

Kodolányi János University

Újmédia és Kreatívipari Kar



IMDB Index szimulátor – avagy egy robot-filmproducer fejlesztése

IMDB-simulator or how to develop a robot-producer

Kolonics Zsolt

ÜZEMMÉRNÖK-INFORMATIKUS

Dr. Pitlik László

Egyetemi docens, Tanszékvezető

2024.11.22.

Tartalomjegyzék

<u>1. BEVEZETÉS</u>	4
<u>1.1 MOTIVÁCIÓ</u>	4
<u>1.2 CÉLOK</u>	4
<u>1.3 CÉLCSOPORTOK</u>	4
<u>1.4 HASZNOSSÁG</u>	4
<u>1.5 A DOLGOZAT SZERKEZETÉRŐL</u>	5
<u>1.6 A MEGOLDÁSRÓL</u>	5
<u>2. SZAKIRODALMI HÁTTÉR</u>	5
<u>2.1 ÖSSZEGZÉS</u>	9
<u>2.2 KORÁBBI MŰVÉSZETI ÉS SPORT ELEMZÉSEK</u>	11
<u>2.3 HASONLÓSÁGELEMZÉS</u>	12
<u>2.4 BENCHMARKING</u>	12
<u>3. SAJÁT FEJLESZTÉSEK</u>	13
<u>3.1 ADATGYŰJTÉS</u>	13
<u>3.1.1 Az adatvagyon forrása</u>	13
<u>3.1.2 A kiválasztás folyamata</u>	13
<u>4. A MODELL NAIV (NEM OPTIMALIZÁLT) ALKALMAZÁSA</u>	15
<u>4.1 OAM ÉRTÉKEK SORBA RENDEZÉSE</u>	15
<u>4.1.1 Az online szakértői rendszer bevonása</u>	16
<u>4.1.2 A szakértői ábrázolás</u>	16
<u>4.1.3 Kimenő jelek értelmezése</u>	16
<u>4.2 A MODELL GYAKORLATI</u>	17
<u>4.2.1 A megbecsülni kívánt objektum kiválasztása</u>	17
<u>4.2.3 A becslés menete</u>	17
<u>4.3. KEZDETI EREDMÉNYEK BEMUTATÁSA</u>	19
<u>4.3.1. Eredmények és a problémák felvázolása</u>	19
<u>4.3.3 Út az optimalizálás felé (a Renitens objektumok kiszűrésével)</u>	20
<u>4.3.4 Út az optimalizálás felé (származtatott adatok bevonásával)</u>	21
<u>4.3.5 Út az optimalizálás felé (Származtatott adatok bevonásával és a renitens objektumok ignorálásával)</u>	22
<u>4.4 BECSLÉS EREDMÉNYESSÉGE</u>	24
<u>5. VITA</u>	25
<u>ÁLLÍTÁS</u>	25
<u>ÁLLÍTÁS 2</u>	26
<u>ÁLLÍTÁS 3</u>	27
<u>ÁLLÍTÁS 4</u>	28
<u>6.KÖVETKEZTETÉSEK ÉS JÖVŐKÉP</u>	30
<u>7.ÖSSZEFOGLALÁS</u>	30
<u>8. MELLÉKLETEK</u>	31
<u>9. IRODALOMJEGYZÉK</u>	31

<u>10. RÖVIDÍTÉSJEGYZÉK</u>	31
<u>11. ÁBRA JEGYZÉK</u>	32
<u>12. DEFINÍCIÓS JEGYZÉK</u>	33

1. Bevezetés Röviden bemutatásra kerül a feladat motivációja, célja és célcsoportja, valamint felépítése az olvasó számára.

1.1 Motiváció Az algoritmus értelmezés atyjaként emlegetett Donald Ervin Knuth szavaival élve „Tudás csak az, ami forráskódba átírható, minden más emberi aktivitás, művészet.”

(Vö. https://miau.my-x.hu/miau2009/index_tki.php3?_filterText0=*knuth)

De hol a határ? Tehetjük fel a kérdést. **Avagy átírható-e forráskódba egy filmélmény**, emberi érzelmek sokasága vagy ez már/még a művészet azon szintje, amit soha nem fogunk tudni megérteni egy robottal? Más szavakkal, lehetséges-e egy producer hatásokat kiváltani akaró gondolatrendszerét szimulátorra alakítani, ha már pl. repülőgép-szimulátor létezik? Képessé tehető-e ezen szimulátor arra, hogy a rendelkezésre álló mérhető/megfigyelhető adatok kölcsönhatásait objektíven ellenőrizhetően nagy helyességgel kezelje? Stb. Amennyiben igen, úgy a robotizáció égető szükségességét a filmipar sem kerülheti el, ennek kiderítése pedig mindannyiunk feladata, hiszen „a változás a természet alaptörvénye” (*Charles Darwin, The Origin of Species, 1859. november 24*)

1.2 Célok A szimulátor célja, hogy az attribútumok esetleges módosításával (Később akár a döntéstámogatáson is túlmutatóan) **elősegítse a filmek potenciális sikerét (vö. produceri hatás-tervezés)**, megbecsülje pl. azok várható nézettségét, valamint IMDB pontszámát, ill. bármely nézettség/IMDB-indexérték-befolyásoló attribútum rel. hiányát, a többlet várható hatását.

1.3 Célcsoportok Célcsoportját alkothatják filmproducerek, rendezők, forgatókönyvírók, színészek, hangmérnökök, dramaturgok, stúdiók vagy akár befektetők, akik érdekeltek lehetnek abban, hogy előre jelezzék egy film várható fogadtatását és bevételét.

1.4 Hasznosság A szimulátor lehetőséget adhat arra, hogy virtuális környezetben különböző forgatókönyvek sikerességét vagy kudarcát modellezzük le fiktív értékelésekkel. Ezáltal jobban megérthetjük a közönség preferenciáit és a trendeket. Ezen előzetes információ segítheti a színészek munkáját, hogy hatékonyabban építsék fel karrierjüket, (Vö. Minél nagyobb a produktum nézettsége, annál ismertebb és/vagy minél jobb az elért IMDB index annál jobb hírnévnek örvendő), de a szimulátor által generált eredmények inspirációt adhatnak a rendezőknek és producereknek is. Sőt a kapott értékelések alapján kialakított marketingstratégiák hatásosabbá tehetik a

filmek vagy sorozatok promócióját, (Vö. Minél jobb az attribútumok eloszlása, annál vonzóbb lesz a produktum.) így **nagyobb nézettséget és profitot eredményezhetnek.**

1.5 A dolgozat szerkezetéről A dolgozat szerkezetének bemutatásakor az alábbi szempontokat kívánom hangsúlyozni. A célcsoport meghatározását és a kutatás hasznosságának körvonalazását követően ismertetem a szakirodalomban fellelhető, hasonló témájú, már létező IMDB-szimulátort, majd bemutatom a saját alternatív megközelítésemet a témában. Külön figyelmet fordítok a saját fejlesztéseim logikai alapvetéseire, a renitens elemek felkutatására, illetve azokra a módszertani megoldásokra, amelyek az IMDB-bebecslési eredmények minél optimálisabb elérését célozzák.

A dolgozatban bemutatott ábrák digitális képek, melyek tetszés szerint nagyíthatók, valamint forrásuk a csatolt mellékletben is eredeti méretükben megtekinthetők.

Fontos kiemelni, hogy a dolgozat keretein belül nem foglalkozom a tanulási modell genetikai potenciáljával kapcsolatos kérdésekkel, illetve azok esetleges korrekciójára irányuló kísérletekkel. Továbbá a terjedelmi korlátok miatt nem kerül részletes bemutatásra a szimulátor nézettség-bebecslési képessége sem. Ezekkel a kérdésekkel kapcsolatban az érdeklődő olvasók a mellékletben találhatnak további információt.

