokból. Napjainkra ugyanis sok tekintetben az adat lett az alapanyag a termeléshez, ezáltal komoly gazdasági és társadalmi értékkel is bír.” (Takács, 2018, 6.o)

Egyik alkalmazási területén kiemelt helyen szerepel a prediktív analitika, másnéven az előrejelző elemzés, amely alkalmazható akár a bűnmegelőzésben vagy terrorelhárításban. Aggályos része a módszernek, ha közben szenzitív információkkal kell dolgozni (egészségi adat, faj, vallási nézet, szexuális beállítottság), melyből következik, hogy a magánélethez és a személyes adatok védelméhez való jogok és garanciák biztosításának meg kell történnie oly módon, hogy a nemzetbiztonsági vagy rendvédelmi érdekek se legyenek hátrányban. (Takács, 2018)

Az IBM Watson Discovery szolgáltatása is ezekre az eljárásokra épít, vagyis különböző digitális formátumokban tárolt struktúrált és nem struktúrált adatokat gépi tanulási algoritmusok segítségével elemzi, rendszerezi, majd NLP eljárásokkal vizsgálja, hogy az adott forrásban milyen kontextusban találhatóak meg a számunkra fontos információk. Jó példa a Watson Discovery működésére egy, a koronavírus-járvány idején végzett demo verziós teszt. A rendszer beállításainak megfelelően kizárólag a járvány kezdete utáni két hónap forrásait vette figyelembe az elemzés során. (Németh-Virágh, 2023)

A „Sentiment Analysis”, vagyis hangulatelemzés része megmutatja, hogy a keresett kifejezés pozitív vagy negatív kontextusban jelenik-e meg a szövegben. Az „Anomaly Detection”, azaz rendellenesség-észlelés rész pedig kiemeli, mely napokon mutatkozott szokatlanul magas előfordulási szám az adott időszakon belül. Ezek a grafikonok alkalmasak különféle trendek elemzésére, mivel segítségükkel például mérhető egy adott hír vagy esemény társadalmi fogadtatása, illetve egy szervezet aktuális megítélése is. Ez a megoldás lehetővé teszi a Magyar Honvédség társadalmi kapcsolatrendszerének vizsgálatát, amiből meghatározható a megfelelő kommunikációs stratégia, a katonai hivatás elismertségének erősítése, valamint a stabil utánpótlás biztosítása. A Watson Discovery másodpercek alatt elvégzi ezeket az elemzéseket, a felhasználónak csupán meg kell adnia a rendszer számára a kívánt forrásokat és az időszakot. Bár a szoftver jelenleg nem támogatja a magyar nyelvet, a Magyar Honvédség számára egy hasonló megoldás kifejlesztése további előnyökkel járhatna, például missziós területeken a CIMIC (Civil-Military Cooperation – civil-katonai együttműködés) vagy a PSYOPS (Psychological Operations – lélektani hadviselés) műveletek támogatására.

A felderítés korábban említett területei a 21. században már nem választhatók el élesen egymástól, és a különböző módszerekkel gyűjtött adatok integrált feldolgozása jelentősen növeli a kapcsolódó tevékenységek hatékonyságát. Ez az ún. összadatforrású felderítés lényege, amely komplex módszereket foglal magában a lehető legszélesebb adat- és információforrások, valamint kapcsolódó technikák használatára a cél elérése érdekében. Az eltérő formátumú adatok egységes rendszerbe foglalása úgynevezett fúziós adatfeldolgozással történik, csak ezt követően nyerhetők ki belőlük hatékony információk. Az információfeldolgozás terén a mai mesterséges intelligencia rendszerek már jelentős támogatást nyújtanak, s fejlődésük alapvetően befolyásolja majd a jövőbeli fegyveres konfliktusokban az információs fölény megszerzésére irányuló műveletek sikerességét. *„A globális katonai erőviszonyokat a következő évtizedekben alapjaiban fogja meghatározni, hogy ki milyen ütemben, és milyen mértékben képes a mesterséges intelligenciát saját szolgálatába állítani, az általa kínált lehetőségeket, technikai megoldásokat fegyver- és vezetés-irányítási, valamint döntéstámogató rendszereibe integrálni”.* (Németh-Virág, 2023, 5.o)

2.3 A HR és Big Data kapcsolata

A gyakorlatban az analitika magában foglalja mind a hagyományos relációs adatbázisokat és táblázatkezelő programokon alapuló elemzéseket, mind az új típusú adatbázis-szoftvereket, amelyek lehetővé teszik a nagyon nagy mennyiségű adat hatékonyabb tárolását és szervezését,

valamint az adatok vizualizáción keresztül történő ábrázolásának és megértésének új technikáit. (Angrave et al, 2016)

Pontosabban, a HRIS-ben tárolt adatok általában a felvett munkavállalókra (munkatörténet, készségek és kompetenciák, formális végzettség és demográfiai adatok) és a fel nem vett pályázókra vonatkozó információkból állnak. Miután egy munkavállalót felvesz egy vállalat, a ledolgozott órákra és a fizetésre vonatkozó adatokat rendszeresen gyűjtik és tárolják. A munkakörből függően lehetnek adatok a munkavállalók teljesítményéről is (eladások, ügyfeleknek számlázott órák, egyéni teljesítmény mérése stb.) Ezenkívül számos „puha” teljesítményadat is gyűjthető az értékelési és teljesítménymenedzsment rendszerekből, valamint a munkavállaló képzésével és fejlődésével kapcsolatos információk, a panaszokkal, képességekkel és fegyelmi ügyekkel kapcsolatos információk, a vitás ügyek rendezése, a belső kommunikáció, a részvételi programok és a személyzet attitűdjeinek felmérése. Korábban ezeket az adatokat különálló, különböző HR-folyamatok végrehajtására tervezett szoftverek tárolták, de egyre inkább összegyűjtik és felhőalapú adattárolókban tárolják őket, mivel a szervezetek beruháznak a HRIS (Human Resources Information System) fejlesztésébe. Elképzelhető, hogy ezeket az adatokat kombinálják a munkavállalók tevékenységére vonatkozó „nagyobb” adatokkal (mobiltelefonok helyadataival, internetes böngészési előzményekkel, elektronikus naptárakkal és más, a termelésben vagy szolgáltatásnyújtásban használt kézi elektronikus eszközökkel), azzal, hogy kivel kommunikálnak (e-mailek és telefonhívások, valamint online együttműködési eszközök), és hogy mit kommunikálnak. (Angrave et al, 2016)

Mindazonáltal, miközben a HR-hez kapcsolódó nagy adatok teljesebb kiaknázását lehetővé tevő technológia fejlődik, a HRIS-ben tárolt meglévő relációs adatokkal még mindig sokat lehet elérni. Valójában ezeknek az adatoknak a jobb kihasználása az értékteremtés és -megőrzés érdekében elengedhetetlen előfeltétele a fejlesztés alatt álló, fejlettebb nagy adatelemzési formáknak. Rasmussen és Ulrich két példát hoznak felegy offshore fűtővállalatnál végzett kifinomult HR-elemzési projektekre, amelyek jelentősen javították az üzleti teljesítményt. Először is, a HR-elemzést arra használták, hogy kapcsolatot teremtsenek a vezetői minőség és az alacsonyabb fluktuációs szint között, ami magasabb szintű operátor kompetenciát eredményezett, ami viszont kevesebb balesetet, kevesebb karbantartási időt és magasabb ügyfél-elégedettséget eredményezett. Másodszor, az analitikát arra használták, hogy bemutassák a vállalat diplomás képzési programjából származó jelentős előnyöket; ennek eredményeként a program mérete megduplázódott. Példának még a Tesco, itt hogyan

alkalmazta az ügyfeleinek megértésére kifejlesztett analitikai eszközöket a munkaerő jobb megértése érdekében, és hogy a McDonalds hogyan tudta azonosítani, hogy a személyzet demográfiai adatai, a vezetői magatartás és a munkavállalói attitűdök hogyan hatnak egymásra az étterem teljesítményének optimalizálása érdekében. A legfontosabb kérdés az, hogy hogyan lehet az analitikát felhasználni az HR-adatokból származó érték létrehozására, megragadására, kiaknázására és védelmére, és hogyan lehet a meglévő, alapvetően leíró jellegű HR-analitikai programokat úgy fejleszteni, hogy azok a humán tőke befektetéseinek stratégiai hatásának mérésére és modellezésére összpontosítsanak, ezáltal jobb vezetői döntéshozatali eszközöket teremtvén. (Angrave et al, 2016)

2.4 Toborzás és megtartás a Magyar Honvédségben

2.4.1 Jelenlegi helyzet bemutatása

Ahhoz, hogy megértsük az ilyen helyzet kialakulását, szükséges áttekinteni a magyarországi munkaerő-piaci viszonyokat. Az elmúlt időszakban az olyan tényezők, mint a generációk munkához való hozzáállása, gazdasági változások és technikai fejlődés, ráébresztették a vállalatok vezetőit és HR-szakembereit, hogy az emberi erőforrások kezelése ugyanolyan fontos, mint a profit termelése. A 2008-as gazdasági válság, a munkaerőhiány és a folyamatos toborzási, kiválasztási és betanítási költségek hatására a humán tőke jelentősége megnőtt. Magyarországon is súlyos munkaerőhiány alakult ki, amely komoly gondokat okoz a munkaadóknak a megbízható munkaerő megtalálásában. Ez kedvez az álláskeresőknél, akik egyre magabiztosabban lépnek fel állásinterjúkon vagy bértárgyalásokon. A kialakult helyzet rávilágít a toborzás, kiválasztás, megtartás és motiválás fontosságára egy szervezet életében. (Resperger-Resperger, 2025)

Kihívást jelent a katonai pálya vonzóvá tétele a fiatalok számára, a személyi állomány változik, és a pótlás sosem egyszerű, bizonyos szakterületeken szinte lehetetlennek tűnik. A Magyar Honvédségnek is szükséges lépni, hogy a munkaerőpiac negatív tendenciáját enyhítse és biztosítsa a megfelelő emberi erőforrást. Itt kiemelendő a generációk közötti ellentétek csökkentése, hiszen a generációk együttműködése fontos tényezője a meglévő személyi állomány megtartásához. (Ambrus, 2020)

A működéssel kapcsolatos pozitív és negatív visszajelzések, illetve fejlesztési javaslatok nagyon fontosak egy szervezet hatékonyságának növelése érdekében. Ezért egy modern

szervezetben – így a honvédségnél is – a proaktivitás egyre nagyobb szerepet kap. Így „azoknak, akik korábban sok éven keresztül utasítások alapján dolgoztak, ez az elvárás frusztrációt okoz. Ahhoz ugyanis, hogy valaki proaktívan tudjon dolgozni, a kreativitáson kívül komoly önbizalomra és szakmai önismeretre is szüksége van.” (Ambrus, 2020)

Az eltérő kompetenciák feszültségeket okozhatnak, de egy jól működő közösségben a tagok segítőkészsége és támogatása elősegítheti az egyéni képességek fejlődését. A generációtól függetlenül negatív hatással lehet a szolgálati kultúra változása, például egy helyőrségváltás, amikor a katona nemcsak egy összetartó közösséget és barátokat hagy maga mögött, hanem megszokott környezetét is. Az új helyen egy új szocializációs kihívással kell szembenéznie. Ez a terhelés nőhet, ha a katona családos, és a család nem tud vele együtt költözni, ezért hosszú távon a hétvégi ingázás fokozott terhet ró mind a katona, mind a családja számára. A helyőrségváltás kevésbé érinti megterhelően az egyedülálló fiatalokat, de az eset mindhárom generációban csökkentheti a szervezeti elkötelezettséget és a honvédség megtartó képességét. Ebben a helyzetben a befogadó szervezet pozitív hozzáállása és támogatása nagyon fontos segítséget nyújthat. (Ambrus, 2020)

A pozitív szocializáció egyik alapja a katonai pályafutás során a műveleti szintek megfelelő bejárása. Ilyenkor a katona taktikai szinten kezdi pályáját, majd ranglétrán feljebb haladva folyamatosan szerzi tapasztalatait, fejleszti kapcsolatait és bővíti látókörét. Amikor felettesei érdemesnek találják, hadműveleti, majd stratégiai szinten folytatja tevékenységét. Ez a folyamat biztosítja, hogy megértse a honvédség szerepét, saját alakulata feladatait és saját feladatait a teljes rendszerben. A szintek kihagyása azonban feszültségeket okozhat; például egy közvetlenül hadműveleti vagy stratégiai szinten kezdő fiatal tiszt megzavarhatja a jól működő közösséget. (Ambrus, 2020)

Kiváló összekovácsoló erő lehet egy katonai misszió, amelynek során minimum fél évig, de akár egy évig is együtt él és dolgozik az állomány az éles, hadműveleti környezetben. A misszió a szocializáció egyik legnagyobb kihívása, mert míg az első misszió során az Y és Z generáció tagjai számára az otthonuktól való hosszabb távolság újdonság, a tapasztaltabb katonáknak is szembe kell nézniük váratlan nehézségekkel, amelyek befolyásolhatják a jól működő kapcsolatrendszereket. (Ambrus, 2020)

Misszió alatt kiemelten fontos a közösségi szellem fenntartása, az egymásra való figyelem, valamint a fiatal katonák támogatása az idősebb, tapasztaltabb társaik által, nemcsak a feladatok teljesítésében, hanem azon túl is. A külföldi misszió előtti hazai felkészítés során a vezetők felelőssége és feladata a közösségbe nem illő vagy bomlasztó egyének kiszűrése és cseréje, hiszen ilyen beavatkozásokat az éles műveleti környezetben már nehezebb vagy lehetetlen végrehajtani. (Ambrus, 2020)

Befolyásoló tényezők:

- Demográfiai jellemzők: Életkor, nem, családi állapot.
- Szolgálati feltételek: Munkakörülmények, előrelépési lehetőségek, juttatások.
- Munkahelyi elégedettség: A munka és a magánélet egyensúlya, munkahelyi légkör.

2.4.2 Toborzás

Amikor egy szervezetnél megüresedik egy vagy több pozíció, elindul a toborzás folyamata, amelyhez belső (már meglévő munkavállalók), külső (bármely alkalmas jelentkező), pályakezdő vagy frissen végzett fiatal, illetve álláskeresők közül választhatnak. Mindegyik lehetőségnek vannak előnyei és hátrányai, amelyeket gondosan mérlegelni kell a toborzás megkezdése előtt. Kifejezetten katonai létszámbővítés esetén a belső toborzás kevésbé hatékony, míg a külső toborzás költséges és nem elég a hagyományos célcsoportokra, mint pályakezdők vagy álláskeresők alapozni. A munkaerőhiány miatt még konkrétabb, célzottabb toborzási stratégiák kialakítása szükséges, amely nem egyszerű feladat. (Resperger-Resperger, 2025)

A nehézségek mögött több tényező is állhat. Egyrészt nem szabad figyelmen kívül hagyni, hogy a több mint három éve tartó orosz–ukrán konfliktus, valamint az izraeli–palesztin válság elrettenti az átlagembert a háborútól és a fegyveres szolgálattól. Másrészt a különböző generációk eltérően viszonyulnak a háború fogalmához. (Resperger-Resperger, 2025)

„Az általános haderőreformok általában nyolc-tíz évig tartanak, egy ország védelme egy nap alatt nem építhető újjá.” (Resperger-Resperger, 2025, 82.o) Ebben a folyamatban a legfontosabb feladat a védelem legértékesebb elemének, a jól kiképzett katonának megóvása és megőrzése. A magyar nemzet 1100 éves története során kevésszer volt képes teljes mértékben saját sorsa alakítására. Az elfogadott honvédelmi törvények és a fegyveres erők szervezeti, felelősségi alapjai megszülettek, és megindult a haderőreform végrehajtása. Ez ritka sikernek számít a Magyar Honvédség számára, hiszen adott a politikai akarat, rendelkezésre áll a

törvényi háttér, a gazdasági források, valamint megkezdődött a magyar hadiipar kiépítése. (Resperger-Resperger, 2025)

A demokratikus rendszerben a közvélemény kiemelkedő szerepet játszik a politikai és állami szervezetek hatékony támogatásában. A honvédelem eszméje, a honvédelmi szervek szükségessége, valamint az erre fordított nemzeti költségvetési összegek csak a társadalom pozitív támogatásával valósíthatók meg. A korábbi, sorozásos rendszerben szolgáló katonák nem mindig értékelték a katonai szolgálat célját és eredményeit. Ezzel szemben a honvédelem iránti igény és a honvédség megbecsülését tükröző pozitív közvélemény kialakítása nélkülözhetetlen. (Resperger-Resperger, 2025)

A Magyar Honvédség elkötelezett haderőreformjának részletes tervezése folyamatban van. Magyarországnak a védelmi rendszer és a szükséges haderőreform alapelveit a nemzet történelmi hagyományaira, elfogadható nemzetbiztonsági kockázatára, az elvárható nemzetközi helyzetre, a politikai és gazdasági körülményekre, az anyagi forrásokra, valamint a fejlett demokráciák hadseregeinek tapasztalataira kell alapoznia. (Resperger-Resperger, 2025)

A haderőreform alapelvei a következők:

- a demokráciákban a civilek vezetik és határozzák meg a hadsereg fejlesztésének irányait, szervezetét, fegyverzetét, pénzügyi támogatását és alkalmazását;
- a reform céljaihoz olyan, hosszabb időszakot felölelő, előkészítő és rendszerépítő munka szükséges, amely jelentős anyagi és szellemi változásokat igényel;
- a honvédség környezeti átalakítása csak egységes koncepció alapján valósítható meg;
- a védelem megerősítéséhez szükséges politikai döntéseket részletes katonai és gazdasági elemzések alapján, átfogó haderőreform-stratégia kidolgozásával kell meghozni;
- a tervezett stratégiai célokat és alapelveket meg kell vitatni, egyeztetni, majd a reformot jól tagolt fázisokra bontva, meghatározott döntési pontokkal végrehajtani. (Resperger-Resperger, 2025)

Boros Lajos tengerészgyalogos ezredes 1994-es jelentésében foglalt javaslatok a mai napig időszerűek a Magyar Honvédség számára. Fontos megvizsgálni Európa és Magyarország

általános geopolitikai és geostratégiai helyzetét, valamint értékelni a nemzetközi tér változásait. Ajánlatos átvenni a szomszédos országok és a NATO-tagállamok sikeres tapasztalatait, amelyekre alapozva fejleszthető a tartalékos rendszer, illetve szükség esetén mérlegelhető a sorkatonai szolgálat visszaállítása. (Resperger-Resperger, 2025)

Az elmúlt években számos ország tapasztalta a katonai toborzás nehézségeit. Három évvel ezelőtt az Egyesült Államokban a legrosszabb toborzási eredményeket érték el 1973 óta.⁴⁵ Európában is hasonló a helyzet: a német védelmi minisztérium 2024 elején jelentette be, hogy 2023-ban 1500 fővel csökkent a haderő létszáma; az Egyesült Királyság 2010 óta nem éri el a toborzási célokat; Olaszország, Spanyolország és Franciaország is hasonló gondokkal küzd. Ezek fényében érdemes fontolóra venni Jacques Chirac egykori francia köztársasági elnök szavait: „*Mire lenne jó egy túlméretezett katonai gépezet, ha a működtetés feltételrendszerének biztosításában megmutatkozó társadalmi megosztottság és a deficitese államháztartás a hanyatlás útjára vinné az országot?*” (Resperger-Resperger, 2025)

Összességében a területvédelmi erők kulcsfontosságúak a helyi közösségek biztonságának és stabilitásának megőrzésében, különösen a kritikus infrastruktúra és a lakóhelyek védelme területén. A toborzás, kiegészítés és a fiatalok megtartása nemcsak a hadsereg működőképességét és hatékonyságát biztosítja, hanem nagyban hozzájárul a nemzet biztonságához és stabilitásához is. (Resperger-Resperger, 2025)

2.5 A hasonlóságelemzésről általában

Az adatok feldolgozását COCO (component based object comparison for objectivity) módszerrel hajtottam végre, amellyel kimutatható az objektumok tulajdonságaik alapján az egymáshoz történő viszonyulásuk szerinti rangsorolás. A COCO-elemzés futtatásánál a három elérhető módszer közül a tisztán Excel-alapú megoldást választottam, amely az Excel Solver bővítményét használja. https://miau.my-x.hu/miau/189/coco_demo.pdf Ez egy olyan bővítmény, ami matematikai modellek alapján határozza meg az adott probléma vagy elemzés kapcsán a legkedvezőbb megoldást. Egy optimalizálandó célfüggvény megadásával, valamint a változók és korlátozások megadása után különböző algoritmusok segítségével kiszámítja az optimális értékeket. Az előbb leírt módszer és az OAM (Objective Aggregation Model) kombinálásával lehetővé teszi több attribútum alapján az előrejelzést a katonák szerződeshosszabbítási hajlandóságának kapcsán. (Pitlik, 2013)

2.6. A tantárgyak és a szakdolgozat kapcsolata

2.6.1 Adatbázisok

Segített a normalizálásban, a „cél” és „becslés” táblák logikus szétválasztásában, amire az egész Excel épült.

2.6.2 Adatszerkezetek és algoritmusok

A hibák normalizálásával az azonos nagyságrendre hozás biztosítja, hogy a hiba1–hiba2–hiba3 összevethető legyen, így egyik sem dominál és az eredmények nem adnak torz képet.

2.6.3 Az elektronika fizikai alapjai

A tárgyhoz kapcsolódó mérési hiba, zaj és tűrés fogalmához lehet társítani az általam használt tűrés értékek meghatározását.

2.6.4 Elektronikus áramkörök

A hibákat és súlyokat olyannak lehet tekinteni, mint az ellenállásarányok a feszültségosztóban, vagyis amelyik „ellenállás” (súly) nagyobb, ott arányosan nagyobb „feszültség” figyelhető meg.

2.6.5 Emberi viselkedés és kommunikáció

A későbbiekben valós adatok esetén felhasználói visszajelzések szerint lehet hangolni a küszöbértékeket, hogy a rendszer a terepi tapasztalatokhoz is igazodjon.

2.6.6 Európai civilizáció és identitás

Az identitásról tanultak alapján a motiváció és a narratíva képezi a kulcsot. A modell a „mely tényezők kapcsolódnak leginkább a maradási döntéshez?” kérdésre számszerű választ ad, így a toborzási üzenetek célzottan hangolhatók a nagyobb hatékonyság érdekében.

2.6.7 Felhasználói interfészek és vizualizáció

Az értékek közötti eltéréseket feltételes formázással oldottam meg ($\pm 0,2$ zöld, $0,2-0,4$ sárga, $>0,4$ piros), úgy gondolom, hogy a jó vizualizáció segít a megértésben.

2.6.8 Hálózatok és számítógép architektúrák

A Hálózatok és számítógép-architektúrák tárgyból hozott redundancia-szemléletet követve az Excel-fájlokat nagyobb módosításoknál külön verzióra nevezve mentettem el (v1...vN), így hiba vagy félreszámítás esetén bármikor vissza tudtam állni adatvesztés nélkül.

2.6.9 Informatikai védelem és biztonság

A tárgy magába foglalja az adatvagyonok kapcsán a CIA-triádát, valamint a GDPR előírásoknak való megfelelést.

2.6.10 Innovatív információs és kommunikációs technológiák

Későbbiekben az esetleges honvédségi felhasználás esetén az elemzést végző állomány tabletjeit MDM-mel és teljes lemeztitkosítással lehet üzemeltetni, távoli törléssel. Így terepen sincs adatvesztés valamint elvesztés esetén azonnal törléssel minimalizálható a szenzitív adatok harmadik félhez kerülésének veszélye.

2.6.11 IT biztonsági fejlesztések minőség és projektmenedzsmentje

Ütemtervvel, mérföldkövekkel és tesztelési protokollokkal csökkenthető a hibázások aránya, illetve gyorsíthatja a döntés-előkészítést.

2.6.12 A jog szerepe a modern társadalmakban

A jog szükségesség-arányosság mércéjén kéri számon az adatvagyon felhasználásával készült elemzéseket: A legitim célhoz feltétlenül szükséges mennyiségű adatot kell felhasználni. A valós adatokkal való munka során, tehát a gyakorlatban ez kontrollált adatfelhasználást, modelldokumentációt és kötelező emberi döntési pontokat jelent, hogy az eszköz továbbra is eszköz maradjon, ne döntéshozó szerepet töltsön be. Továbbá az adatok megszerzésének, bárminemű felhasználásának módja igazodjon a GDPR-ban előírtaknak.

2.6.13 Komplex társadalomtudományi ismeretek

A katonák demográfiai vonatkozású adatainak az elemzése.

2.6.14 Kultúra, sport, munkahelyi jólét

A hosszabbítási hajlam pontosabb jelzéséhez a kohézió és a fizikai–mentális állapot mérhető változó, ezeket jóléti proxykként (sport-részvétel, hiányzás, elégedettség) lehet felvenni.

2.6.15 Matematikai alapok

Statisztikai mutatók (minimum, maximum, átlag, szórás, medián, módusz, kvartilis) alkalmazása a korrelációs mátrixok számításánál.

2.6.16 Mentori óra

Dr. Pitlik László tanár úrral való hatékony egyeztetések során hasznos információkat és tanácsokat kaptam a számítási feladatokhoz, valamint a szakdolgozat egyéb részeivel kapcsolatban.

2.6.17 Mesterséges intelligencia az IT biztonság területén

Mesterséges intelligencia használatával több képlet/függvény opció számbavétele az adott feladatnál, valamint a fellépő hibák hatékonyabb kezelésére.

2.6.18 Operációs rendszerek

Az Operációs rendszerek tárgyhoz kapcsolódó, hogy a számításokat Windows, macOS és iPadOS Excelben végeztem, a fájlokat felhőbe mentettem, így platformtól függetlenül ugyanonnan folytathattam a munkát. A környezetek eltérései nem okoztak gondot, leszámítva a Solver futást, hiszen ez a funkció iPadOS esetén nem elérhető.

2.6.19 Programozás

A programozás tantárgyak erősítették, fejlesztették az algoritmikus gondolkodást, amiket a számításaim során a hiba-függvények, aggregálások és a Solver futások esetében tudtam hasznosítani. A függvények rendszerezett használata pedig átláthatóvá tette a számításokat.

2.6.20 Programozási alapelvek és módszertanok

Az összeg/szorzat aggregátorváltás csak paramétercseré, így a képletek nem módosulnak, emellett védelmi szabályként a „HAHIBA” függvény használatával biztosítottam a stabilitást egy esetleges nagyon adathalmaz esetén.

2.6.21 Rendszermodellezés

A 20 Solver-futás eredményeképp kapott értékeket összesítő kísérlet tábla kapcsán KPI-alapú összevetést alkalmaztam, így nem egy, hanem az összes érték alapján történt az optimalizáció.

2.6.22 Rendszertervezés

Moduláris felépítésben készült a megoldás, tehát bemenet, feldolgozás, majd értékelés és kiválasztás. Az átláthatóságot szem előtt tartva egységes névkonvenciót használtam.

2.6.23 Szoftverarchitektúrák

A tárgyból hozott elvek szerint modulárisan épült fel az Excel. A rétegezés lehetővé tette a 6x3-mas tábla nagyobbra skálázhatóságát.

2.6.24 Szoftvertesztelés

A szoftvertesztelés logikájának mentén feltételes formázással ellenőriztem a küszöbértékeknek való megfelelést, valamint stabilitást a súlyértékek is adták.

2.6.25 Szoftverüzemeltetés

Üzemeltetésben már ismert monitoring analógia követése, ahol a KPI-panel folyamatos állapotjelzést adott vissza a követhetőség miatt.

2.6.26 Tudásmenedzsment az IT biztonság területén

A teljes elemzési folyamat dokumentálásával az eredmény megismételhető, valamint átadható későbbi felhasználásra.

2.6.27 Vállalati gazdaságtan

Az elemzés során aggregátorok és súlyok mellett figyeltem, hogy melyik megoldás jó akkor is, ha a feltételek változnak, ez közel áll egy valós üzleti döntéshez, ahol nem csupán egy, hanem több szempontnak is meg kell felelni.

2.6.28 Vezetési és vállalkozási ismeretek

A vezetési alapelvek közül a prioritizálás és az átláthatóság jelenik meg, ahol a súlyok döntések a KPI-k pedig követhetővé teszik az eredményt, alkalmazható döntéstámogatásra is.