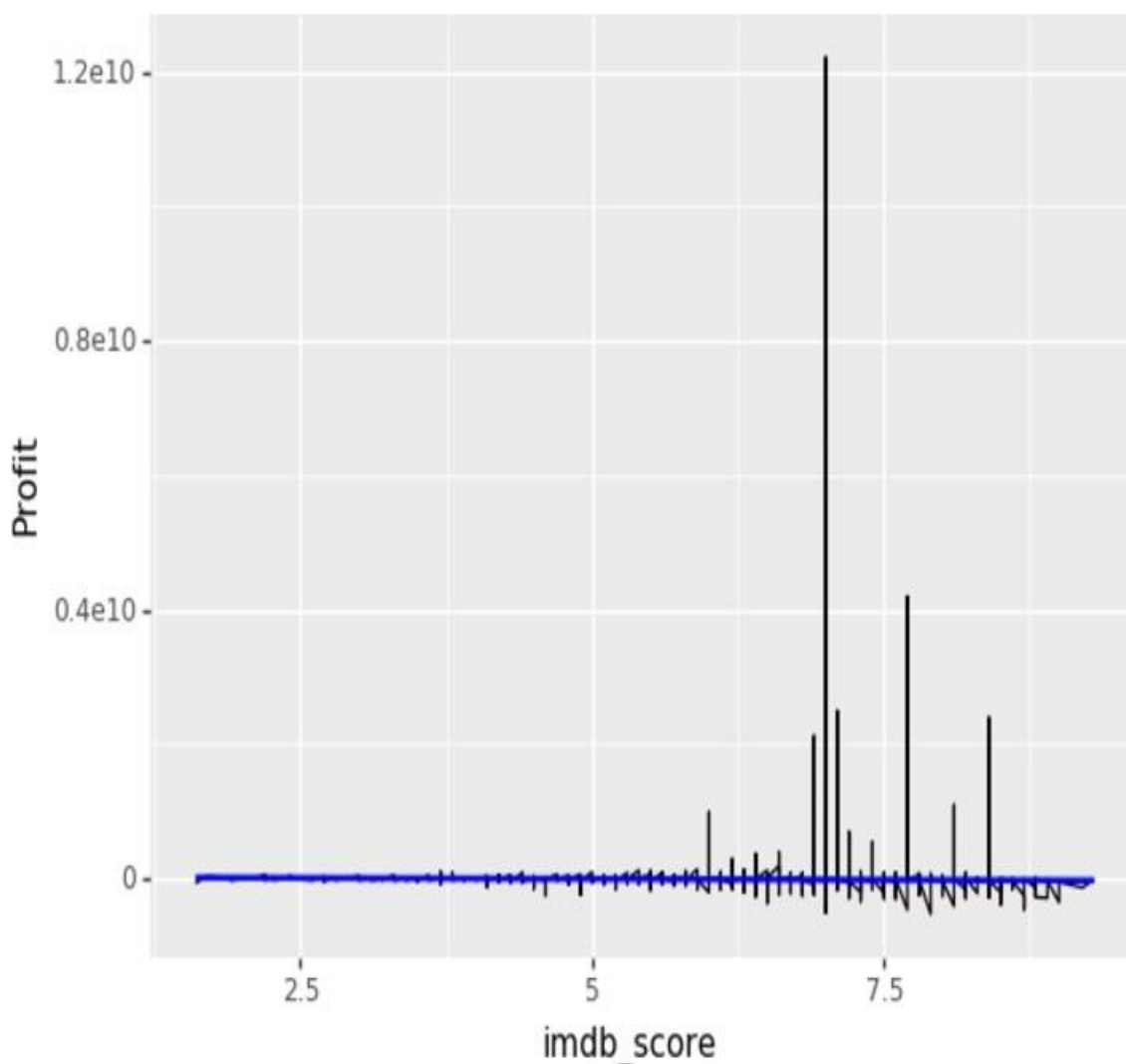
1.6 A megoldásról A szimulátor egy OAM (Objektum-Attribútum Mátrix) -alapú és hasonlóság-elemzés (pl. (<https://miau.my-x.hu/myxfree/coco/index.html>)) által generált rendszer, mely felfogható pl. induktív online szakértői rendszerként is. Az elemzési/szimulátorépítési folyamat több, mint 30 objektum (film) és azok attribútumai alapján összehasonlításokat végez és utólag objektíven tesztelhető mintákat azonosít be.

A szimulátor önellenőrző (ön-validáló) működési fejlesztési elvek mentén készül. Minden egyes rendelkezésre álló attribútum kapcsán (n darab), a mindenkor $n-1$ attribútum kapcsán min. 2-2 elemi (egymást validáló) modell építhető. A modellek láncolata konzisztens rendszerműködésre bírható a későbbiekben bemutatott lépéseken keresztül, vagy a szimulátor képes saját határainak tételes felmutatására/kimutatására.

2. Szakirodalmi háttér A szakdolgozat keretében kitérek Victory (Chibuike) Onumaku által 2023. november 26.-án publikált módszertanra és eredményeire is (benchmark-ként), összehason-

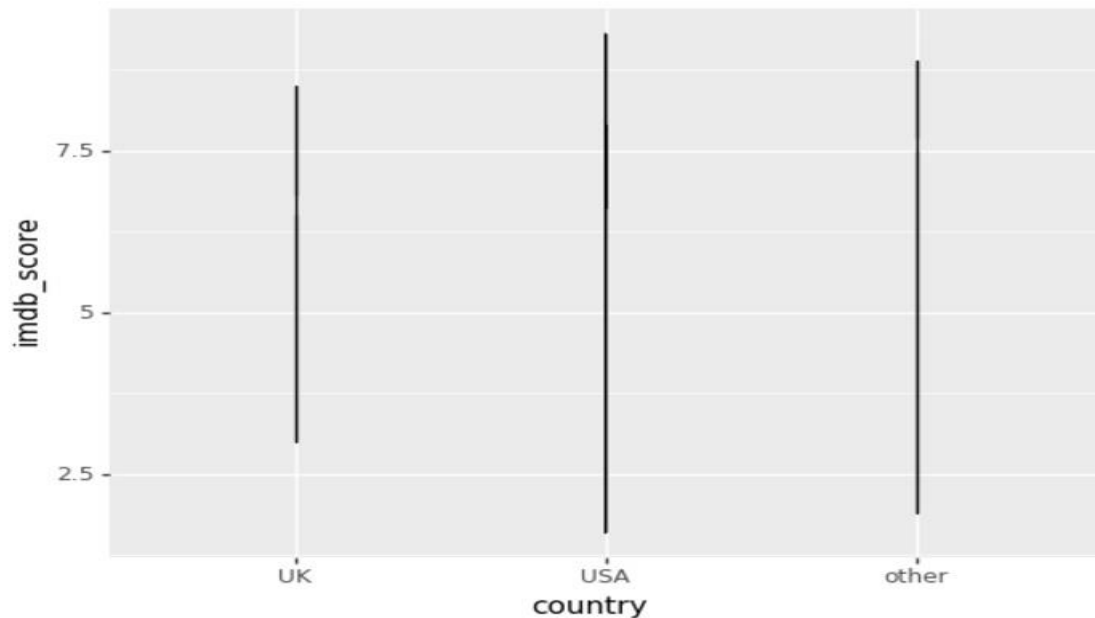
lítva azokat saját kutatási eredményeimmel és megközelítésemmel. Továbbá a szakirodalomban fellelhető kapcsolódó korábbi művészeti és sport előzmények kerülnek fókuszba.

Onmaku az Edge Hill Egyetem Művészeti és Tudományos Karán kutatásokba kezdett egy 5000 filmet tartalmazó adatvagyon segítségével azzal a céllal, hogy megoldást kínáljon a filmipar valós problémáira. Meglátása szerint a filmek potenciális kereskedelmi sikere számos tényezőtől függ, (pl.: rendezők, színészek, kritikusok értékelése, nézők reakciója) vizsgálatai alapján erős korrelációt fedezett fel a filmek IMDB pontszáma és azok profitja között. (Vö. 1. ábra)



1. Ábra: Az IMDB pontszámok és a filmek profitja közötti kapcsolat: A vízszintes tengelyen a filmek IMDB értékelése.. Míg a függőleges tengely a filmek bevételeit adja meg (10^{10} dollárban). – forrás (<https://www.kaggle.com/code/saurav9786/imdb-score-prediction-for-movies>),