3. fejezet: Saját fejlesztés bemutatása a szerződéshosszabbítás-elemzéshez

3.1. Alapadatok

A kutatásomban jelenleg 6x3-mas táblázatban láthatók adatok. Ezek az adatok soronként személyeket jelképeznek, oszlopok szerint három attribútummal. Természetesen a dolgozatomban már említett személyügyi adatbázissal is működik a bemutatott módszer. A cél táblában található értékek megjelenítése a szemléltetés, valamint a visszaellenőrzés, tehát a számítási eredmények helyességének meghatározása céljából történt. A számítási modell az eredményeket az elérendő értékek ismerete nélkül – mivel ezek az adatok megsemmisülés miatt nem állnak rendelkezésre - adja vissza, aminek jelentősége a fő kérdéshez kapcsolódik: kizárólag statisztikai mutatók ismeretében, hogyan szimulálható valósághoz közelítő adatvagyon? Az átláthatóságot szem előtt tartva dolgoztam kevesebb adatokkal. Az értékek egységen mindenhol 0 és 1 között jelennek meg, ezzel segítve a számítások átláthatóságát, könnyebb értelmezését és segítséget nyújt a valós kép kialakításában. A kialakítás szerint skálázható nagyobb táblákra is, például 60x30, ez azt jelenti, hogy a logika és a szabályok változtatás nélkül továbbvihetők, alkalmazhatók.

adatok1 - cél											
max (pontoszm)	0,60	0,70	0,90								
min (pontoszm)	0,10	0,10	0,10								
átlag (pontoszm)	0,35	0,45	0,58								
szórás (pontoszm)	0,24	0,23	0,31								
medián (pontoszm)	0,35	0,45	0,70								
módusz (pontoszm)	0,10	0,70	0,70								
kvartilis Q1 (pontoszm)	0,13	0,33	0,40								
kvartilis Q2 (pontoszm)	0,35	0,45	0,70								
kvartilis Q3 (pontoszm)	0,58	0,65	0,78								
A (pontoszm)	B (pontoszm)	C (pontoszm)	max (pontoszm)	min (pontoszm)	átlag (pontoszm)	szórás (pontoszm)	medián (pontoszm)	módusz (pontoszm)	kvartilis Q1 (pontoszm)	kvartilis Q2 (pontoszm)	kvartilis Q3 (pontoszm)
1	0,10	0,40	0,90	0,90	0,10	0,47	0,40	0,40	0,25	0,40	0,65
2	0,60	0,10	0,30	0,60	0,10	0,33	0,25	0,30	0,30	0,30	0,45
3	0,50	0,50	0,80	0,80	0,50	0,60	0,17	0,50	0,50	0,50	0,65
4	0,20	0,70	0,70	0,70	0,20	0,53	0,29	0,70	0,70	0,45	0,70
5	0,60	0,30	0,10	0,60	0,10	0,33	0,25	0,30	0,20	0,30	0,45
6	0,10	0,70	0,70	0,70	0,10	0,50	0,35	0,70	0,70	0,40	0,70

2. ábra adatok1 – cél táblázat

Saját forrás: „tablázat” munkalap A1-M18 tartománya / Jelmagyarázat: táblában látható az elérendő cél, valamint a későbbi becsléshez szükséges statisztikai mutatók

3.1.1. Az adatvagyon koncepciója és szükségessége

Az adatvagyon egy olyan gondosan feldolgozott, adattisztításon átesett és strukturált információhalmaz, amely segíti a stratégiai szintű döntéshozatalt, alkalmas előrejelzésre, valamint nagy szerepet játszik a szervezeti fejlesztésekben.

A Magyar Honvédség vonatkozásában jelen esetben az adatvagyon a szerződéses állomány szerződéshosszabbítási hajlam előrejelzését segítő információ tömeget jelenti. Ezen halmaz

adatgyűjtési és feldolgozási folyamatának sikerességétől függ, hogy a COCO, OAM modell mennyire ad megbízható és pontos eredményt az előrejelzésnél, hogy valójában stratégiai szinten a döntéshozatal tényleges pozitív eredményt hozzon.

Az adatvagyonnal történő munkafolyamatnak három kritikus fázisa állapítható meg.

3.1.2. Adatgyűjtési fázis

Az adatgyűjtés során más-más forrásokból kerülnek összegyűjtésre az adatok, információk, melyek ebben a fázisban még nem látható módon, de összefüggnek a szerződeshosszabbítási hajlandósággal. Az adatok származhatnak belső forrásokból, ebben az esetben a Magyar Honvédség által nyilvántartott személyügyi adatbázisból, ahol megjelennek az alapvető demográfiai és szolgálati adatok egyénekre bontva. Ebben az adatbázisban található a teljes katonai állomány adatai. A teljesség igénye nélkül jellemzően az alábbiakat tartalmazza:

- név
- születési idő
- nem
- jelenlegi lakóhely
- családi állapot
- végzettség szintje
- nyelvismeret
- szolgálati idő kezdete
- eddigi és jelenlegi beosztás
- szolgálati idő
- külszolgálatban eltöltött idő

Ezek mind-mind olyan hasznos adatok, melyek bevonásával még tisztább képet mutathat az előrejelzési folyamat eredménye. Természetesen kiegészíthető akár teljesítményértékelés során megállapított és rögzített adatokkal, ahol kimutatható az egyén munkahelyi teljesítménye az aktuális beosztásában. Ez már most előrevetíti azt, hogy az alacsonyabb értékelési szinttel rendelkező katona motivációja valószínűleg már nem elegendő egy szerződeshosszabbítás aláírásához.

A belső források feldolgozása után külső forrásokból is nyerhető hasznos adat, amelyekkel érdemes kalkulálni a számítások során. Ilyen lehet például egy közösségi oldalra feltöltött információ, konkrétan egy LinkedIn-profil, ahonnan megismerhető, hogy a katonák leszerelést követően jellemzően milyen munkahelyen tudtak elhelyezkedni, vagyis kontextust ad a munkaerőpiaci kereslet megértéséhez.

3.1.3. Az adattisztítás és validálás fázisa

A sikeres adatgyűjtést követően az információ szennyezettsége és megbízhatatlansága feloldandó feladat. Egy valós adathalmazban törvényszerű, hogy előfordulnak hiányzó adatok. Ennek oka lehet akár adminisztratív mulasztás vagy adatmigráció során elcsúszott adatok. Figyelembe kell venni, hogy az egészhez viszonyítva hány hiányzó részről van szó, hiszen egy rekordeltávolítás esetén információveszteség állhat fenn. Amennyiben a hiányzó adatok aránya elég magas egy téves eredmény megjelenéséhez, akkor becslésen alapuló pótlás szükséges, hogy ne torzuljon a kép.

Az imputáció lehet átlag-imputáció, ahol a hiányzó értékek helye került feltöltésre az adott attribútum átlagával. Ez viszonylag egyszerű megoldása a pótlásnak, de előfordulhat, hogy az adatok varianciáját csökkenti. Az átlag-imputációhoz hasonló módszer a medián-imputáció, amely a medián értéket veszi alapul. A hiányzó adatok kiemelten magas arányt mutatnak, ebben az esetben egyetlen megoldás az elemzésből való kizárás, hiszen teljesen megbízhatatlan eredményeket adna.

3.2. Számítások magyarázata

A modell felépítésekor első lépés a “valóság” szétválasztása volt, hiszen jelen esetben az elérendő értékek informatív és ellenőrzési jelleggel jelennek meg, mivel adatvesztés miatt az elemzés során nincsenek valós adatok, ennek megfelelően hoztam létre a cél és a becslés táblákat. A cél tábla egyfajta horgonyként látja el feladatát, hiszen ebben találhatóak a kiinduló statisztikai mutatók, vagyis a minimum, maximum, átlag, szórás, medián és kvartilisek. Ezek előzetesen meghatározott statisztikai célértékek. A becslés tábla az a változó rács, amiben a véletlenszámokkal feltöltött adatokat hasonlítottam össze a célmutatóhoz viszonyítva. Ezek az értékek a már említett statisztikai mutatók alapján lett “korlátok” közé beszorítva, hogy a megfelelő tartományból legyenek a véletlenszámok megjelenítve. Ezzel biztosítottam a realitást, hogy az értékek ne mutassanak túl a tartományon, de mégis megtartva egyfajta szabadságot, hogy a Solvernek legyen min javítania. A becslés tábla statisztikai mutatói közül a módusz esetében “HAHIBA” függvényt is alkalmaztam arra az esetre, ha a hiba keletkezne, de a számítások ne akadjanak el és a Solver stabilan fusson. Főbb előnye, hogy egyértelműen azonosítható, hogy mi a cél és mit próbálok ehhez az értékhez igazítani. A cél tábla mutatóit akár olyan kérdésekhez is lehet igazítani, mint a “mi történne, ha az átlagos elégedettség emelkedést mutatna?”.

adatok2 - becsült érték												
max (pontszám)	0,57	0,67	0,75									
min (pontszám)	0,11	0,40	0,25									
átlag (pontszám)	0,37	0,55	0,59									
szórás (pontszám)	0,20	0,10	0,19									
medián (pontszám)	0,40	0,58	0,65									
módusz (pontszám)	0,40	0,60	0,70									
kvartilis Q1 (pontszám)	0,21	0,47	0,54									
kvartilis Q2 (pontszám)	0,40	0,58	0,65									
kvartilis Q3 (pontszám)	0,53	0,61	0,72									
	A (pontszám)	B (pontszám)	C (pontszám)	max (pontszám)	min (pontszám)	átlag (pontszám)	szórás (pontszám)	medián (pontszám)	módusz (pontszám)	kvartilis Q1 (pontszám)	kvartilis Q2 (pontszám)	kvartilis Q3 (pontszám)
1	0,11	0,40	0,73	0,73	0,11	0,41	0,31	0,40	0,40	0,26	0,40	0,57
2	0,37	0,56	0,25	0,56	0,25	0,39	0,16	0,37	0,40	0,31	0,37	0,47
3	0,44	0,62	0,53	0,62	0,44	0,53	0,09	0,53	0,50	0,48	0,53	0,57
4	0,57	0,44	0,60	0,60	0,44	0,54	0,08	0,57	0,60	0,50	0,57	0,58
5	0,16	0,67	0,70	0,70	0,16	0,51	0,31	0,67	0,70	0,41	0,67	0,69
6	0,56	0,59	0,75	0,75	0,56	0,64	0,10	0,59	0,60	0,58	0,59	0,67

3. ábra adatok2 – becsült érték

Saját forrás: „tablázat” munkalap A28-M45 tartománya / Jelmagyarázat: a tábla 6x3 részében az 1. ábra szerinti statisztikai mutatók korlátai közé szorított véletlenszámok jelennek meg a későbbi feldolgozáshoz, valamint az oszlopra és sorokra vonatkozó statisztikai adatok

3.2.1. Korrelációs mátrixok

A korrelációs mátrixok szükségessége abban rejlik, hogy téves gondolatot ébreszthet, ha az illesztés csak külön-külön megfelelő az oszlopokra. A lényeg éppen az, hogy a mátrix megmutatja az oszloppárok együttjárását, vagyis az egyik adat változása milyen hatással van a többi adatra. Ha kizárólag a statisztikai mutatók szerint hasonlított volna becsült érték a célértékekhez, akkor előfordulhat, hogy olyan becslés eredményét kapom vissza, amely bár eloszlásban jó, de a szerkezete teljesen eltérő, ezért fontos az oszloponkénti korreláció.

adatok1 - korrelációs mátrix			
	A (pontszám)	B (pontszám)	C (pontszám)
A	1,00	-0,72	-0,75
B	-0,72	1,00	0,61
C	-0,75	0,61	1,00

4. ábra adatok1 – korrelációs mátrix

Saját forrás: „tablázat” munkalap A20-D25 tartománya / Jelmagyarázat: megmutatja, hogy egyes változók mekkora mértékben és milyen irányba mozognak egymáshoz viszonyítva

adatok2 - korrelációs mátrix			
	A (pontszám)	B (pontszám)	C (pontszám)
A	1,00	-0,03	-0,21
B	-0,03	1,00	0,28
C	-0,21	0,28	1,00

5. ábra adatok2 – korrelációs mátrix

Saját forrás: „tablazat” munkalap A20-D25 tartománya / Jelmagyarázat: megmutatja, hogy egyes változók mekkora mértékben és milyen irányba mozognak egymáshoz viszonyítva

3.2.2. Korrelációs eltérések elemzése

A hiba1 értékét a két korrelációs tábla, tehát a cél és a becült tábla elemenkénti különbségeinek az eredménye adja.

adatok1 és adatok2 korrelációs mátrix közti eltérés			
	A (pontszám)	B (pontszám)	C (pontszám)
A	0,0E+00	4,8E-01	2,9E-01
B	4,8E-01	0,0E+00	1,1E-01
C	2,9E-01	1,1E-01	4,9E-32
hiba1 (pontszám)	1,75		

6. ábra adatok1 és adatok2 korrelációs mátrix eltérés

Saját forrás: „tablazat” munkalap A54-D67 tartománya / Jelmagyarázat: megmutatja a cél tábla mátrix és a becült tábla mátrix közötti eltérést

A hiba2 az oszlopok statisztikai profiljára helyezte a fókuszot, minden oszlopra a leíró statisztikákat a cél és a becült táblák vonatkozásában számítottam ki és ezek különbségének szorzatösszege adja eredményül a hiba2-t. Ezzel biztosítva lett, hogy a becült tábla esetében ne csak az átlag értéke legyen megfelelő határon belül, hanem az eloszlás széle is megközelítse a célértéket.

oszlop elteres minmax.... adat-adat2			
	A (pontszám)	B (pontszám)	C (pontszám)
max	0,03	0,20	0,01
min	-0,21	-0,03	-0,14
átlag	-0,13	0,16	-0,03
szórás	0,13	0,09	0,05
medián	-0,18	0,16	0,04
módusz	0,40	-0,40	0,20
kvartilis Q1	-0,28	0,15	-0,02
kvartilis Q2	-0,18	0,16	0,04
kvartilis Q3	0,03	0,29	-0,06

hiba2 (pontszám)	0,85
------------------	------

7. ábra adatok1 és adatok2 korrelációs mátrix eltérés oszlopok szerint

Saját forrás: „tablazat” munkalap F54-I67 tartománya / Jelmagyarázat: megmutatja a cél tábla mátrix és a becsült tábla statisztikai mutatóinak oszlop szerinti eltérései

A hiba3 a hiba2-nél már ismert módszer szerinti számításokat végeztem el, de itt nem az oszloponkénti, hanem soronkénti számítással. Ennek célja az volt, hogy az egyes sorok is valóságot megközelítő mintázatot mutassanak.

sor elteres minmax.... adatok-adatok2						
	1 (pontszám)	2 (pontszám)	3 (pontszám)	4 (pontszám)	5 (pontszám)	6 (pontszám)
max	0,28	-0,09	0,26	-0,18	0,03	-0,19
min	-0,15	-0,27	0,37	0,06	-0,14	-0,40
átlag	0,07	-0,15	0,26	0,02	-0,05	-0,15
szórás	0,20	0,07	-0,03	-0,08	0,08	0,14
medián	0,09	-0,08	0,14	0,17	-0,03	0,16
módusz	0,10	-0,10	0,10	0,20	0,00	0,20
kvartilis Q1	-0,03	-0,17	0,26	0,12	-0,08	-0,12
kvartilis Q2	0,09	-0,08	0,14	0,17	-0,03	0,16
kvartilis Q3	0,18	-0,08	0,20	0,00	0,00	-0,02
hiba3 (pontszám)						
1,34						

8. ábra adatok1 és adatok2 korrelációs mátrix eltérés sorok szerint

Saját forrás: „tablazat” munkalap K54-Q67 tartománya / Jelmagyarázat: megmutatja a cél tábla mátrix és a becsült tábla statisztikai mutatóinak sor szerinti eltéréseit

3.2.3. A hibák normalizálása és tűrések értelmezése

A három hiba értéke egymáshoz viszonyítva akkor mutat valós képet, ha azonos súlyozással jelennek meg a számításokban. Ehhez hibánként kiszámoltam egy olyan maximum értéket, amely felső határként funkcionál. Ezután a három hiba értékeit a saját maximumához való viszonyítással egy-egy százalékos érték jelent meg, így ez a normalizálás lehetővé tette a súlyozott aggregálást.

hiba123 maximuma, azonos nagyságrendre hozása		
hiba1max (pontszám)	8,65	0,20
hiba2max (pontszám)	1,80	0,47
hiba3max (pontszám)	3,90	0,34

9. ábra hiba értékek maximum meghatározása és azonos súlyozás beállítása

Saját forrás: „tablazat” munkalap A73-C76 tartománya / Jelmagyarázat: A korábban említett számítással megállapított értékek közötti eltérést adó hiba1, hiba2 és hiba3 maximumának megjelenítése és azonos nagyságrendre hozása

3.2.4. A túrések jelentősége

A túrés a számításokban két szerepet is betölt, egyrészt a korrelációs eltérésekben, ahol a 0.2-es túrés az elfogadható különbséget jelenti, másrészt a cellák eltéréséhez használt feltételes formázás alapját adta. A kapott értékek olyan küszöbértékek, amik megmutatják a cellák közti eltérések fokát. Ez a fajta vizuális megjelenítés segítséget nyújt a könnyebb értelmezés érdekében.