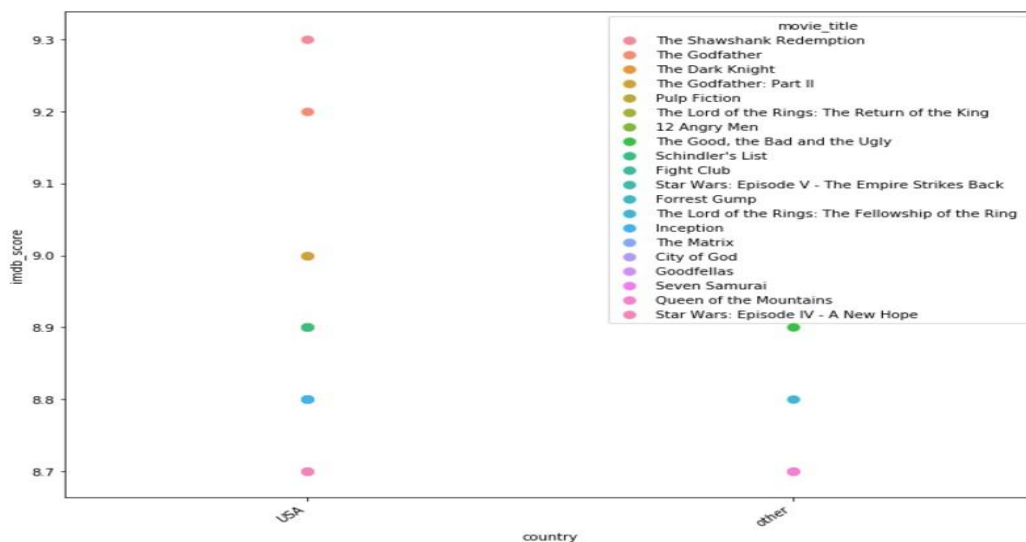
Továbbá, azt, hogy régiós bontásban a legjobb és a legrosszabb filmek többsége az Egyesült Ál-



lamokból származtatható. (Vö. 2. ábra)

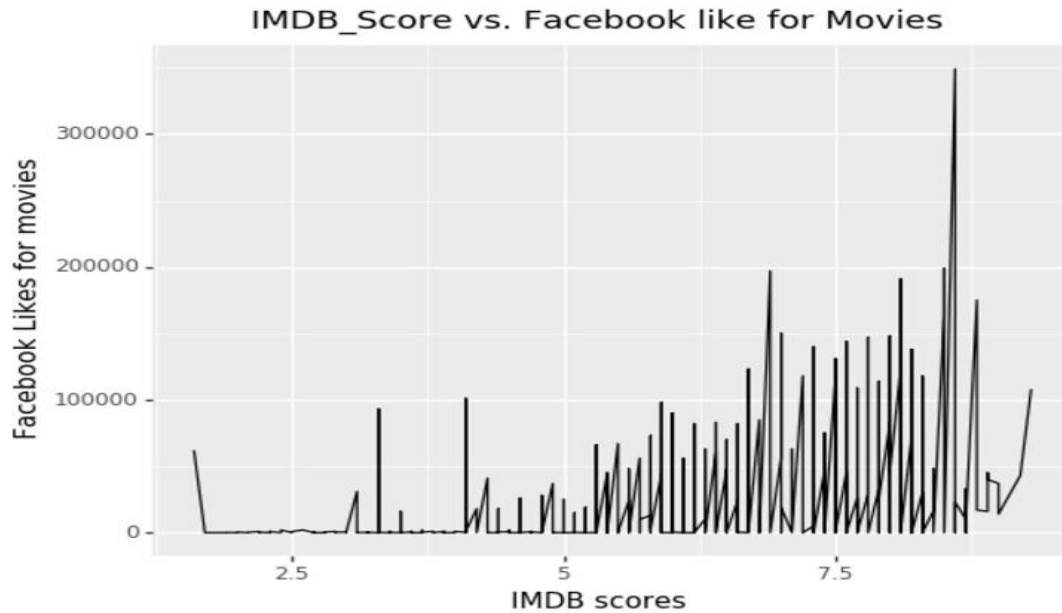
2. Ábra: Filmek IMDb pontszáma régiós bontásban. A 8.75 pontot meghaladó filmek többségét az USA-ban készítették. A vízszintes tengely az országgódot a függőleges tengely az IMDb pontszámot jelöli. --forrás (<https://www.kaggle.com/code/saurav9786/imdb-score-prediction-for-movies>)

Azonban megállapítása szerint, az USA-ban leadott értékelések nagyságrendileg megegyezöek a világ bármely más pontján élöök szavazataival, tehát a nézők véleménye nem területspecifikus. (Vö. 3. ábra)

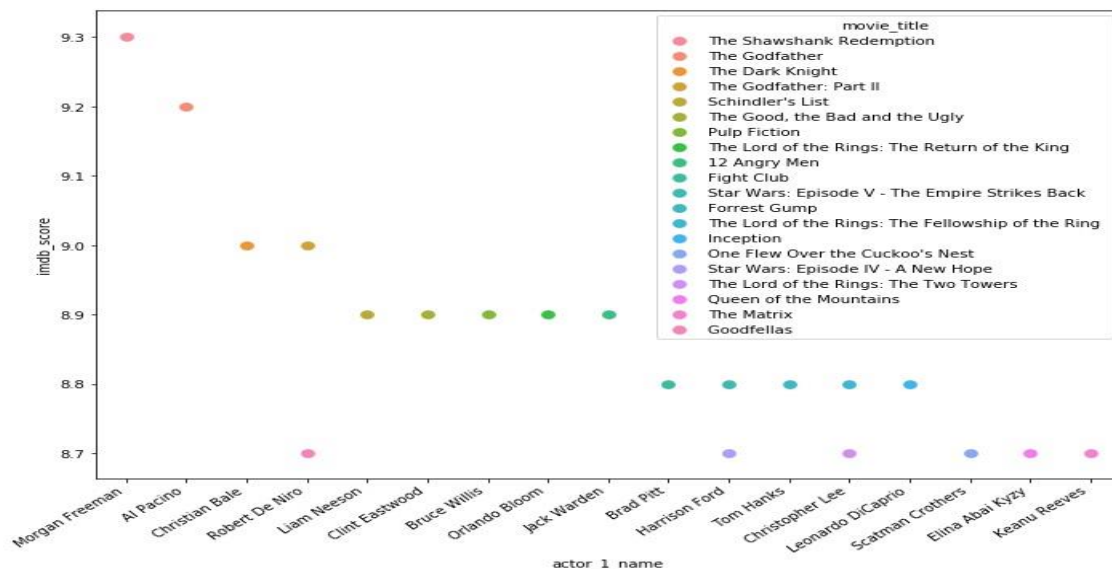


3. Ábra: Az USA és egyéb országok lakóinak szavazatát összehasonlító mutató. A vízszintes tengely az országgódot, a függőleges tengely az IMDb pontszámot, míg a színes jelzések az egyes filmek ezen a skálán való elhelyezkedését jelölik. --forrás (<https://www.kaggle.com/code/saurav9786/imdb-score-prediction-for-movies>)

Kapcsolatot vélt felfedezni a közösségi médiában szerzett like-ok és a produktum későbbi IMDB értékelése között. (Vö. 4. ábra) Arra jutott, hogy a like-okat, így a későbbi profitot és az IMDB



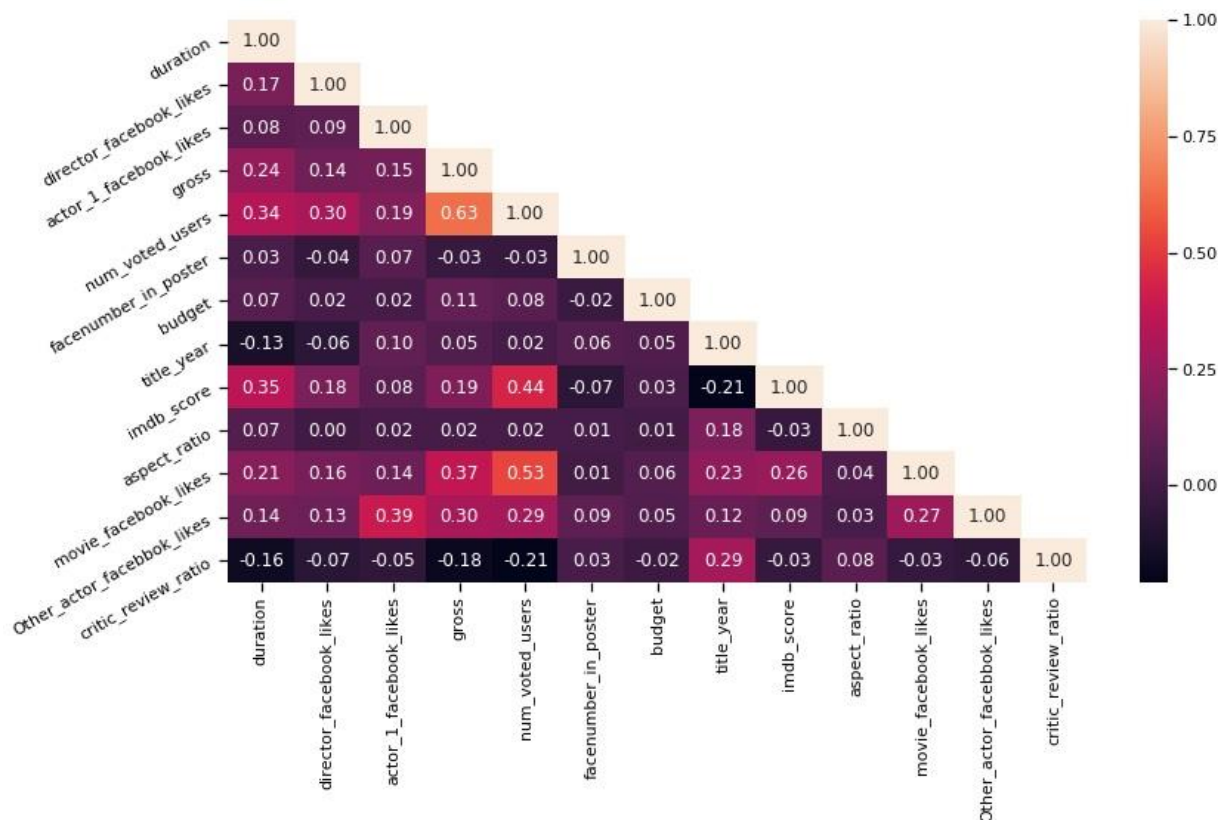
4. Ábra: A facebook like-ok és az IMDb pontszám közt növekvő kapcsolat figyelhető meg. A vízszintes tengely az IMDb pontszámokat, míg a függőleges az egyes filmekre leadott facebook like-okat ábrázolja. --forrás (<https://www.kaggle.com/code/saurav9786/imdb-score-predicti>)



értékelést a stáb ún. „húzó neveinek” (rendezők, színészek) hírneve generálja. (Vö. 5. ábra)

5. Ábra: A kiválasztott színészek és a filmek IMDB pontszám kapcsolatának mutatója. A vízszintes tengely színészek neveit, a függőleges tengely az IMDB pontszámot, A színes jelölők pedig ezek térbeli kapcsolatát ábrázolja. --forrás (<https://www.kaggle.com/code/saurav9786/imdb-score-prediction-for-movies>)

Ezen lépéseket követően a végső attribútum készlete a 6. ábra szerint ábrázolható, melynek segítségével több hasonlóságelemző algoritmust (KNN, SVC, Decision tree, Ada Boosting, Random Forest) alkalmazva hajtotta végre a predikciós kísérleteket.



6. Ábra: Az ábra a vizsgált attribútumok közötti korrelációs együtthatókat mutatja. A színskála a kapcsolat erősségét és irányát jelöli: a világosabb színek erős pozitív, míg a sötétebb színek gyenge vagy negatív korrelációt jeleznek."--forrás (<https://www.kaggle.com/code/saurav9786/imdb-score-prediction-for-movies>)

2.1 Összegzés: A Victory Onomaku által bemutatott IMDb szimulátor elmondása szerint a vizsgált filmek IMDb-értékeit 0,74%-os pontossággal sikerült előre jelezni a Random Forest algoritmus alkalmazásával. Azonban az általa publikált dokumentációkban több lényeges részlet tisztázatlan maradt, ami megnehezíti az eredmények objektív értékelését és reprodukálhatóságát. Például nem derül ki egyértelműen, hogy mit értett 100%-os pontosság alatt. (Vajon a szimulátor

ténylegesen tökéletes egyezést produkált a valós IMDb-értékekkel, vagy bizonyos megengedett hibahatárral számolt?)

Hasonlóan, az általa sikeresnek ítélt szcenáriók esetében nem publikált részletes információkat arról, hogy milyen statisztikai mutatók vagy kritériumok alapján értékelte ezeket az eredményeket sikeresnek. Nem ismert például, hogy a hibahatár milyen mértékben játszott szerepet az értékelésben. Azokban az esetekben pedig, amikor a szimulátor tévedett, az eltérés nagysága szintén nem került dokumentálásra, így nem áll rendelkezésre információ arról, hogy az eltérés szisztematikus vagy véletlenszerű jellegű volt-e, illetve, hogy milyen mértékben befolyásolhatta ez az algoritmus általánosíthatóságát.

Ezen hiányosságok miatt Onomaku eredményei csak korlátozottan értelmezhetők, és a módszertan reprodukálhatósága komoly akadályokba ütközik. Továbbá felmerül a kérdés, hogy a szimulátor fejlesztése során milyen mértékben került sor az algoritmus optimalizálására. Ezek a nyitott kérdések arra utalnak, hogy a publikált eredmények további mélyreható vizsgálatot igényelnek annak érdekében, hogy megbízhatóan összehasonlíthatók legyenek más IMDb-predikációs modellekkel.

Fontos megjegyezni, hogy Onomaku szimulátora nem vizsgálja az objektumot művészeti alkotásként, azaz nem veszi figyelembe annak esztétikai, narratív vagy kulturális értékeit. Ez komoly korlátozást jelent, mivel egy ilyen megközelítés nem képes megfelelően értékelni azon alkotásokat, amelyek egyedi művészi minőségükkel emelkednek ki, de alkotóik vagy közreműködőik kevésbé ismertek, illetve nem rendelkeznek számottevő iparági támogatottsággal.

Egy feltörekvő, széles körben még nem ismert stáb által készített, de korszakalkotónak bizonyuló mű esetében a szimulátor várhatóan alábecsülné a film sikerpotenciálját. Ez a hiányosság rávilágít arra, hogy Onomaku modellje elsősorban a filmipar általánosan elfogadott mintázatait helyezi előtérbe, és lényegében azt sugallja, hogy nem maga az alkotás minősége vagy mondanivalója, hanem az alkotók ismertsége és a környezeti tényezők határozzák meg egy film esélyeit.

Ezzel szemben az általam vizsgált szimulációs megközelítés középpontjában magának a produktumnak a részletes elemzése áll. Ez a módszer arra törekszik, hogy a siker kulcsát magában a műalkotásban azonosítsa, figyelembe véve annak tartalmi, strukturális és esetleg esztétikai jellemzőit. Ezáltal a modell képes lehet a filmes sikeresség pontosabb előrejelzésére, függetlenül az

alkotók ismertségétől vagy a filmipar konvencionális elvárásaitól. Ez nemcsak szélesebb körű alkalmazhatóságot, hanem az előrejelzési pontosság további javítását is ígéri.

2.2 Korábbi művészeti és sport elemzések A *Moneyball* című film Michael Lewis azonos című könyvének adaptációja, amely a baseball világában alkalmazott hagyományos, tapasztalati alapú döntéshozatali módszerek és a statisztikai alapú megközelítések közötti feszültséget vizsgálja. A hagyományos szemlélet helyett, amely az edzők és „szkennelők” tapasztalataira és intuíciójára alapozott, Beane a statisztikai elemzéseket helyezte előtérbe. A filmben az alapvető filozófia az volt, hogy az emberek hajlamosak túlbecsülni a hagyományos teljesítménymutatókat, míg a statisztikák olyan rejtett mintákat tárnak fel, amelyek alapvetően más döntéseket hozhatnak. (Vö. A Go játék szimulátorához hasonlóan, amely képes volt legyőzni a nagymestereket, mivel a hagyományos döntésekkel szemben alternatív, statisztikai alapú döntéseket hozott)

A *Moneyball* által képviselt analitikai szemlélet egyértelmű párhuzamot mutat a filmek IMDB pontszámának előrejelzésére szolgáló szimulátorok működésével, amelyek nem csupán a szubjektív, közönség általi értékelésére támaszkodnak, hanem statisztikai modellek segítségével próbálják meghatározni a filmek várható teljesítményét. Mindkét esetben az **objektív adatfeldolgozás** központi szerepet kap, miközben minimalizálják az intuitív, emberi döntéshozatalból adódó torzításokat.