A feltételes formázás az alábbi küszöbértékek szerint működik:

1. Zöld: ha az eltérés ≤ 0.2 , akkor elfogadható pontosságú,
2. Sárga: ha a $0.2 < X \leq 0.4$, akkor közepes erősségű az eltérés,
3. Piros: ha az eltérés $0.4 >$, akkor nagy eltérésről van szó.

Az értékek szerinti feltételes formázást használva létrehoztam a cél és becslő értékek közti eltérést megállapító táblázatot.

cél és becslő értékek különbsége (adatok-adatok2)			
	A (pontszám)	B (pontszám)	C (pontszám)
1	-0,21	0,15	0,28
2	0,23	-0,28	-0,39
3	-0,04	0,37	0,44
4	-0,33	0,56	-0,18
5	0,03	-0,03	-0,14
6	-0,44	0,20	-0,19
színmagyarázat			
0-0,2	0,2-0,4	0,4 -	

10. ábra a cél tábla és a becslő tábla értékeinek különbsége

Saját forrás: „táblázat” munkalap A80-h87 tartománya / Jelmagyarázat: A két tábla közti küszöbök szerinti különbség megjelenítése színnel

3.2.5. Solver

A feladat lényege egy olyan nem lineáris és többcélú optimalizálás létrehozása, használata, ami egyszerre képes javítani a hiba1, hiba2 és hiba3 értékein, miközben megfelel a statisztikai korlátok határértékeinek. Az Excel Solver pontosan alkalmas arra, hogy sok változás és nem lineáris függvényt minimalizáljon a megadott korlátok szerint. A célfüggvényem egy aggregált hiba volt szorzat vagy összeg formájában. Mivel a Solver egy munkalapon képes futni, ezért

minden egyes Solver futtatást külön munkalapon végeztem el, amelyekre természetesen az értékek összevetésénél és összegzésénél sikeresen lehet hivatkozni. A változó cellák mindig a becsült tábla adatai voltak, nem szabad a cél táblát használni, hiszen ekkor előfordulhat, hogy saját magunkat csapjuk be a vonzónak vélt értékekkel. Természetesen minden külön munkalap felépítése azonos, vagyis található egy cél tábla, egy becsült tábla, a korrelációs mátrixot, azok eltérései, valamint a statisztikai mutatók eltérései is, megegyezően hiba1, hiba2 és hiba3 megjelenésével. Ezek a későbbiekben látható húsz soros táblázat értékeit adják, eltérő aggregátorral, súlyozással és az ez alapján futtatott Solveres értékekkel. A Solver munkalapokon a becsült tábla kezdeti értékei mindig véletlen számok voltak. Ez a megoldás közelebb áll egy Monte Carlo jellegű feltáráshoz, mint egy egyszerű optimalizálási folyamathoz.

3.2.6. 20 soros kísérleti panel

Minden sor egy Solver futtatást jelent, összeg és szorzat aggregátorral, valamint változó súlyokkal. Mindegyik sor tartalmaz statisztikai adatot, tehát minimum, maximum, átlag, szórás, medián, módusz és kvartilisek. A Solver az adott célfüggvény szerint próbálta célhoz igazítani a becsült értékeket. A 20 kísérlet eredményét az összehasonlítás érdekében a KPI-ket 0 és 1 közé normalizáltam. Súlyozásokat tekintve alkalmaztam 1:1:1 arányban is, hogy egyenlő súlyokkal kapjak értékeket, valamint eltérő, például 2:1:1 vagy 1:1:2, de a hangsúlyozás érdekében megfigyelhető 1:3:1 is. Ez a panel megmutatja, hogy „lehet-e minden hiba-aggregálás másként egyformán perspektivikus?” vagyis: ha máshogy súlyozok / mást tekintünk célnak, ugyanaz a megoldás marad-e „legjobb”?

ID	Aggregátor	Súly_hiba1	Súly_hiba2	Súly_hiba3	Zöld db	Hibátlan %	Min (absz.) (pontoszám)	Max (absz.) (pontoszám)	Átlag (absz.) (pontoszám)	Szórás (absz.) (pontoszám)	Medián (absz.) (pontoszám)	Módusz (absz.) (pontoszám)	Q1 (absz.) (pontoszám)	Q3 (absz.) (pontoszám)	Aktiv cél (pontoszám)
1	összeg	1	1	1	1	0	0,04	0,49	0,26	0,16	0,28	0,10	0,08	0,40	0,87
hiba1+hiba2+hiba3 (pontoszám)		0,87		solver											

11. ábra Solver futtatással elért értékek szerinti statisztikai mutatók eltérései

Saját forrás: „tablazat1” munkalap A90-P91 tartománya / Jelmagyarázat: Solver futtatást követően kapott értékek összehasonlítása a cél és becsült táblák vonatkozásában bizonyos aggregátorral és bizonyos súlyozás szerint

A “zöld db” megmutatja, hogy hány cella került túrésen belüli, azaz az eltérés maximum 0.2 volt. A minimum és maximum abszolút eltérés a cél és becsült táblák közötti legkisebb és legnagyobb értékét jelzi.

A 20 soros kísérlet szerint megállapítható, hogy az eltérő hiba-aggregálások, tehát különböző hiba1, hiba2 és hiba3 súlyok, valamint összeg vagy szorzat aggregátor értékek nem egyformán perspektivikusak. Nem automatikusan az a sor tekinthető a legerősebbnek, ahol a feltételes formázás szerint is a legtöbb zöld, tehát a legtöbb célhoz közeli érték jelenik meg.

Leggyengébb KPI maximalizálásra azért volt szükség, hogy olyat megoldást találjak, ahol a leggyengébb láncszem is viszonylag erős. Ebből következik, ha van egy viszonylag súlyosnak számító gyenge pont, akkor hiába mutatkozik átlagosan jónak a kép, az egész sebezhető marad.

ID	Aggregátor	Súly_hiba1	Súly_hiba2	Súly_hiba3	Zöld db	Hibátlan %	Min (absz.) (pontszám)	Max (absz.) (pontszám)	Átlag (absz.) (pontszám)	Szórás (absz.) (pontszám)	Medián (absz.) (pontszám)	Módusz (absz.) (pontszám)	Q1 (absz.) (pontszám)	Q3 (absz.) (pontszám)	Aktív cél (pontszám)	leggyengébb KPI (pontszám)
1	összeg	1	1	1	7	0	0,04	0,49	0,26	0,16	0,29	0,10	0,08	0,40	0,87	0,22
2	szorzat	1	1	1	14	0	0,01	0,48	0,12	0,15	0,05	0,00	0,01	0,11	0,49	0,17
3	összeg	2	1	1	12	0	0,01	0,42	0,16	0,13	0,11	0,00	0,04	0,27	0,36	0,26
4	szorzat	2	1	1	8	0	0,01	0,61	0,24	0,19	0,22	0,00	0,07	0,41	3,69	0,19
5	összeg	1	2	1	7	0	0,01	0,85	0,34	0,24	0,36	0,10	0,09	0,50	1,95	0,00
6	szorzat	1	2	1	11	0	0,01	0,48	0,18	0,17	0,14	0,00	0,03	0,32	0,95	0,14
7	összeg	1	1	2	13	0	0,01	0,45	0,14	0,14	0,08	0,00	0,04	0,19	0,44	0,17
8	szorzat	1	1	2	11	0	0,00	0,65	0,23	0,20	0,13	0,10	0,08	0,41	3,95	0,02
9	összeg	2	2	1	10	0	0,01	0,55	0,20	0,17	0,13	0,10	0,06	0,30	0,96	0,25
10	szorzat	2	2	1	5	0	0,01	0,58	0,33	0,20	0,38	0,50	0,11	0,50	2,03	0,00
11	összeg	2	1	2	7	0	0,01	0,64	0,27	0,20	0,27	0,00	0,06	0,46	1,55	0,13
12	szorzat	2	1	2	8	0	0,03	0,75	0,31	0,22	0,23	0,10	0,11	0,53	6,31	0,00
13	összeg	1	2	2	9	0	0,00	0,52	0,23	0,17	0,20	0,00	0,07	0,36	2,00	0,00
14	szorzat	1	2	2	5	0	0,04	0,66	0,33	0,20	0,30	0,30	0,15	0,48	6,62	0,00
15	összeg	3	1	1	12	0	0,01	0,49	0,19	0,16	0,14	0,10	0,05	0,30	1,02	0,17
16	szorzat	3	1	1	5	0	0,00	0,59	0,32	0,16	0,34	0,30	0,19	0,46	2,28	0,00
17	összeg	1	3	1	6	0	0,06	0,54	0,27	0,17	0,26	0,10	0,11	0,42	1,82	0,25
18	szorzat	1	3	1	9	0	0,02	0,57	0,24	0,18	0,20	0,20	0,15	0,32	2,12	0,25
19	összeg	1	1	3	9	0	0,01	0,55	0,21	0,20	0,16	0,00	0,03	0,37	1,56	0,19
20	szorzat	1	1	3	10	0	0,01	0,53	0,20	0,19	0,11	0,00	0,03	0,31	1,92	0,15

12. ábra 20 soros kísérlet tábla, Solver futások
Saját forrás: „tablazat1” munkalap A90-Q110 tartománya / Jelmagyarázat: Minden Solver által kapott értékek megjelenítése egy táblázatban és a soronkénti leggyengébb KPI-k meghatározása

3.2.7. COCO Y0-modell alkalmazása

Mivel nem csak a statisztikai illeszkedés volt a fő kérdés, hogy a becsült adatok mennyire hasonlítanak a célhoz, hanem az is, hogy az attribútumok mennyire “magyarázzák” az Y-változót, jelen esetben tekintsük szerződeshosszabbítási hajlandóságnak. A COCO szerinti elemzéshez alapvetően szükség van egy OAM-ra, ez a 20 soros kísérlet táblázat. A módszer lényegi része, hogy a 20 kísérletet a leíró attribútumok rangsora alapján hasonlítja össze, ezzel megmutatja az egymáshoz viszonyított helyét. Megállapítható a módszerrel, hogy a 20 kísérlet közül meg a céladatokhoz legjobban illeszkedő és melyikben látható az a mintázat, ami a szerződeshosszabbítási hajlandóságot mutatja. Minden attribútumhoz tartozik egy irány, ami 0 vagy 1. A 0, tehát minél nagyobb, annál jobb, például zöld db. Az 1 pedig a minél kisebb, annál jobb, például minimum eltérés. Az irány azért jelentős, mert a későbbiekben eszerint lehet a rangsort felállítani.

Számításaim során a COCO Y0-változatot alkalmaztam, ami anti-diszkriminációs modell. A működése szerint az Y0 vizsgálja azt, hogy “lehet-e minden objektum másként egyforma”. Pontosabban, minden az ábrán látható X-értékhez egy lépcsős függvény tartozik, amely a rangsorhoz pontszámot rendel, attribútumonként ceteris paribus szabályelv szerint, tehát szakaszokkal való megközelítésről van szó.

Rangsor	X(A1)	X(A2)	X(A3)	X(A4)	X(A5)	X(A6)	X(A7)	X(A8)	X(A9)	Y(A10)
O1	19	6	13	6	16	10	12	11	4	1000000
O2	8	3	1	3	1	1	1	1	3	1000000
O3	15	1	3	1	3	1	6	3	1	1000000
O4	12	15	12	13	12	1	10	13	17	1000000
O5	10	20	20	20	19	10	14	18	12	1000000
O6	5	4	4	8	7	1	2	7	5	1000000
O7	9	2	2	2	2	1	5	2	2	1000000
O8	2	17	10	18	6	10	13	12	18	1000000
O9	14	10	7	9	5	10	8	4	6	1000000
O10	13	13	18	17	20	20	17	19	14	1000000
O11	4	16	14	14	15	1	9	16	8	1000000
O12	17	19	16	19	13	10	15	20	19	1000000
O13	1	7	9	10	11	1	11	9	13	1000000
O14	18	18	19	16	17	18	19	17	20	1000000
O15	7	5	5	5	8	10	7	5	7	1000000
O16	3	14	17	4	18	18	20	15	16	1000000
O17	20	9	15	7	14	10	16	14	10	1000000
O18	16	12	11	11	10	17	18	8	15	1000000
O19	11	11	8	15	9	1	3	10	9	1000000
O20	6	8	6	12	4	1	4	6	11	1000000

13. ábra COCO-módszer szerinti rangsorolás

Saját forrás: „Munka1” munkalap A7-K27 tartománya / Jelmagyarázat: O1-O20 sorok a korábban számított kísérleti tábla értékeit tartalmazza, valamint Y-érték, amelynek segítségével definiálhatók a rangsorok

A lépcsők nagyságát tekintve az a megkötés vehető észre, hogy a következő lépcső mindig legalább 1 egységgel rosszabb legyen az előzőnél. Az Y0 értéke minden esetben azonos, tehát nem változik. A Solver az optimalizálás szerint keresi azokat az értékeket, amelyeknél a becslés a legjobban illeszkedik az Y0-hoz. Ez nem egy egyszerű regressziós becslés, hanem egy lépcsős, rangsor alapú skála. A számítások során, minél kisebb volt a delta, annál közelebbi állapotot jelent az adott kísérlet az ideális viszonyokhoz.