Gergics Miléna 2021-ben publikált szakdolgozata, amely a *Robotedző* címet viseli, az adatvezérelt döntéstámogatás alkalmazási lehetőségeit vizsgálja a sport területén, különös tekintettel a labdarúgásra. A kutatás célja annak bemutatása volt, hogy a nagy mennyiségű adat elemzése és megfelelő feldolgozása miként járulhat hozzá az eredményesség növeléséhez és a döntéshozatal optimalizálásához a sportban.

A *Robotedző* algoritmusának alapját a játékosok különféle attribútumai képezik, amelyek elemzése révén képes meghatározni az egyes jellemzők közötti összefüggéseket és hatásokat. Az eszköz megállapíthatja például, hogy egy adott játékos esetében egy adott attribútumnak – például fizikai állapotnak, technikai készségnek vagy taktikai érettségnek – milyen mértékű hatása van a piaci értékre, vagy akár más teljesítménymutatókra.

Továbbá, a *Robotedző* prediktív képességei lehetővé teszik a játékos eredményességének előrejelzését, a fejlesztendő területek azonosítását, valamint a potenciális szerep meghatározását a csa-

paton belül. Ez azt jelenti, hogy az algoritmus képes olyan javaslatokat tenni, amelyek segíthetik az edzői stáb döntéshozatalát, például annak eldöntésében, hogy a játékos bekerülhet-e a kezdőcsapatba, vagy mely területeken szükséges további képzés.

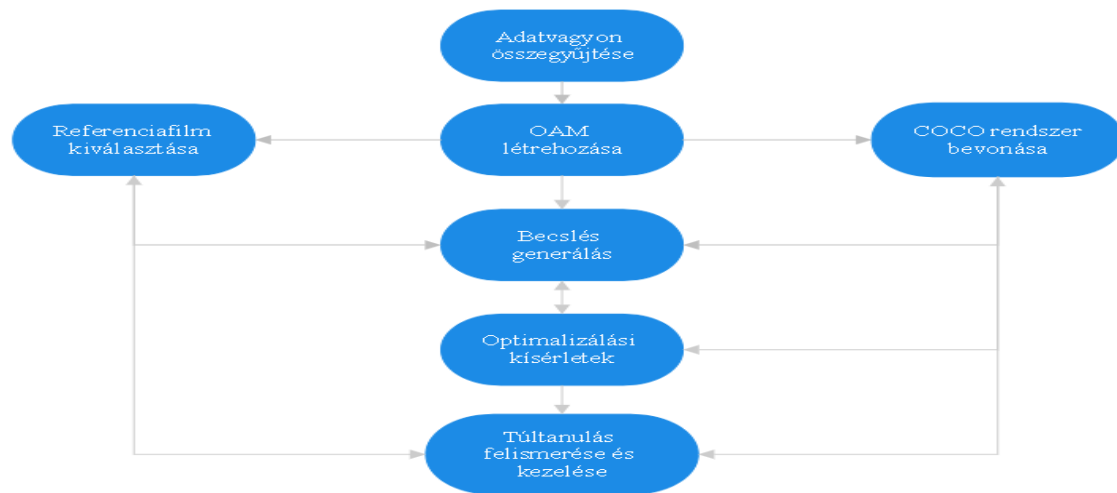
Gergics Miléna munkája rávilágít arra, hogy az adatalapú megközelítések nem csupán a teljesítmény mérését, hanem a döntéshozatali folyamatok finomhangolását is támogathatják, mind a játékosok egyéni fejlődése, mind a csapat stratégiai céljai szempontjából.

Továbbá jelentős párhuzam vonható zenei területen a *Dalfutár* című műsor koncepciójával is. (Vö. https://www.youtube.com/results?search_query=dalfut%C3%A1r)

2.3 Hasonlóságelemzés A hasonlóságelemzés egyik legismertebb gyakorlati alkalmazása a tartalom alapú képkeresés (CBIR), ahol a felhasználó által megadott referenciaképek alapján kereshek hasonló vizuális tartalmak. Az orvosi diagnosztikában például hasonló képletek (pl. daganatos szövetek) azonosításában segíthet. Sőt az autonóm járművek képfeldolgozó rendszerei is gyakran hasonlóságelemzést használnak a tárgyak felismerésére és követésére/elkerülésére.

2.4 Benchmarking A filmiparban a siker előrejelzése általában hagyományos adatelemzési módszereken alapul, mint például: közönségkutatás, tapasztalati becslések vagy marketing kampányok elemzése. Azonban ezek a módszerek bizonyos szempontból korlátozottak, hiszen az elemzések jelentős része emberi döntéshozataltól függ, amely esetben felmerülhet az elfogultság, valamint gyakran nehéz az eltérő formátumú, strukturálatlan adatok, például szöveges értékelések vagy közösségi média interakciók feldolgozása. Ezzel szemben a szimulátor azáltal nyújthat információs többletértéket, hogy nemcsak a strukturált, hanem a strukturálatlan adatokat is képes kombinálni. Ezáltal sokkal komplexebb és pontosabb előrejelzéseket nyújt. Illetve a szimulátor algoritmikus alapú döntéseket hoz, amelyek mentesek az emberi elfogultságtól. Az eredményei bármikor újrafuttathatók, ezáltal transzparenssek és validálhatók.

3. Saját fejlesztések.: Az alábbi pontokban bemutatásra kerül az adatvagyon forrása, a kiválasztás-



7. Ábra: Egyszerűsített folyamatábra, mely a dolgozat keretein belüli lépéseket kívánja beutatni.-(forrás: Saját ábrázolás)

tásának mikéntje, felhasználása, valamint ábrázolása a reprodukálhatóság érdekében.

3.1 Adatgyűjtés Felvázolásra kerül az adatvagyon forrása, kiválasztásának folyamata, ábrázolása és értelmezése. Fontos megjegyezni, hogy A dolgozat ábrái magyar és angol nyelvi elemeket egyaránt tartalmaznak, mivel azok a közeljövőben nemzetközi fórumokon is bemutatásra kerülnek.

3.1.1 Az adatvagyon forrása A projekt objektumaiként az X-akták c. sorozat epizódjai kerültek kiválasztásra, mivel a többszörös Golden-Globe és Emmy díjas produktum megfelelő alapot kínálhat a jót, jobbat, legjobbat halmazok értelmezéséhez és létrehozásához.

A modellhez szükséges adatvagyon egyrészt az IMDB-ről, (<https://www.imdb.com>) másrészt pedig a (<https://www.fandom.com>) híroldalról származtatott. Mivel előbbi a legismertebb és legnagyobb méretű filmes adatbázis, utóbbiban pedig elérhetőek friss nézettségi adatok is.

3.1.2 A kiválasztás folyamata Kezdetben **a vizsgált objektumok 3/4-ét oly módon válogattam ki, hogy azok a spektrum minden szélső értékét tartalmazzák**, hogy egy megfelelően széles skálán tudjak kísérleteket folytatni. Különösképpen a legnézettebb és a legkevésbé nézett epizódok, valamint legmagasabb és a legalacsonyabb IMDB értékkel bíró részek voltak a fókuszban. Azért, hogy a vizsgált objektumok ne legyenek ennyire szélsőségesek, további 1/4 -üket **véletlenszerűen választottam ki**, némi hektikusságot is injektálva az adatvagyonba.

Az attribútumok kiválasztásánál azon érzelmeket előteremtő mozzanatokot dokumentáltam, mely a vizsgált objektumok legalább egyikében több, mint 3%-át kitöltötték a teljes idő intervallumnak. Az elemzés során külön figyelmet fordítottam a másodpercekben mért attribútumok összegére, amelyet a D-O (Duration and Operation) jelölés reprezentál.

Az attribútum értékeket az objektivitás megőrzése érdekében úgy rögzítettem, ahogy azt a

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	type	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	y
2	direction	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	n/a
3	correlation	0.50	0.32	0.51	0.25	-0.49	0.08	0.45	0.08	0.40	-0.06	0.45	-0.06	0.34	0.19	0.60	0.67	1.00
4	unit of measurement	million	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	1/10 score	1/10 score
	Episode name	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation	Reality
5	Detour	22.88	258	211	139	254	40	43	348	0	204	75	47	82	2678	1701	9.2	8.1
6	Redux	27.34	91	184	9	40	0	254	358	220	396	90	0	0	2693	1642	9.2	8.6
7	Redux II	24.84	112	485	24	10	0	153	65	98	267	273	0	20	2798	1507	9.1	9
8	Clyde Bruckman's Final Repose	15.38	235	115	223	30	73	95	290	43	216	125	40	10	2719	1495	9	9.2
9	Paper Clip	17.2	63	156	6	30	0	108	604	30	107	334	0	20	2648	1458	9	9
10	Dreamland II	17.01	56	246	132	35	33	142	163	77	196	231	29	21	2728	1361	9	8.9
11	Bad Blood	19.24	166	27	405	54	70	105	255	0	100	35	97	10	2615	1324	8.9	9.2
12	Musings of a Cigarette Smoking Man	17.09	104	203	12	37	62	205	75	10	65	478	5	15	2640	1271	8.8	9
13	Ghoul	17.38	154	79	56	85	0	80	492	0	220	60	44	0	2630	1270	8.6	5.9
14	Dreamland	17.48	39	100	163	18	106	113	280	115	96	145	2	60	2726	1237	8.3	8.8
15	First Person Shooter	3.64	107	202	34	153	0	99	306	75	233	17	0	0	2693	1226	8.1	7.9
16	Teso Dos Bichos	29.15	117	84	46	101	0	74	324	108	111	224	0	0	2669	1189	8	8.3
17	Leonard Betts	16.6	78	170	0	105	0	148	227	60	166	120	0	0	2729	1074	8	9.1
18	Invocation	13.9	98	28	2	89	0	179	283	35	180	170	0	0	2685	1064	7.9	7.6
19	Anasazi	15	72	94	53	126	0	137	383	0	67	108	0	22	2729	1062	7.6	6.3
20	3	15.31	65	16	84	184	12	61	290	0	198	62	35	0	2679	1007	7.2	6.1
21	My Struggle IV	3.43	22	180	3	190	0	96	88	4	320	55	0	0	2666	958	6.7	7.2
22	Babylon	7.07	61	123	105	34	101	66	109	55	244	57	0	0	2636	955	6.6	6.4
23	Space	10.7	85	97	32	151	0	64	310	0	139	46	8	0	2810	932	6.4	6.2
24	Fight Club	11.7	18	36	112	168	116	61	159	31	225	0	0	0	2684	926	6.4	5.9
25	My Struggle III	5.15	45	2	2	177	0	124	75	0	145	252	0	0	2526	822	6.3	6.4
26	El Mundo Oira	22.37	50	100	46	88	0	69	272	9	176	0	0	0	2688	810	6.2	8
27	Rm9sbG93ZUz	3.23	140	28	77	66	7	22	10	84	65	20	208	0	2667	727	6.1	6.7
28	Nothing Lasts Forever	3.01	18	18	13	41	16	22	194	4	210	152	0	0	2666	688	5.9	6.6
29	Familiar	3.46	97	40	3	60	0	19	231	26	16	193	0	0	2663	685	5.9	8

8. ábra: A hőtésképpel ellátott és az irány figyelembevételével sorbarendezett OAM mátrix. Minél zöldebb, annál pozitívabb hatással bír az Y értékre (forrás: saját ábrázolás, [IMDb szimulátor.xlsx](#), OAM TDK munkalap)

szerző mindenkor gondolta vagy szándékozta, esetleg utalt rá, még akkor is, ha az már nem tűnik aktuálisnak. (pl.: A technológia előrehaladtával egy jelenet, bár már nem minősülne félelmetesnek, de a szerző az adott pillanatban annak szánta, így ennek megfelelően került dokumentálásra. (Vö. <https://www.youtube.com/watch?v=QSLm-IqB0oE>) Az így létrehozott, sorbarendezett és hőtésképpel ellátott korai Objektum-Attribútum Mátrix (nyers adatok) a 8-as ábra szerint ábrázolható:

A fenti 8. ábra megértéséhez az alábbi jelmagyarázat nyújt további segítséget.:

1-es sor: Megadja az attribútumuk típusát az adott oszlopban (Az X értékek változók, az Y valós IMDB érték)

2-es sor: Megmutatja az attribútumok irányát az adott oszlopban (0-ha az irány pozitív tehát minél több időegységet tartalmaz, annál jobb, 1-es amennyiben ennek az ellenkezője vélelmezhető. Az irány kiszámításához a 3. sorban részletezett korrelációt alkalmaztam.)

3-as sor: Megadja a korreláció mértékét azaz, hogy hogyan viszonyul az X az Y célértékhez az adott oszlopban.

4-es sor: Megadja az adott oszlop mértékegységét. (Nézőszámok millióban, jelenetek másodpercben, a vélt és valós imdb pontszámok 1-től 10-ig terjedő skálán kifejezve.)

5-ös sor: A vizsgált attribútumokat és objektumokat nevesíti meg. (Kiválasztásuk mikéntje fentebb a 3.1.2-es pontban részletezve olvasható)

Az ábra **C-N oszlopai** az élmény **mennyiségét fejezik ki másodpercben**, naív értelmezés szerint az érték minél nagyobb, annál jobb, **ezek összeadott értékét a P oszlop mutatja meg**.

Az ezen elv szerinti naív **IMDB elvárásokat a Q oszlop** tartalmazza, míg a **tényleges adatoka az R oszlopban** tekinthetők meg.

4. A modell naiv (nem optimalizált) alkalmazása Alkalmazásra és megértésre kerülnek a bemeneti és kimeneti jelek a szakértői rendszer bevonásával. (folyamatábra, szavak)

4.1 OAM értékek sorba rendezése Az objektum attribútum mátrixából származó kimeneti jelek értelmezéséhez a hasonlóságelemző szakértői rendszer számára elengedhetetlen azok megfelelő átalakítása. Ehhez első lépésként szükséges az objektumok attribútumainak sorba rendezése egy sorszámfüggvény segítségével. Ezt követően az objektumok Y értékét – vagyis a valós IMDB pontszámukat – oly módon kell megadni, hogy azt 1000-rel megszorozzuk, biztosítva ezzel az

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Episode name	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation based on the Q column	Reality
1																		
2	Detour	4	1	2	3	20	5	17	16	15	12	11	3	1	10	1	1	8100
3	Redux	2	11	5	16	5	9	1	17	1	20	10	9	8	14	2	1	8600
4	Redux II	3	7	1	13	1	9	3	2	3	18	3	9	3	19	3	3	9000
	Clyde Bruckman's																	
5	Final Repose	11	2	9	2	2	2	10	11	7	14	7	5	6	16	4	4	9200
6	Paper Clip	8	16	8	17	2	9	6	20	9	6	2	9	3	4	5	4	9000
7	Bad Blood	6	3	18	1	7	3	7	9	15	5	16	2	6	1	6	4	9200
	Musings of a Cigarette																	
8	Smoking Man	9	9	3	15	4	4	2	3	11	2	1	8	5	3	7	7	9000
9	Ghoul	7	4	14	7	10	9	11	19	15	15	13	4	8	2	8	8	5900
	First Person																	
10	Shooter	16	8	4	11	16	9	8	13	5	17	18	9	8	14	9	9	7900
11	Teso Dos Bichos	1	6	13	9	12	9	12	15	2	7	4	9	8	9	10	10	8300
12	Leonard Betts	10	13	7	20	13	9	4	7	6	9	8	9	8	17	11	10	9100
13	Anasazi	13	14	12	8	14	9	5	18	15	4	9	9	2	17	12	12	6300
14	3	12	15	20	5	18	7	15	11	15	11	12	6	8	11	13	13	6100
15	My Struggle IV	18	18	6	18	19	9	9	4	13	19	14	9	8	6	14	14	7200
16	Space	15	12	11	12	15	9	14	14	15	8	15	7	8	20	15	15	6200
17	Fight Club	14	19	16	4	17	1	15	5	8	16	19	9	8	12	16	16	5900
18	Rm9sbG93ZDÜz	19	5	17	6	9	8	18	1	4	2	17	1	8	8	18	17	8000
19	El Mundo Gira	5	17	10	9	11	9	13	10	12	10	19	9	8	13	17	18	6700
	Nothing Lasts																	
20	Forever	20	19	19	14	6	6	18	6	13	13	6	9	8	6	19	19	6600
21	Familiar	17	10	15	18	8	9	20	8	10	1	5	9	8	5	20	19	8000

9. ábra: Az ábrán a rangsorolást követő állapot látható, miután az R oszlop, azaz a filmek valós IMDB eredményeit az 1000 szorozásra növeltük. A sorbarendezést az irány határozta meg. Minél kisebb a sorszám a korrelációs értékek szerint, annál több befolyással bír az IMDB értékelésre. (forrás: saját ábrázolás, [IMDb szimulátor.xlsx](#) Sorbarendezett OAM TDK munkalap)

adatok megfelelő skálázását.