Lépcsők(1)	X(A1)	X(A2)	X(A3)	X(A4)	X(A5)	X(A6)	X(A7)	X(A8)	X(A9)
S1	$(69+66)/(2)=67.5$	$(31+19)/(2)=25$	$(19+19)/(2)=19$	$(19+33)/(2)=26$	$(999858.3+24)/(2)=499941.15$	$(19+19)/(2)=19$	$(19+999932.3)/(2)=499975.65$	$(19+19)/(2)=19$	$(19+19)/(2)=19$
S2	$(68+65)/(2)=66.5$	$(23+18)/(2)=20.5$	$(18+18)/(2)=18$	$(18+32)/(2)=25$	$(999857.3+23)/(2)=499940.15$	$(18+18)/(2)=18$	$(18+999931.3)/(2)=499974.65$	$(18+18)/(2)=18$	$(18+18)/(2)=18$
S3	$(67+64)/(2)=65.5$	$(17+17)/(2)=17$	$(17+17)/(2)=17$	$(17+31)/(2)=24$	$(999856.3+22)/(2)=499939.15$	$(17+17)/(2)=17$	$(17+999930.3)/(2)=499973.65$	$(17+17)/(2)=17$	$(17+17)/(2)=17$
S4	$(66+63)/(2)=64.5$	$(16+16)/(2)=16$	$(16+16)/(2)=16$	$(16+30)/(2)=23$	$(999855.3+21)/(2)=499938.15$	$(16+16)/(2)=16$	$(16+999929.3)/(2)=499972.65$	$(16+16)/(2)=16$	$(16+16)/(2)=16$
S5	$(36+62)/(2)=49$	$(15+15)/(2)=15$	$(15+15)/(2)=15$	$(15+29)/(2)=22$	$(999854.3+20)/(2)=499937.15$	$(15+15)/(2)=15$	$(15+999928.3)/(2)=499971.65$	$(15+15)/(2)=15$	$(15+15)/(2)=15$
S6	$(35+61)/(2)=48$	$(14+14)/(2)=14$	$(14+14)/(2)=14$	$(14+28)/(2)=21$	$(999853.3+19)/(2)=499936.15$	$(14+14)/(2)=14$	$(14+999927.3)/(2)=499970.65$	$(14+14)/(2)=14$	$(14+14)/(2)=14$
S7	$(34+60)/(2)=47$	$(13+13)/(2)=13$	$(13+13)/(2)=13$	$(13+27)/(2)=20$	$(999852.3+18)/(2)=499935.15$	$(13+13)/(2)=13$	$(13+999926.3)/(2)=499969.65$	$(13+13)/(2)=13$	$(13+13)/(2)=13$
S8	$(12+59)/(2)=35.5$	$(12+12)/(2)=12$	$(12+12)/(2)=12$	$(12+26)/(2)=19$	$(999851.3+17)/(2)=499934.15$	$(12+12)/(2)=12$	$(12+999925.3)/(2)=499968.65$	$(12+12)/(2)=12$	$(12+12)/(2)=12$
S9	$(11+58)/(2)=34.5$	$(11+11)/(2)=11$	$(11+11)/(2)=11$	$(11+25)/(2)=18$	$(999850.3+16)/(2)=499933.15$	$(11+11)/(2)=11$	$(11+999924.3)/(2)=499967.65$	$(11+11)/(2)=11$	$(11+11)/(2)=11$
S10	$(10+57)/(2)=33.5$	$(10+10)/(2)=10$	$(10+10)/(2)=10$	$(10+24)/(2)=17$	$(999849.3+15)/(2)=499932.15$	$(10+10)/(2)=10$	$(10+999923.3)/(2)=499966.65$	$(10+10)/(2)=10$	$(10+10)/(2)=10$
S11	$(9+56)/(2)=32.5$	$(9+9)/(2)=9$	$(9+9)/(2)=9$	$(9+23)/(2)=16$	$(999848.3+14)/(2)=499931.15$	$(9+9)/(2)=9$	$(9+999922.3)/(2)=499965.65$	$(9+9)/(2)=9$	$(9+9)/(2)=9$
S12	$(8+55)/(2)=31.5$	$(8+8)/(2)=8$	$(8+8)/(2)=8$	$(8+22)/(2)=15$	$(999847.3+13)/(2)=499930.15$	$(8+8)/(2)=8$	$(8+999921.3)/(2)=499964.65$	$(8+8)/(2)=8$	$(8+8)/(2)=8$
S13	$(7+54)/(2)=30.5$	$(7+7)/(2)=7$	$(7+7)/(2)=7$	$(7+21)/(2)=14$	$(999846.3+12)/(2)=499929.15$	$(7+7)/(2)=7$	$(7+999920.3)/(2)=499963.65$	$(7+7)/(2)=7$	$(7+7)/(2)=7$
S14	$(6+53)/(2)=29.5$	$(6+6)/(2)=6$	$(6+6)/(2)=6$	$(6+20)/(2)=13$	$(999845.3+11)/(2)=499928.15$	$(6+6)/(2)=6$	$(6+999919.3)/(2)=499962.65$	$(6+6)/(2)=6$	$(6+6)/(2)=6$
S15	$(5+52)/(2)=28.5$	$(5+5)/(2)=5$	$(5+5)/(2)=5$	$(5+19)/(2)=12$	$(999844.3+10)/(2)=499927.15$	$(5+5)/(2)=5$	$(5+999918.3)/(2)=499961.65$	$(5+5)/(2)=5$	$(5+5)/(2)=5$
S16	$(4+51)/(2)=27.5$	$(4+4)/(2)=4$	$(4+4)/(2)=4$	$(4+18)/(2)=11$	$(999843.3+9)/(2)=499926.15$	$(4+4)/(2)=4$	$(4+999917.3)/(2)=499960.65$	$(4+4)/(2)=4$	$(4+4)/(2)=4$
S17	$(3+50)/(2)=26.5$	$(3+3)/(2)=3$	$(3+3)/(2)=3$	$(3+11)/(2)=7$	$(999842.3+8)/(2)=499925.15$	$(3+3)/(2)=3$	$(3+999916.3)/(2)=499959.65$	$(3+3)/(2)=3$	$(3+3)/(2)=3$
S18	$(2+49)/(2)=25.5$	$(2+2)/(2)=2$	$(2+2)/(2)=2$	$(2+2)/(2)=2$	$(999841.3+2)/(2)=499921.65$	$(2+2)/(2)=2$	$(2+999915.3)/(2)=499958.65$	$(2+2)/(2)=2$	$(2+2)/(2)=2$
S19	$(1+1)/(2)=1$	$(1+1)/(2)=1$	$(1+1)/(2)=1$	$(1+1)/(2)=1$	$(999840.3+1)/(2)=499920.65$	$(1+1)/(2)=1$	$(1+999914.3)/(2)=499957.65$	$(1+1)/(2)=1$	$(1+1)/(2)=1$
S20	$(0+0)/(2)=0$	$(0+0)/(2)=0$	$(0+0)/(2)=0$	$(0+0)/(2)=0$	$(999839.3+0)/(2)=499919.65$	$(0+0)/(2)=0$	$(0+999881.3)/(2)=499940.65$	$(0+0)/(2)=0$	$(0+0)/(2)=0$

14. ábra COCO-módszer szerinti rangsorolás

Saját forrás: „Munka1” munkalap A7-K27 tartománya / Jelmagyarázat: O1-O20 sorok a korábban számított kísérleti tábla értékeit tartalmazza, valamint Y-érték, amelynek segítségével definiálhatók a rangsorok

A COCO-módszerrel tehát a 20 kísérletnél nem csak a hibák és azok mértéke, hanem rangsor- és pontszám-szintű kép is láthatóvá vált. Abban az esetben, ha több olyan kísérleti sor van, amelyek közel vagy szinte azonos hibaértéket mutatnak, közöttük a COCO alapján lehet különbséget tenni, vagyis hol volt nagyobb hitelesség, kiegyensúlyozottság és a 0-hoz legközelebbi delta érték.

3.2.8. Számítás összegzése

A számítási feladat megmutatta, amennyiben a teljes adatvagyon megsemmisül és csupán a statisztikai értékek ismertek (pl. a hiányzó adatvagyon más számításokhoz történő felhasználásának nyíltan elérhető értékeiből), akkor ez alapján egy becsült tábla szerinti korrelációs mátrixok létrehozásával mennyire is lehet visszaadni a valósnak vélt adatokat. Mindehhez Excel Solver használatával három hibateneggellyel dolgozva megállapítható a korrelációs mátrixok, az oszlopokra vonatkozó statisztikai mutatók, valamint a sorokra

vonatkozó statisztikai mutatók különbsége, eltérései. Ezen hiba adatok százalékos formára alakítása, majd egységes súlyozással számítva összeg és szorzat aggregátorral 20 külön Solver megoldást rögzítettem, ami képezte az OAM-táblát. Ez a 20 megoldás egységesen ugyanazt a célt próbálta megközelíteni más-más súlyozással, aminek eredményeképp rögzítésre került az itt kapott összeg mutató értéke, ez adta a kísérleti táblát. A kísérleti tábla nagyon jó kiindulási alap, de önmagában nem megállapítható, hogy a 20 sor közül melyik is van legközelebb a célhoz. Ha csupán ezekkel az adatokkal történne egy megállapítás, akkor teljesen félrevezetné a döntéshozókat, mivel csak úgy tűnik, hogy jók az adatok és egyértelmű a legjobb sor, de ez valójában nem igaz. Erre a kérdésre adott pontosabb választ a COCO Y0-modell, amely a kísérleteket az elért, rögzített mutatók alapján rangsorolta. Így valós képet kaptam arról, hogy melyik sorból állapítható meg a legbiztosabban a szerződés hosszabbítási-hajlandóság.

eltérés (sorrend)	sorrend	Zöld db	ID	Min (abszolút) (pontszám)	Max (abszolút) (pontszám)	Átlag (abszolút) (pontszám)	Standard (abszolút) (pontszám)	Medián (abszolút) (pontszám)	Módusz (abszolút) (pontszám)	Q1 (abszolút) (pontszám)	Q3 (abszolút) (pontszám)	Adat cs. (pontszám)	Y0 (pontszám)	naiv (sorrend)	sorrend	validálás (1=igen, 0=nem)	becslés (pontszám)	sorrend	eltérés (sorrend)
-5	15	7	1	18	6	13	6	18	10	12	11	4	1000000	11	10	1	999998.8	15	-5
0	1	14	2	18	3	13	3	18	10	12	11	4	1000000	2	1	0	1000001.2	14	0
0	12	8	3	18	1	12	1	18	10	12	11	4	1000000	3	2	0	1000001.1	12	0
0	12	8	4	12	15	12	12	12	10	10	13	17	1000000	11.67	13	0	999998.4	12	0
2	12	7	5	12	10	10	20	18	12	14	18	12	1000000	16	17	1	999999.6	12	0
-1	9	11	6	18	4	12	4	18	10	12	11	4	1000000	6	2	0	1000000.6	9	0
0	2	13	7	6	2	2	2	2	1	5	24	24	1000000	3	2	0	1000000.4	2	0
8	3	11	8	12	12	10	18	18	10	14	12	18	1000000	12	14	0	1000001.4	3	0
1	7	10	9	14	10	12	9	14	10	10	14	18	1000000	12	8	1	1000010.1	7	1
1	18	5	10	12	13	14	17	20	20	17	18	14	1000000	17	10	0	999993.8	18	0
-6	15	7	11	18	16	14	14	15	10	10	15	1	1000000	11	10	1	1000011.2	15	0
0	12	8	12	17	18	18	18	18	10	15	20	18	1000000	18	18	1	999994.1	12	0
-2	8	9	13	14	7	7	8	10	11	11	11	13	1000000	8	7	1	1000042.3	8	0
2	18	5	14	18	18	18	18	18	18	18	17	20	1000000	18	20	1	999997.1	18	0
3	12	15	15	7	5	5	5	8	10	10	15	17	1000000	7	6	1	1000040.8	12	1
-2	18	5	16	9	14	17	4	18	18	20	15	18	1000000	14	16	1	999970.8	18	0
2	12	8	17	20	8	15	15	18	18	18	14	18	1000000	13	14	1	999999.9	12	0
0	9	18	18	18	12	11	11	10	10	10	10	10	1000000	13	15	0	999971.5	9	0
0	9	9	19	11	11	8	15	9	1	1	10	9	1000000	9	9	1	1000012.2	10	1
-2	7	10	20	6	8	8	12	4	4	4	8	11	1000000	6	5	1	1000041.8	7	1

15. ábra COCO-módszer szerinti rangsorolás eredménye, legjobb eredmény megállapítása

Saját forrás: „tablazat” munkalap D114-X134 tartománya / Jelmagyarázat: A COCO-modell futása utána kapott eredmények rangsormegállapítása, ahol a naiv oszlop a nem optimalizált eredmény jelenti. A becslés pedig az értékek optimalizált, legkevesebb ellentmondást mutató alakzata. A színskálán a zöld jelenti a legjobb értéket, a piros a legrosszabb értéket.

3.5 Adatvédelem és GDPR

Ahhoz, hogy adatvédelmi szempontból megfelelően elemezhesünk, először is tisztában kell lennünk néhány alapvető adatvédelmi és jogi fogalommal. Az egyik legfontosabb ilyen fogalom maga az adatvédelem, amely „a személyes adatok jogszerű kezelését, az érintett személyek védelmét szolgáló alapelvek, szabályok, eljárások, adatkezelési eszközök és módszerek összessége” (NAIH, 2025).

Az érintett személy vagy más néven adatalany az a természetes személy, aki bármely meghatározott személyes adat alapján beazonosítható. A személyes adat olyan információ, amely az érintett személyhez kapcsolható, például a neve vagy bármilyen azonosító jele (NAIH, 2025).

Vannak azonban úgynevezett különleges adatok is, amelyek kezelése szigorúbb szabályokhoz kötött, és csak meghatározott környezetben engedélyezett. Ilyen különleges adatnak számítanak többek között az egészségügyi információk, a biometrikus adatok, a genetikai adatok, valamint a káros szenvedélyekre vonatkozó adatok (Horváth, 2018).

A gyűjtött adatokat különböző módokon kezelhetik vagy feldolgozhatják. Az adatkezelés egy igen tág értelemben vett fogalom, amely így fogalmazható meg: „Az adatokon végzett bármely művelet, például az adatok gyűjtése, felvétele, rögzítése, rendszerezése, tárolása, megváltoztatása, felhasználása, lekérdezése, továbbítása, nyilvánosságra hozatala, összehangolása vagy összekapcsolása, zárolása, törlése és megsemmisítése, valamint az adatok további felhasználásának megakadályozása, fénykép-, hang- vagy képfelvétel készítése, valamint a személy azonosítására alkalmas fizikai jellemzők (ujj- vagy tenyérnyomat, DNS-minta, íriszkép stb.) rögzítése.” (NAIH, 2025)

Az előírások nemcsak azokra vonatkoznak, akik tárolják a begyűjtött adatokat, hanem azokra is, akik csupán lekérdezik azokat. Az adatkezelő az a személy vagy szervezet, amely a természetes személyek adatait kezeli, meghatározva az adatkezelés módját, elveit, céljait és eszközeit. Minden munkáltató és vállalkozás, amely ügyfelei adatait nyilvántartja, adatkezelőnek minősül. Az adatkezelő megbízásából az adatfeldolgozó végzi el az adatkezelési műveletekhez kapcsolódó technikai feladatokat; ilyenek azok a cégek, amelyek más vállalkozások számára nyújtanak például IT üzemeltetési, könyvelési vagy bérszámfejtési szolgáltatásokat. Az adatok gyűjtéséhez, feldolgozásához és kezeléséhez elengedhetetlen az érintettek hozzájárulása: „Az érintett akaratának önkéntes és határozott kinyilvánítása, amely megfelelő tájékoztatáson alapul, és amellyel félreérthetetlen beleegyezését adja a rá vonatkozó személyes adatok – teljes körű vagy egyes műveletekre kiterjedő – kezeléséhez. Különleges adatok esetében szükséges az írásos forma.” (NAIH, 2025)

Mint fentebb említettük, a megfelelő tájékoztatás az adatvédelem alapvető része: „az érintettel az adatkezelés megkezdése előtt közölni kell, hogy az adatkezelés a hozzájárulásán alapul-e vagy kötelező, továbbá egyértelműen és részletesen tájékoztatni kell az adatai kezelésével kapcsolatos minden tényről, így különösen az adatkezelés céljáról és jogalapjáról, az adatkezelésre és az adatfeldolgozásra jogosult személyéről, az adatkezelés időtartamáról, illetve arról, hogy kik ismerhetik meg az adatokat. A tájékoztatásnak ki kell terjednie az érintett adatkezeléssel kapcsolatos jogaira és jogorvoslati lehetőségeire is.” (NAIH, 2025)

A felhasználók védelmét különféle jogszabályok és hatóságok biztosítják. Ezek közül a legfontosabb és legátfogóbb a GDPR. Itt azonban a GDPR előtti, illetve vele egyidejűleg hatályos szabályozásokat és hatóságokat mutatjuk be. Ilyen hatóság például a NAIH, vagyis

a Nemzeti Adatvédelmi és Információszabadság Hatóság (Holló, 2018), amely 2012. január 1-jétől működik, és amely az adatvédelmi biztos intézményét váltotta fel (NAIH, 2025).