Az így előállított, sorbarendeztett **B-R oszlop értékei** már kompatibilisek a COCO rendszer bemeneti jelrendszerével, így azokat a rendszer megfelelően tudja feldolgozni és értelmezni. Ez az átalakítási folyamat kulcsfontosságú a rendszer működésének pontossága és hatékonysága szempontjából. (Vö 9. ábra)

4.1.1 Az online szakértői rendszer bevonása A sorbarendeztett B-R oszlop értékeit betápláljuk a COCO STD felületén (https://miau.my-x.hu/myx-free/coco/beker_std.php), hogy az hasonlóság-

The screenshot shows the COCO STD web interface. At the top left is the COCO logo and version information: "Verzió: v2.17 (2012.10.25.)". The interface is divided into several sections:

- Azonosító**: A text input field for identification.
- Mátrix**: A table with 8 rows and 8 columns of numerical values. The values are: Row 1: 19, 10, 4, 1, 10, 25, 21, 21; Row 2: 5, 20, 12, 12, 14, 12, 17, 13; Row 3: 16, 12, 24, 1, 10, 10, 22, 22; Row 4: 24, 5, 20, 9, 12, 11, 23, 25; Row 5: 5, 2, 22, 25, 10, 16, 23, 23; Row 6: 25, 24, 23, 17, 9, 9, 23, 17; Row 7: 17, 17, 9, 1, 10, 17, 24, 24; Row 8: 22, 11, 18, 21, 11, 12, 25, 15; Row 9: 14, 1, 7, 1, 10, 19, 25, 24.
- Vizuális beállítások**: A section with two tabs: "Objektum elnevezések" and "Attribútum elnevezések". Below the tabs is a large empty text area.
- A rangsor mátrix megadása kötelező!**: A section with a "Lépcsők" input field.
- Modell**: A dropdown menu set to "Standard".
- Fájlok megtartása**: A checkbox that is currently unchecked.
- Futtatás** and **Alapállapot** buttons at the bottom right.

10. ábra: A Coco standard modell kezelőfelületén a felhasználó betáplálja a sorszámozott értékeket (forrás: https://miau.my-x.hu/myx-free/coco/beker_std.php)

elemzésen alapuló mintákat azonosítson. (Vö 10. ábra)

4.1.2 A szakértői ábrázolás A futtatást követően **megtekinthetjük a rendszer által végzett számításokat, súlyozott értékeket, becsült és tény adatokat**, valamint **az ezek közti eltérést a delta/tény oszlopban**. Továbbá **az általa vélelmezett skála maximumát**, avagy a szakértői rendszer szerint benne rejlő genetikai potenciált, pedig az S1 összeg cella jelöli. (Vö 11. ábra)

4.1.3 Kimenő jelek értelmezése Az így végzett predikciós kísérletek eredményeit a delta/tény oszlop alapján értékeljük, az eredmények akkor tekinthetők optimálisnak ha az értékek minél inkább a 0-hoz közelítenek. (Vö 11. ábra)

4.2 A modell gyakorlati alkalmazása A predikciós feladatok alkalmazását és az ebből származó eredményeket demonstrálok.

COCO-STD	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation based on the Q column	Becslés	Tény=0	Delta	Delta/Tény
Detour	0	767	1367	67	0	5833	0	0	0	0	0	67	0	0	0	0	8101	8100	-1	-0.01
Redux	0	0	1367	0	0	5833	700	0	700	0	0	0	0	0	0	0	8600	8600	0	0
Redux II	0	0	1367	0	0	5833	333	0	700	0	767	0	0	0	0	0	9000	9000	0	0
Clyde Bruckman's Final Repose	0	767	733	1067	0	5833	0	0	0	0	733	67	0	0	0	0	9200	9200	0	0
Paper Clip	0	0	1367	0	0	5833	333	0	0	133	1333	0	0	0	0	0	8999	9000	1	0.01
Bad Blood	0	0	0	3167	0	5833	0	0	0	133	0	67	0	0	0	0	9200	9200	0	0
Musings of a Cigarette Smoking Man	0	0	1367	0	0	5833	333	0	0	133	1333	0	0	0	0	0	8999	9000	1	0.01
Ghoul	0	0	0	0	0	5833	0	0	0	0	0	67	0	0	0	0	5900	5900	0	0
First Person Shooter	0	0	1367	0	0	5833	0	0	700	0	0	0	0	0	0	0	7900	7900	0	0
Teso Dos Bichos	867	0	0	0	0	5833	0	0	700	133	767	0	0	0	0	0	8300	8300	0	0
Leonard Betts	0	0	1367	0	0	5833	333	0	700	133	733	0	0	0	0	0	9099	9100	1	0.01
Anasazi	0	0	0	0	0	5833	333	0	0	133	0	0	0	0	0	0	6299	6300	1	0.02
3	0	0	0	67	0	5833	0	0	0	133	0	67	0	0	0	0	6100	6100	0	0
My Struggle IV	0	0	1367	0	0	5833	0	0	0	0	0	0	0	0	0	0	7200	7200	0	0
Space	0	0	167	0	0	5833	0	0	0	133	0	67	0	0	0	0	6200	6200	0	0
Fight Club	0	0	0	67	0	5833	0	0	0	0	0	0	0	0	0	0	5900	5900	0	0
Rm9sbG93ZDZ	0	0	0	0	0	5833	0	1267	700	133	0	67	0	0	0	0	8000	8000	0	0
El Mundo Gira	0	0	733	0	0	5833	0	0	0	133	0	0	0	0	0	0	6699	6700	1	0.01
Nothing Lasts Forever	0	0	0	0	0	5833	0	0	0	0	767	0	0	0	0	0	6600	6600	0	0
Familiar	0	0	0	0	0	5833	0	0	0	1400	767	0	0	0	0	0	8000	8000	0	0
SI összeg:	17468																			

11. ábra: Coco standard modell futtatást követő állapota, mely felvázolja a hasonlóságelemzéssel vizsgált 20 objektum és azok attribútumain végzett mintázat felkutató értékeléseket és az általa vélelmezett genetikai potenciált. Az attribútumok mértékegysége az objektumok IMDb indexe ezerrel megszorozva. (forrás: <https://miau.my-x.hu/myx-free/coco/engine3.php>)

4.2.1 A megbecsülni kívánt objektum kiválasztása Ahhoz, hogy megfelelő képet kapjunk a szimulátor várható hatásosságáról érdemes önellenőrzésként elsőként egy már meglévő objektumot kihagyni a képletből és ezen objektumot megkísérelni megbecsülni. Azaz szándékosan adathiányt generálni.

4.2.3 A becslés menete A kiválasztott megbecsülni kívánt objektum kihagyásával a sorszámozott adatok rögzítésre kerülnek a szakértői rendszerben, miközben a lépcsők száma a kiválasztott objektumokhoz képest 1 értékkel növelésre kerül. A futtatási folyamat lezárását követően az eredmények Excelbe másolása történik meg. Az adatok rendezésénél a Lépcsők (2) oszlop legutolsó értékét követő üres sor cellái 1-től növekvő sorrendben kerülnek beszámozásra, az átláthatóság és

további feldolgozhatóság érdekében (vö. 12. ábra)

	Lépcső(2)	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation based on the Q column
101	Detour	887	767	1367	1167	0	5833	700	1267	700	133	1333	67	0	0	0	0
102	Redux	0	767	1367	1067	0	5833	333	0	700	133	1333	67	0	0	0	0
103	Redux II	0	0	1367	67	0	5833	333	0	700	133	767	67	0	0	0	0
104	Clyde Bruckman's Final Repose	0	0	1367	67	0	5833	333	0	700	133	767	67	0	0	0	0
105	Paper Clip	0	0	1367	67	0	5833	333	0	700	133	767	67	0	0	0	0
106	Bad Blood	0	0	1367	0	0	5833	333	0	700	133	767	67	0	0	0	0
107	Musings of a Cigarette Smoking Man	0	0	1367	0	0	5833	0	0	0	133	733	67	0	0	0	0
108	Ghoul	0	0	1367	0	0	5833	0	0	0	133	733	0	0	0	0	0
109	First Person Shooter	0	0	733	0	0	5833	0	0	0	133	0	0	0	0	0	0
110	Teso Dos Bichos	0	0	733	0	0	0	0	0	0	133	0	0	0	0	0	0
111	Leonard Betts	0	0	167	0	0	0	0	0	0	133	0	0	0	0	0	0
112	Anasazi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
113	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
114	My Struggle IV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
115	Space	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
116	Fight Club	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
117	Rm9sbG93ZUz	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
118	El Mundo Gira	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
119	Nothing Lasts Forever	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
120	Familiar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121																	
122	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