A NAIH feladata a személyes adatok védelmének biztosítása, az adatvédelmi jogok érvényesülésének ellenőrzése és elősegítése. A GDPR hatályba lépése óta Magyarországon a NAIH felel a GDPR betartásáért, az ellenőrzések lefolytatásáért, valamint a szabályszegések szankcionálásáért (Holló, 2018).

Fontos volt továbbá az az uniós szintű önálló munkacsoport, amelyet az 95/46/EK irányelv 29. cikke alapított (NAIH, 2025). Ez a csoport 2018. május 25-éig, a GDPR hatálybalépéséig felügyelte a magánélet és a személyes adatok védelmét (EDPB, 2019).

Az adatvédelem jogi védelme már régóta jelen van: Magyarországon például 1992-ben hozták meg az 1992. évi LXVI. törvényt, amely a polgárok személyi adatait és lakcímét védi (Jogtár, 2025a). Ez volt az első ilyen adatvédelmi törvény a rendszerváltás után, amely már lehetővé tette a természetes személyek számára, hogy tiltakozzanak a róluk tárolt adatok kiadása ellen (Jogtár, 2025a). Ez a törvény 2012. január 1-jén hatályát veszítette (NAIH, 2025).

Emellett fontos még az 1995. évi CXIX. törvény, amely a kutatás és a közvetlen üzletszerzés célját szolgáló név- és lakcímadatok kezeléséről szól (Jogtár, 2025b). Ez a törvény kiterjed tudományos kutatásokra, közvélemény- és piackutatásokra, valamint a direkt marketingre is, és szabályozza az adatgyűjtést és adatfelhasználást ezekben a célokban (Jogtár, 2025b).

Az Európa Tanács valamennyi tagállama által aláírt 1998. évi VI. törvény az egyének védelmét szolgálja a személyes adatok gépi feldolgozása során, biztosítva, hogy az érintettek jogait tiszteletben tartsák (Jogtár, 2025c).

Végül az Infotv., vagyis a 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról 2012. január 1-je óta hatályos, és helyébe lépett az 1992-es törvénynek (NAIH, 2025). Célja egyrészt a természetes személyek magánszférájának védelme, másrészt a közügyek átláthatóságának elősegítése (NJT, 2025).

3.6 Automatizálási lépések

Az automatizálási lényege az, hogy a korábban bemutatásra került számítási folyamat egy futással eredményt adjon lehetőleg minimális felhasználói beavatkozás mellett.

Adatfolyam és komponensek:

A megoldás három, jól elválasztott egységre épül, melyek az alábbiak:

1. Paraméterezés:

a. tűrések meghatározása:

- i. alsó és
- ii. felső értékhatár kijelölése
- iii. a tűrések értékeinek rögzítése (minden attribútumra érvényesen):
 1. tűrés1: 0,2
 2. tűrés2: 0,4

b. súlyvektorok (pl. 1:1:1; 2:1:1; 1:3:1) meghatározása:

- i. a három hiba-érték relatív fontosságának kijelölése
 1. hiba1
 2. hiba2
 3. hiba3
- ii. súlyvektor definiálása (minden Solver-futáshoz):
 1. Súly_hiba1
 2. Súly_hiba2
 3. Súly_hiba3
- iii. legalább 20 súlyozási beállítás megadása az OAM-hoz
 1. Súlyozási beállítás 1
 20. Súlyozási beállítás 20

c. aggregátor típusok megjelölése(összeg/szorzat):

- i. KPI-komponensek összeg és
- ii. szorzat jellegű aggregálása
- iii. aggregátor kiválasztása a súlyvektor-változathoz

d. Solver-beállítások:

- i. célfüggvény megadása a Solver-ben
- ii. a becsült 6x3-as tábla celláinak megadása

- iii. korlátok megadása:
 - 1. statisztikai célérték sáv
 - 2. korrelációs tűrés
 - 3. attribútum értéktartomány

2. Számítási sablon:

- a. külön cél-tábla és
- b. becsült tábla,
- c. korrelációs mátrixok:
 - i. cél-tábla korrelációs mátrix
 - ii. becsült tábla korrelációs mátrix
 - iii. korrelációs mátrixok közötti eltérés
- d. hiba_{1/2/3}:
 - i. hiba₁:
 - 1. a cél- és
 - 2. becsült korrelációs mátrix elemenkénti eltérése
 - ii. hiba₂:
 - 1. a cél- és
 - 2. becsült tábla oszloponkénti eltérése
 - iii. hiba₃:
 - 1. a cél- és
 - 2. becsült tábla soronkénti eltérése
 - iv. a három hiba értékének előállítása a későbbi normalizáláshoz:
 - 1. hiba₁ érték
 - 2. hiba₂ érték
 - 3. hiba₃ érték
- e. hibák normalizálása:
 - i. értékek 0-1 tartományba alakítása
 - ii. az értékek azonos nagyságrendre hozása
 - iii. hibák súlyozható, egységes skálára hozása
 - iv. a három normalizált hibaérték előállítása a KPI-számításhoz
- f. KPI-panel:
 - i. a normalizált hibák és súlyok szorzatainak összege adja a KPI-értéket
 - ii. leggyengébb KPI azonosítása

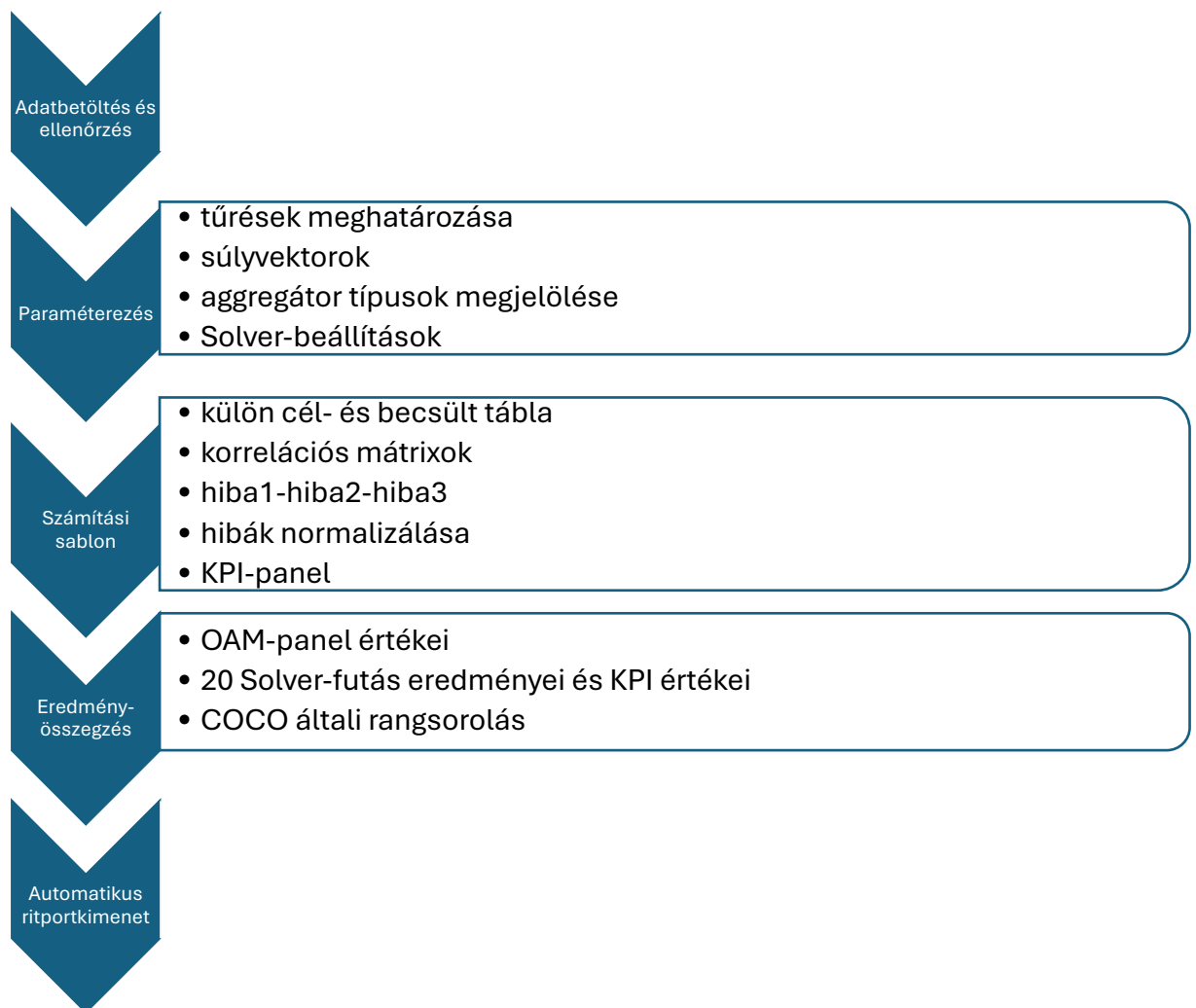
3. Eredmény-összegzés:

- a. OAM panel értékei:
 - i. táblázatos összefoglaló rész
 - ii. egy helyen jeleníti meg a 20 Solver-futás értékeit
- b. a 20 Solver futás eredményei és
 - i. a 20 Solver-futás során előállított becsült értékek,
 - ii. az előállított becsült értékekhez tartozó eltérések rögzítése:
 - 1. statisztikai mutatók eltérései
 - 2. túréssen belüli értékek száma
- c. azok KPI-értékei:
 - i. a 20 Solver-futás során létrejövő összesített mutatók meghatározása:
 - 1. KPI-érték
 - 2. KPI-státusz (színjelölés)
 - ii. az összesített mutató egységes megjelenítése
- d. COCO általi:
 - i. rangsorolás, valamint
 - ii. a futások végső sorrendjének meghatározása,
 - iii. leggyengébb KPI szerint rendezés
- e. automatikus riportkimenet:
 - i. legjobb és
 - ii. leggyengébb értékek kijelölése,
 - iii. kulcsmutatók PDF-be nyomtatása.

3.6.1. Automatizált lépések sorrendje

1. Adatbetöltés és ellenőrzés: a bemeneti tábla a cél és becsült sablonba kerül, a statisztikai mutatók szerinti minimum és maximum közötti értékekkel, tartománynév-frissítéssel.
2. Statisztikai mutatók frissítése: mindkét táblához tartozó leíró statisztikák (minimum, maximum, átlag, szórás, medián, módusz, kvartilisek) automatikusan újraszámolódnak.
3. Inicializálás: a becsült tábla kezdőértékei véletlenszámokkal előállnak és értékként rögzülnek, így azonos paraméterezéssel az eredmény ismételtető.
4. Korrelációk és hibák: a két adattáblából képződik a három korrelációs mátrix, ezekből pedig a hiba1 (táblák közötti eltérés), az oszlop- (hiba2) és sorprofil-eltérések (hiba3).
5. Normalizálás: a hiba1-2-3 azonos nagyságrendre hozása saját maximumukhoz viszonyítva történik, ez képezi a súlyozott, aggregált célfüggvényt,

6. Solver-futás, változó cellák: a becsült tábla esetében a cél: az aggregált hiba minimalizálása, korlátok: a statisztikai célértékek sávjai és a korrelációs tűrések,
7. Kísérleti panel: az előre definiált súlyok és aggregátorok mellett 20 külön futás készül; a fő KPI-k (pl. zöld cellák száma, min/max absz. eltérés, aggregált hiba) az OAM-panelre gyűjtve jelennek meg,
8. COCO Y0 rangsor: a lépcsős függvényekkel számított pontszámok alapján létrejön a rangsor, azonos pontszám esetén a kisebb ellentmondás (delta) élvez elsőbbséget,
9. Riport és archívum: a legjobb megoldás és a kulcsmutatók összegzése táblázatosan és PDF-ben kerül kiírásra, minden futás időbélyeggel archiválódik a visszakövethetőség érdekében.



16. ábra Automatizálási lépések folyamatára – Jelmagyarázat: a folyamatábra megmutatja az automatizálási lépések közötti kapcsolatokat - saját forrás

3.6.2. Integrációs lehetőség

A bemenet CSV/Excel exporttal illeszthető az értékelő által kezelt személyügyi rendszerhez, az automatizált pipeline változtatás nélkül futtatható, a riportok menetrend szerint generálhatók (pl. havi lezáráskor).

Ezen automatizálási lépések sorozatával a szerződeshosszabbítási hajlandóság becslése egységesen, gyorsan és visszakereshetően készül el, a 20 különböző optimalizálási kísérlet COCO Y0 rangsora objektív kiválasztást tesz lehetővé a döntéstámogatáshoz.

3.6.3. Felhasználási terület

A modellt funkciójából adódóan a Magyar Honvédség személyügyi állománya, valamint döntéshozók vennék használatba, civil felhasználás esetében ugyanígy a HR és a vezetőség.

A megoldás használatának gyakoriságát tekintve alkalmazható az alábbiakhoz:

havi zárás

negyedéves tervezés

éves tervezés

eseti jellegű futás, például egy nagyobb változás utáni megfigyelés.

Az automatizáció lényege ezeken a területeken a reprodukálhatóság, az időmegtakarítás, valamint a nyomonkövetés.

Megtérülését tekintve, ha az adott alakulatnál 500 fős szerződéses katona lát el szolgálatot és az éves fluktuáció 15% és az elemzés alapján célzott beavatkozással akár ez 1%-kal csökken, akkor az 5 katonát nem kell pótolni.

3.7 Tesztelés

A tesztelés célja, hogy igazolja az Excel-alapú számítási módszert (leíró statisztikák, korrelációk, hiba1-hiba2-hiba3, normalizálás, Solver futások), hogy minden eleme elvárás szerint működik és nincs eltérés vagy hibaüzenet.

Tesztterületek és elfogadási kritériumok:

1. Szerkezeti integritás: a cél és becsült táblák tartományai, a hivatkozások és a névkonvenciók konzisztensen szerepelnek, nincs törött hivatkozás, nincs hibajelzés, például #ÉRTÉK
2. Leíró statisztikák: minimum, maximum, átlag, szórás, medián, módusz, kvartilisek képletei helyes értéket adnak, manuális ellenőrzéssel,

3. Korrelációs mátrixok: szimmetria és főátló helyes, az 1.00 érték teljesül,
4. Hiba-komponensek: hiba1 (két tábla közötti eltérés), hiba2 (oszlopprofil-eltérés), hiba3 (sorprofil-eltérés) irányhelyesen változnak, bármely részeltérés növelése rontja az aggregált célt,
5. Normalizálás: a hiba-értékek saját maximumukhoz viszonyítva 0–1 közé esnek; monotonitás teljesül,
6. Solver futások: összeg/szorzat aggregátorral és többféle súlyozással (pl. 1:1:1; 2:1:1; 1:3:1) minden futás konvergált, az optimalizált megoldásokban az aggregált hiba csökkent, a „zöld db” ($\leq 0,2$ eltérés) nőtt, a statisztikai értékek közötti eltérés mérséklődött,
7. OAM-panel konzisztencia: a 20 futás mutatói rögzítve, a számított KPI-k belső összhangja figyelhető meg,
8. COCO Y0 rangsor: a lépcsős függvények monotonitása és a delta-minimalizálás teljesül; a rangsorok manuális sorrendezéssel is bizonyíthatók,
9. Vizualizációs ellenőrzések: A feltételes formázás a tűrési küszöbök szerint működik (zöld $\leq 0,2$; sárga 0,2–0,4; piros $> 0,4$), a vizuális jelzések megfelelnek a küszöbök szerinti értékeknek,

Eredmény: A tesztek alapján a modell teljes számítási lánc hibamentesen működik: a képletek helyesek, az optimalizálás hatására javulás figyelhető meg, az OAM- és COCO-eredmények egymással és a manuális ellenőrzésekkel összhangban vannak. A modell a szerződeshosszabbítási hajlandóság becslésének megállapítására stabilan és megbízhatóan használható.

3.8 IT-biztonsági kérdések

Ebben a fejezetben bemutatom az IT-biztonsági kockázatokat és ezekre elérhető védelmi megoldásokat, kitérve a technikai kontrollra, titkosításra, eszköz-és platformbiztonságra, fájl- és verziókezelésre, valamint a hálózatra és adatmozgásra.

3.8.1. Fenyégetésmodell

Kulcskockázatok: jogosulatlan hozzáférés, adatszivárgás, integritássérülés, hordozható eszközök elvesztése, valamint anonim vagy szimulált adatokból pontos személy meghatározásának a kockázata.

3.8.2. Technikai kontroll

Hozzáféréskezelés, melynek célja a cél értékekhez és becsült táblákhoz csak elemzői jogosultság megadása, a Solver-paraméterek módosítása „négy szem elve” szerint. A COCO kimenetek megtekintéséhez vezetői és elemzői szerepkör szükséges.

3.8.3. Titkosítás

- Tárolt adatok: lemeztitkosítás, Windows BitLocker,
- Adatátvitel: VPN használata,
- Excel szint: munkafüzet-védelem erős jelszó alkalmazásával.

3.8.4. Eszköz- és platformbiztonság

Vállalati MDM (mobil-/eszközmenedzsment, ahol a távoli formázás engedélyezve van, Windows eszközök használata ajánlatos a Solver-futások miatt. Erős jelszó kikényszerítése a felhasználótól és a feladat ellátásához nem szükséges alkalmazások letiltása.

3.8.5. Fájl- és verziókezelés

A futások külön, időbélyeggel elnevezett fájlokban (v1...vN) kerülnek mentésre, a kiinduló becsült tábla értékei rögzülnek futás előtt, így biztosított a visszaellenőrzés és a reprodukálhatóság. Mentési stratégia: 3-2-1, 3 példány, 2 eltérő médium, 1 offsite.

3.8.6. Hálózat és adatmozgás

Az elemzői és riportszegmens között hálózati szétválasztás van, az adatok pedig csak egy manuálisan jóváhagyott exportkapun keresztül mehetnek át. Fájlmegosztás csak vállalati tárhelyen lehetséges.

3.8.7. Összegzés

A fenti kontrollok a számítási módszer mellé olyan védelmi rétegeket adnak, amelyek egyszerre kezelik a bizalmasság, a sértetlenség és a rendelkezésre állás követelményeit. Így a modell biztonságosan használható és alkalmas az esetleges honvédségi környezetbe történő beillesztésre.

3.9. Mesterséges intelligencia aspektusok

A dolgozatban a mesterséges intelligencia alapvetően nem szöveg LLM-formában van jelen, hanem numerikus, optimalizáció jellegű modellként. A COCO-modell Solver eredmények

felhasználásával többattribútumos döntéstámogató szerepet tölt be, vagyis eszerint a COCO-modellt numerikus MI-eszköznek lehet tekinteni. Nem nyelvi szövegeket, hanem statisztikai mutatókból alkotott adatvagyonot felhasználva rangsorol az elérendő célhoz legközelebbi érték szerint.

A módszer egyik előnye, hogy a lépései determinisztikusak és teljesen visszakövethetők, ezért biztonságosabb egy feketedoboz jellegű MI-modellnél.

4. fejezet: Vita

A szakdolgozatban kifejtett modell célja az volt, hogy egy teljes adatvagyon megsemmisülés miatt a valós adatok elérhetetlensége estén, miként is lehet statisztikai mutatókkal számítva valósághű adatvagyon létrehozni, ami szerint kimutatható a szerződés hosszabbítási hajlandóság. A Solver-futások értékeit tartalmazó 20 soros kísérleti panel, valamint a COCO Y0 általi rangsorok alapján nem feltétlen az a legjobb megoldás, ahol a legtöbb tűréshatáron belüli érték jelenik meg. A COCO-modell lényege pontosan abban rejlik, hogy minden értéket számításba véve, tehát az összképet nézve állapítja meg a legjobb futást. Ez döntéstámogatás szempontjából előnyös, mivel így megismerhető, hogy mely tényezők milyen mértékben befolyásolják vagy nem befolyásolják a végeredményt.

Mint mindennek, természetesen az általam bemutatott módszernek is vannak korlátai. A számítási feladat egy 6x3-mas táblán történt, egy nagyobb (pl. 60x30) táblákon a Solver skálázhatósága egyfajta technika korlát lehet. Továbbá a számítások közben mivel több eszközt is használtam, észrevettem, hogy még ez a kisebb adatokat tartalmazó Excel fájl néhol lassabban nyílik meg, illetve reagál a parancsokra. Egy nagyobb táblával való végrehajtás során ennek fokozódása jelenthet még korlátot.

Érdekes a modellt más alternatívákkal is összehasonlítani. A logisztikus regresszió rugalmasabb előrejelzést adhat, több paraméterrel és automatikus súlytanulással. A jelen megközelítés kiemelt előnye az ellenőrizhetőség: a cél és becsült tábla szétválasztása, a három hibatengető és a COCO Y0 rangsor mind visszakereshető, valamint egyértelműen magyarázható döntést eredményez, ami személyügyi kontextusban kiemelt szempont.

Előnyös megoldás egy olyan számítási feladat során, ahol attribútumok és rangsorok alapján történik az összehasonlítás. Amennyiben előfordul bizonyos adathiány, ezt a modell jól tolerálja. Magasan objektív módszer, vagyis a szubjektivitás nem torzítja el az eredményeket. Hátrányként jelenik meg, hogy a Solver korlátai miatt csak bizonyos méretű mátrixokkal lehet dolgozni.

A rendszer alapvetően egy jelzőrendszerként működik, tehát korai figyelmeztetést adhat a várható leszerelési hullámokra, mindemellett megmutatja, hogy mely tényezők javítására, fejlesztésére lehet szükség.

A szakdolgozat jogi részében tárgyalt szükségesség és arányosság elve itt is irányadó, tehát csak a legitim célhoz feltétlenül szükséges adatot lehet felhasználni. A COCO-modell átláthatósága segíti a GDPR szerinti elszámoltathatóságot.

Érdemes skálázási próbákat végezni nagyobb adatvagyonnal és több attribútummal. esetleg alternatív Solver megoldásokkal, valamint értékelést készíteni a toborzási és megtartási intézkedések hatásairól.

Összességében a bemutatott modell erőssége a transzparens, auditálható rangsorolás és a „leggyengébb láncszem” szemlélet, amely jelentős segítséget nyújt a honvédségi döntéstámogatásban.

5. fejezet: Következtetések

A szakdolgozat témájához kapcsolódó szakirodalmak alapján, valamint az általam számított módszer eredményei kapcsán az alábbi következtetéseket lehet levonni.

- a használhatatlanságot eredményező mértékű adatok megsemmisülése során a valóság-hű adatokat visszaadó számítással nem okoz jelentős hátrányt az ilyen jellegű adatvesztés,
- akár más vállalati környezetben történő számítások esetében is előfordulhat, hogy valamilyen okból megtagadják a valós adatok kiadását, ezért csak a statisztikai mutatók adják az irányt, amivel létrehozható a kívánt adatvagyon,
- a szerződés-hosszabbítási hajlandóság megállapításához használt COCO-módszer alapján elmondható, hogy minimális szubjektivitással mérhető a katonák rendszerben maradási „kedve” vagy ennek ellenkezője,

Mindent összevetve a COCO-módszerrel a korábban említett céladatok hiányában is megadható valósághű érték, így a fentebb említett hajlandóság kiszámítható a Magyar Honvédség szerződéses állományába tartozó katonák személyügyi adatainak ismerete nélkül, felhasználva a hatékony toborzás, illetőleg megtartás érdekében. Ez természetesen a Magyar Honvédségen kívül az adott esetleg vállalatra szabva is működhet, akár a munkavállalók által kitöltött elégedettségi kérdőívek kitöltését beszámítva.

6. fejezet: Összefoglalás

Dolgozatomban megvizsgáltam, hogy a tervek szintjén Big Data útján COCO-módszerrel hogyan is használható fel egy – jelen esetben szimulált, kisméretű – adatvagyon a szerződeshosszabbítási hajlandóság megállapítására vagy annak megközelítésére. Az adatvesztés okozta adatok hiányában, a reális helyzet teremtése érdekében vált fő kérdéssé és egyben megoldandó problémává, hogy milyen módszerrel lehet ezen adatok nélkül valósághoz közeli értékeket szimulálni. Kiindulópontként bemutattam a Big Data fogalmi ismertetőit, főbb jellemzőit és a társadalmi, illetve adatvédelmi vonatkozásait. Kiemeltem, hogy a releváns adatvagyon felépítése nem csupán szimpla adatgyűjtést jelent, hanem ezenfelül adattisztítást, validálást és strukturálást, amelyekre az optimalizálás építhető. A gyakorlati részben az átláthatóság kedvéért a katonai személyügyi adatoknál kisebb, jelképes méretű (6x3-as) táblát használtam, de így is jól látható a modell működése. Létrehoztam egy cél-, valamint egy becslés-táblát statisztikai mutatóikkal kiegészítve. A modell egyik magját a korrelációs mátrixok képezik, amivel nem csak azt vizsgáltam, hogy egyes értékek milyen erősen és milyen irányba mozognak együtt, hanem azt is, hogy az oszlopok, illetve sorok szintjén milyen statisztikai mutatók írják le a céladatok szerkezetét. Mindezek adatok alapján egyenkénti Solver futással létrehozott 20 soros kísérleti tábla alapján COCO-módszerrel megállapítottam a rangsort, ami megmutatta, hogy melyik aggregátorral és súlyozással számított értékek kerültek legközelebb a célértékhez. Ennek beépítése a szerződeshosszabbítási hajlandóság valós vizsgálatába a döntéshozó egyértelmű képet kaphat arról, hogy ez esetben melyik katonánál valószínű a szerződés meghosszabbítása és melyiknél lehet számítani leszerelésre a határozott idő leteltével.

7. fejezet: Jövőkép

Az imént következtetésben leírtakhoz visszatérve, a COCO-módszer lehetőséget nyújt az olyan helyzetek megoldására, ahol a kialakult adatvesztés/adathiány miatt jelentős hátrány keletkezett. Ezt az „űrt” hivatott betölteni a valósághű adatok megállapításával. Továbbá ha az adott cég viszont a számítási feladatokat kiszervezi esetleg egy másik cégnek, de adatvédelemre hivatkozva a munkavállalók adatai nélkül, kizárólag a statisztikai mutatókkal, így a szolgáltatás megrendelést így is képes legyen teljesíteni. Ha a módszer további fejlesztésére kerülne sor, akkor esetleg meg lehetne állapítani azt – más számítási értékekkel -, hogy egy adott csoportba tartozó munkavállaló átlagosan hány évig marad a vizsgálat ideje alatt betöltött pozíciójában. A csoportba tartozás természetesen nem diszkriminatív üzenetű, egyszerűen több adat alapján kialakult mintázatok szerint lehetne csoportokat létrehozni és ezekhez társítani a munkavállalókat.

8. fejezet: Mellékletek

E fejezet az irodalom- és forrásjegyzéket, az ábrajegyzéket, valamint a rövidítési jegyzéket tartalmazza.

8.1 Irodalom- és forrásjegyzék

Ambrus, Péter alezredes (2020): GENERÁCIÓS KIHÍVÁSOK KEZELÉSE A MAGYAR HONVÉDSÉGBEN, Hadtudományi Szemle, HSz 2020/4., 108–120.

https://real.mtak.hu/125803/1/HSZ_2020_148_4_Ambrus_Peter.pdf (Letöltve: 2025.11.01)

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016): HR and analytics: Why HR is set to fail the big data challenge. Human Resource Management Journal, 26(1), 1–11. <https://doi.org/10.1111/1748-8583.12090> (Letöltve: 2025.11.01)

Belényesi, Pál (2016): DIGITÁLIS PLATFORMOK ÉS A BIG DATA

<https://kti.krtk.hu/file/download/vesz2016/bigdata.pdf> (Letöltve: 2025.11.01)

Computerworld (2017): Amit a big data elemzésekről tudni érdemes.

<https://computerworld.hu/biztonsag/amit-a-big-data-elemzesekrol-tudni-erdemes-226889.html>
(Letöltve: 2025.10.31)

De Mauro, A. – Greco M. – Grimaldi, M. (2016), “A Formal Definition of Big Data Based on its Essential Features”, Library Review, Vol. 65., No. 3, pp. 122-135,

<http://www.emeraldinsight.com/doi/pdfplus/10.1108/LR-06-2015-0061>. (Letöltve: 2025.10.31)

EDPB (2019): A 29. cikk alapján létrehozott munkacsoport https://edpb.europa.eu/our-work-tools/article-29-working-party_hu (Letöltve: 2025.11.01)

Teo, E. G.S. (2015): Emergence, Growth, and Sustainability of Bitcoin: The Network Economics Perspective, 191-200 oldal, 2015 Elsevier, ISBN 978-0-12-802117-0, David

Evans, D.S; Schmalensee, R. (2010): Failure to launch: critical mass in platform

businesses. Rev. Netw. Econ. 9 (4), 1–28. DOI: 10.2202/1446-9022.1256. (Letöltve: 2025.10.29)

Dr. Holló, Dóra (2018): A Nagy GDPR Kérdezz-felelek <https://7blog.hu/gdpr/> (Letöltve:2025.10.26)

Dr. Horváth, Katalin (2018): GDPR szótár: 15 fontos alapfogalom, amit jó ha ismersz https://piacesprofit.hu/kkv_cegblog/gdpr-szotar-15-fontos-alapfogalom-amit-jo-ha-ismersz/ (Letöltve: 2025.10.26)

IBM (2017): 10 Key Marketing Trends for 2017. <https://www.studocu.com/en-au/document/monash-university/taxation-law/10-key-marketing-trends-for-2017-and-ideas-for-exceeding-customer-expectations/5963924> (Letöltve: 2025.10.26)

Jogtár, (2025a): 1992. évi LXVI. törvény a polgárok személyi adatainak és lakcímének nyilvántartásáról <https://net.jogtar.hu/jogszabaly?docid=99200066.TV> (Letöltve: 2025.10.26)

Jogtár, (2025b): 1995. évi CXIX. törvény a kutatás és a közvetlen üzletszerzés célját szolgáló név- és lakcímadatok kezeléséről <https://net.jogtar.hu/jogszabaly?docid=99500119.TV> (Letöltve: 2025.10.27)

Jogtár (2025c) 1998. évi VI. törvény az egyének védelméről a személyes adatok gépi feldolgozása során, Strasbourgban, 1981. január 28. napján kelt Egyezmény kihirdetéséről <https://net.jogtar.hu/jogszabaly?docid=99800006.TV> (Letöltve: 2025.10.27)

Marton, Ádám (2023): A big data elemzés alapvető összefüggései <https://www.ludovika.hu/blogok/ludeconblog/2023/08/25/a-big-data-elemzes-alapveto-osszefuggesei/> (Letöltve:2025.10.31)

Mayer –Schönberger V., Cukier K, ford. Dankó Zsolt (2014). Big data : Forradalmi módszer, amely megváltoztatja munkánkat, gondolkodásunkat és egész életünket, HVG könyvek. HVG Zrt., Budapest.

NAIH (2025) Adatvédelmi szótár - NEMZETI ADATVÉDELMI ÉS INFORMÁCIÓSZABADSÁG HATÓSÁG <https://naih.hu/adatvedelmi-szotar> (Letöltve: 2025.10.26).

Dr. Németh, András – Virágh, Krisztián (2023): Mesterséges intelligencia és haderő – További katonai alkalmazási lehetőségek, Haditechnika LVII. évf. – 2023/2 https://real.mtak.hu/175736/1/HT_2023-2_cikk_01.pdf (Letöltve: 2025.11.01)

NJT (2025) 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról http://njt.hu/cgi_bin/njt_doc.cgi?docid=139257.362143 (Letöltve: 2025.10.26)

Dr. Pitlik, László: Hasonlóságelemzés COCO használatával (2013), Oktatási segédanyag, https://miau.my-x.hu/miau/189/coco_demo.pdf (Letöltve: 2025.11.01)

Resperger, István ezredes – Resperger, Viktória (2025): A TOBORZÁS AKTUÁLIS KÉRDÉSEI NEMZETKÖZI ÖSSZEHASZNÁLÁS ALAPJÁN (1.), HSz 2025/5., 82–96. DOI: 10.35926/HSZ.2025.5.7 (Letöltve: 2025.10.28)

Szommer Károly, Balogh Zoltán, Racskó Péter, (2014): Az on-Line Világban Hagyott Virtuális Lábnyomokban Rejlő Információk És Azok Veszélyei. Vezetéstudomány / Budapest Management Review 45, 97–104.

Takács, Gergely (2018): Big Data (adatvezérelt) elemzési módszerek alkalmazása a nemzetbiztonsági szférában I. rész, TERROR& ELHÁRÍTÁS 2018/1. sz., https://epa.oszk.hu/02900/02932/00016/pdf/EPA02932_terror_elharitas_2018_1_005-044.pdf (Letöltve: 2025.10.28)

Toonders, J. (2014): Data Is The New Oil Of The Digital Economy, <https://www.linkedin.com/pulse/20140716110251-7386607-data-is-the-new-oil-of-the-digital-economy> (Letöltve: 2025.10.25)

Zódi, Zsolt (2017): Privacy és a Big Data. Fundamentum. 1-2. sz <http://fundamentum.hu/sites/default/files/fundamentum-17-1-2-02.pdf> (Letöltve: 2025.10.31)

8.2 Ábrajegyzék

1. ábra Kritikus tömeg és az egyensúlyi helyzet a direkt és az indirekt externáliák függvényében (Evans et al., 2010).....	15
2. ábra adatok1 – cél táblázat Saját forrás: „tablázat” munkalap A1-M18 tartománya / Jelmagyarázat: táblában látható az elérendő cél, valamint a későbbi becsléshez szükséges statisztikai mutatók.....	28
3. ábra adatok2 – becsült érték Saját forrás: „tablázat” munkalap A28-M45 tartománya / Jelmagyarázat: a tábla 6x3 részében az 1. ábra szerinti statisztikai mutatók korlátai közé szorított véletlenszámok jelennek meg a későbbi feldolgozáshoz, valamint az oszlopra és sorokra vonatkozó statisztikai adatok	31
4. ábra adatok1 – korrelációs mátrix Saját forrás: „tablázat” munkalap A20-D25 tartománya / Jelmagyarázat: megmutatja, hogy egyes változók mekkora mértékben és milyen irányba mozognak egymáshoz viszonyítva.....	31
5. ábra adatok2 – korrelációs mátrix Saját forrás: „tablázat” munkalap A20-D25 tartománya / Jelmagyarázat: megmutatja, hogy egyes változók mekkora mértékben és milyen irányba mozognak egymáshoz viszonyítva.....	32
6. ábra adatok1 és adatok2 korrelációs mátrix eltérés Saját forrás: „tablázat” munkalap A54-D67 tartománya / Jelmagyarázat: megmutatja a cél tábla mátrix és a becsült tábla mátrix közötti eltérést	32
7. ábra adatok1 és adatok2 korrelációs mátrix eltérés oszlopok szerint Saját forrás: „tablázat” munkalap F54-I67 tartománya / Jelmagyarázat: megmutatja a cél tábla mátrix és a becsült tábla statisztikai mutatóinak oszlop szerinti eltérései	33
8. ábra adatok1 és adatok2 korrelációs mátrix eltérés sorok szerint Saját forrás: „tablázat” munkalap K54-Q67 tartománya / Jelmagyarázat: megmutatja a cél tábla mátrix és a becsült tábla statisztikai mutatóinak sor szerinti eltéréseit.....	34
9. ábra hiba értékek maximum meghatározása és azonos súlyozás beállítása Saját forrás: „tablázat” munkalap A73-C76 tartománya / Jelmagyarázat: A korábban említett számítással megállapított értékek közötti eltérést adó hiba1, hiba2 és hiba3 maximumának megjelenítése és azonos nagyságrendre hozása.....	34
10. ábra a cél tábla és a becsült tábla értékeinek különbsége Saját forrás: „tablázat” munkalap A80-h87 tartománya / Jelmagyarázat: A két tábla közti küszöbök szerinti különbség megjelenítése színnel.....	35

11. ábra Solver futtatással elért értékek szerinti statisztikai mutatók eltérései Saját forrás: „tablazat1” munkalap A90-P91 tartománya / Jelmagyarázat: Solver futtatást követően kapott értékek összehasonlítása a cél és becsült táblák vonatkozásában bizonyos aggregátorral és bizonyos súlyozás szerint	36
12. ábra 20 soros kísérlet tábla, Solver futások Saját forrás: „tablazat1” munkalap A90-Q110 tartománya / Jelmagyarázat: Minden Solver által kapott értékek megjelenítése egy táblázatban és a soronkénti leggyengébb KPI-k meghatározása	37
13. ábra COCO-módszer szerinti rangsorolás Saját forrás: „Munka1” munkalap A7-K27 tartománya / Jelmagyarázat: O1-O20 sorok a korábban számított kísérleti tábla értékeit tartalmazza, valamint Y-érték, amelynek segítségével definiálhatók a rangsorok	38
14. ábra COCO-módszer szerinti rangsorolás Saját forrás: „Munka1” munkalap A7-K27 tartománya / Jelmagyarázat: O1-O20 sorok a korábban számított kísérleti tábla értékeit tartalmazza, valamint Y-érték, amelynek segítségével definiálhatók a rangsorok	39
15. ábra COCO-módszer szerinti rangsorolás eredménye, legjobb eredmény megállapítása Saját forrás: „tablazat” munkalap D114-X134 tartománya / Jelmagyarázat: A COCO-modell futása után kapott eredmények rangsormegállapítása, ahol a naiv oszlop a nem optimalizált eredmény jelenti. A becslés pedig az értékek optimalizált, legkevesebb ellentmondást mutató alakzata. A színskálán a zöld jelenti a legjobb értéket, a piros a legrosszabb értéket.	40
16. ábra Automatizálási lépések folyamatára – Jelmagyarázat: a folyamatára megmutatja az automatizálási lépések közötti kapcsolatokat - saját forrás.....	46

8.3 Rövidítések jegyzék

COCO - component based object comparison for objectivity

OAM - Objective Aggregation Model

KPI - Key Performance Indicator

GDPR - General Data Protection Regulation

NAIH - Nemzeti Adatvédelmi és Információszabadság Hatóság

IT - Information Technology

MH – Magyar Honvédség

MDM - mobil-/eszközmenedzsment – Mobile Device Management

CIA – Confidentiality, Integrity, Access

HR – Human Resources

HRIS - Human Resources Information System

CIMIC- Civil-Military Cooperation – civil-katonai együttműködés

NLP - neuro-lingvisztikus programozás

PSYOPS - Psychological Operations – lélektani hadviselés

NATO - North Atlantic Treaty Organisation

8.4. Releváns LLM-konverzációk teljes szövege

Ebben a fejezetben a szakdolgozathoz felhasznált MI-vel történő beszélgetések szövegét jelenítem meg, hiszen az LLM használata kötelező volt az elkészítés során.

8.4.1. Számítással kapcsolatos kérdés és válasz

Kérdés: A hibátlan % amúgy mind a 20 sornál 0, így is 1 legyen az irány értéke? Vagy mindegy?

Válasz: Röviden: igen, maradjon +1 a „Hibátlan %” iránya - elvben minél nagyobb, annál jobb. De mivel mind a 20 sorban 0, ez a KPI most nem különböztet meg semmit. Mit tehetsz:

A) Hagyd bent, de semlegesítsd a hatását (ajánlott):

a normalizálásnál kezeld a „max = min” (itt $O=0$) esetet semleges értékkel, pl. 0,5

B) Ideiglenesen vedd ki a Q-ból:

ha súlyokat használsz, ennek a KPI-nek adj 0 súlyt;

ha nem, egyszerűen ne vedd bele a Q-min számításba addig, amíg konstans.

8.4.2. Automatizálással kapcsolatos kérdés és válasz

Kérdés: A csatolt számítások szerint milyen automatizálási lépések relevánsak?

Válasz:

- Paraméterezés: *tűrések (0,2/0,4), súlyvektorok (pl. 1:1:1; 2:1:1), aggregátor (összeg/szorzat), futások száma (20).*
- Adatbetöltés és skálakontroll (0-1 tartomány),
Cél/Becsült táblák szétválasztása.
- Célpofil újraszámolása (min, max, átlag, szórás, medián, módusz, kvartilisek).
- Becsült mátrix inicializálása és értékként

rögzítése (reprodukálhatóság).

- Korrelációk képzése és hiba1/hiba2/hiba3

kiszámítása.

- Normalizálás (saját maximumhoz viszonyítva) +

súlyozott, aggregált célfüggvény.

- Solver futtatások a Becsült mátrixon, előre definiált súlyvektorokkal/aggregátorokkal.

- KPI-k gyűjtése (pl. zöld cellák száma, min/max absz. eltérés, aggregált hiba) → OAM-panel.

- COCO YO rangsorolás és a legjobb megoldás kiválasztása.

- Riportkészítés és verzió-archiválás időbélyeggel.

8.4.3. COCO-modell más HR témával kapcsolatos kérdés és válasz

Kérdés: Milyen HR vonatkozású témákban alkalmazható még a COCO-modell?

Válasz: HR-területek, ahol a COCO-modell alkalmazható

- Toborzási prioritások kijelölése (mely profilokat érdemes előre rangsorolni).
- Kiválasztási modell-alkotás (jelöltek pontozása több szempont alapján).
- Kompetencia-értékelés súlyozása (mely tényezők számítanak jobban).
- Munkaköri megfelelés rangsorolása (ki mely pozícióra illik leginkább).
- Teljesítményértékelési indexek képzése (súlyozott értékek összevetése).
- Előléptetési döntéstámogatás (alkalmassági tényezők kombinálása).
- Képzési prioritások meghatározása (ki milyen fejlesztési igényt mutat).
- Kulcsemberek azonosítása (magas értékű, helyettesíthetetlen szereplők).
- Utódlástervezés támogatása (kik a potenciális utánpótlások).
- Jutalmazási/bonusz-elosztási modellek (kinek milyen súlyozott érték alapján jár).
- Hiányposztok kockázati rangsora (hol legkritikusabb a munkaerőhiány).
- Munkahelyi kockázati faktorok összegzése (stressz, túlterhelés mintázatai).
- Fluktuációs kockázat összevetése részlegek között.
- Szervezet egészségi index számítása (morál, stabilitás, terhelés mutatói).
- HR-beavatkozások hatáselemzése (mi változik egy intézkedés után).
- Költség-hatás összevetések (mely HR-lépés ad jobb eredményt).
- Létszámtervezési scenáriók rangsorolása (melyik opció a legkedvezőbb).
- Onboarding eredményességének mérés Vi mennyire illeszkedik be gyorsan).

- *HR-kockázati pontszámok aggregálása (részlegekre, beosztásokra).*

8.4.4. Feketedoboz jellegű MI-modellel kapcsolatos kérdés és válasz

Kérdés: Micsoda pontosan a feketedoboz jellegű MI-modellel?

Válasz: *A feketedoboz jellegű MI-modellel (black-box model) azt jelenti, hogy: a modell belső működése a felhasználó számára nem átlátható, nem lehet pontosan megmondani miért hozott egy adott döntést, és a kimenet oka nem vezethető vissza közérthető, determinisztikus lépésekre. Ez az MI egyik alapvető problémája.*

Mitől „feketedoboz” egy MI? — Három ok

1) A modellek túl bonyolultak

Több millió vagy milliárd paraméter van (pl. LLM, deep learning), amit ember nem tud követni.

2) A döntések logikája nem explicit

A modell nem ad szabályt vagy magyarázatot:

→ nem lehet pontosan tudni, hogyan jutott egy adott eredményre.

3) A működése nem rekonstruálható lépésről lépésre

Ellentétben egy klasszikus eljárással (szabályrendszer, lineáris modell, táblázat), a feketedoboz-modellnél nincs:

konkrét döntési fa

konkrét képlet

egyértelműen követhető számítási lépés

determinisztikus működés

Milyen MI-modellek tipikusan feketedobozok?

Deep learning (neurális hálók)

LLM-ek (mint a ChatGPT)

Random forest (egy részben)

Gradient boosting

Reinforcement learning komplex környezetben

Ezeknél a kimenetet még a fejlesztő sem tudja lépésről lépésre rekonstruálni.