12. ábra: Az adathiány generálást követően megbecsülni kívánt +1.-ik objektum berögzítése az excelbe. (122.színnel jelölt sor) A 2-es,3-as stb. számok a referenciafilm attribútum értékeire hivatkoznak. (forrás: saját ábrázolás, [IMDb szimulátor.xlsx](#) OAM 20 munkalap, B-R oszlop 101-122-es tartomány)

A becslési folyamat során az OAM sorszámozott formájában előkészítésre kerül egy sor az értékek összesítéséhez, amely a szakértői rendszer mintázatelemzése alapján valószínűsíti az IMDB eredményt. Ez az FKERES függvény alkalmazásával történt, amely a vizsgált objektum rangso-

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	type	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	y
2	direction	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	n/a
3	correlation	0.31	0.44	0.46	0.21	-0.50	0.08	0.44	-0.10	0.43	-0.09	0.51	0.12	0.23	-0.07	0.54	0.62	1.00
4	unit of measurement	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank	y0
5	Episode name	Viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O	Naive estimation	rating *1000
27	Detour	4	1	2	3	20	5	17	16	15	12	11	3	1	10	1	1	8100
28	Redux	2	11	5	16	5	9	1	17	1	20	10	9	8	14	2	1	8600
29	Redux II	3	7	1	13	1	9	3	2	3	18	3	9	3	19	3	3	9000
30	Clyde Bruckman's Final Repose	11	2	9	2	2	2	10	11	7	14	7	5	6	16	4	4	9200
31	Paper Clip	8	16	8	17	2	9	6	20	9	6	2	9	3	4	5	4	9000
32	Bad Blood	6	3	18	1	7	3	7	9	15	5	16	2	6	1	6	4	9200
33	Musings of a Cigarette Smoking Man	9	9	3	15	4	4	2	3	11	2	1	8	5	3	7	7	9000
34	Ghoul	7	4	14	7	10	9	11	13	15	15	13	4	8	2	8	8	5900
35	First Person Shooter	16	8	4	11	16	9	8	13	5	17	18	9	8	14	9	9	7900
36	Teso Dos Bichos	1	6	13	9	12	9	12	15	2	7	4	9	8	9	10	10	8300
37	Leonard Betts	10	13	7	20	13	9	4	7	6	9	8	9	8	17	11	10	9100
38	Anasazi	13	14	12	8	14	9	5	18	15	4	9	9	2	17	12	12	6300
39	3	12	15	20	5	18	7	15	11	15	11	12	6	8	11	13	13	6100
40	My Struggle IV	18	18	6	18	19	9	9	4	13	19	14	9	8	6	14	14	7200
41	Space	15	12	11	12	15	9	14	14	15	8	15	7	8	20	15	15	6200
42	Fight Club	14	19	16	4	17	1	15	5	8	16	19	9	8	12	16	16	5900
43	Rm9sbG93ZUz	19	5	17	6	9	8	18	1	4	2	17	1	8	8	18	17	8000
44	El Mundo Gira	5	17	10	9	11	9	13	10	12	10	19	9	8	13	17	18	6700
45	Nothing Lasts Forever	20	19	19	14	6	6	18	6	13	13	6	9	8	6	19	19	6600
46	Familiar	17	10	15	18	8	9	20	8	10	1	5	9	8	5	20	19	8000
47	Estimate	0	0	0	0	0	5833	0	0	0	133	767	0	0	0	0	0	6733

13. ábra: A 47. sorban megtekinthetők a szakértői rendszer által megbecsült +1. objektum (referenciafilm) attribútum értékei (C-R sor), valamint az ezekből származtatott várható IMDB index érték 1000 szorosára növelt alakját (S oszlop). Az objektum valódi IMDB értéke (*1000) az 46. sor S oszlopában látható. (forrás: saját ábrázolás [IMDb szimulátor.xlsx](#) OAM 20 munkalap B-S 27-47 tartomány)

rolt nézettségére hivatkozik. Az adatkeresési folyamat során a lépcsők (2) kezdő és végpontja rögzítésre kerül, biztosítva a cellahivatkozások stabilitását.

Az attribútumok irányának meghatározása (0 vagy 1) és a hivatkozások pontosítása után a függvényt az összes becsült objektumra alkalmazzuk. Az így kapott értékek összege adja meg a szakértői rendszer által előre jelzett IMDB eredményt, amely a mintázatok és a rangsorolási folyamat integrált eredményét tükrözi. (Vö. 13. ábra)

4.3. Kezdeti eredmények bemutatása Ismertetésre kerülnek a korai kísérletek eredményei, a felmerült problémák, valamint az azok orvoslására tett kísérletek.

4.3.1. Eredmények és a problémák felvázolása A1 vizsgált 20 objektum esetében a szakértői rendszer által számított delta/tény érték szinte hibátlan eredményt mutatott (Vö. 10. ábra). Azonban amikor a szimulátort megkísérelte a becslést, az 1,3 ponttal elmaradt a valós IMDB értéktől (lásd 12. ábra, 52-53. sor, R oszlop). Ennek több lehetséges oka lehet pl.: Az objektumok korlátozott száma (Nem elegendő mennyiségű az adatvagyon) miatt nem tudtunk pontosabb becslést végezni, vagy valamely vizsgált objektum olyannyira a többitől eltérő, hogy az befolyásolta a szimulátor tanulási mintázatát. Ezen kívül a kapott genetikai potenciál 7468 értékkel meghaladta a kívánt célt (Vö. 11. ábra).”

4.3.2 Út az optimalizálás felé (az adatvagyon bővítésével) Az el nem ért célértékek egyik lehetséges magyarázata, hogy az adatvagyon mennyisége nem bizonyult elégségesnek, ezért további 11 objektummal bővítettem az Objektum Attribútum Mátrixot (Vö. 3.1.2-es pont). Azonban ez a hektikus többletérték csak tovább növelte a szakadékokat.: Mind a genetikai potenciál, mind a ka-

COCO:STD	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation	Becslés	Tény+0	Delta	Delta/Tény
The Truth	0	765.2	1172.9	0	0	1472.4	765.2	0	357.9	83.4	723.7	1780.4	0	1095.1	0	0	8186.1	8200	13.9	0.17
Detour	315.9	1980	1172.9	1306.2	0	1472.4	0	41.4	0	0	0	1148	648.9	0	0	0	8005.8	8100	14.2	0.18
Redux	315.9	765.2	1172.9	0	2079.9	1472.4	0	41.4	357.9	83.4	0	1780.4	0	648.9	0	0	8718.2	8600	-118.2	-1.37
Redux II	315.9	765.2	1172.9	0	2828.5	1472.4	0	0	357.9	0	723.7	1780.4	0	648.9	0	0	10065.9	9000	-1065.9	-11.84
Clyde Bruckman's Final Repose	315.9	1980	0	1680.6	2079.9	1472.4	0	41.4	357.9	0	607.4	0	0	648.9	0	0	9184.4	9200	15.6	0.17
Paper Clip	315.9	765.2	0	0	2079.9	1472.4	0	840	357.9	0	723.7	1780.4	0	648.9	0	0	8984.2	9000	15.8	0.18
Dreamland II	315.9	765.2	1172.9	149.7	2079.9	1472.4	0	0	357.9	0	723.7	1197.9	0	648.9	0	0	8884.4	8900	15.6	0.18
Bad Blood	315.9	1980	0	3294.2	2079.9	1472.4	0	41.4	0	0	0	0	0	0	0	0	9183.9	9200	16.1	0.18
Nuisings of a Cigarette Smoking Man	315.9	765.2	1172.9	0	2079.9	1472.4	0	0	357.9	0	1622.2	1197.9	0	0	0	0	8984.2	9000	15.8	0.18
Ghouli	315.9	1980	0	0	2079.9	1472.4	0	41.4	0	0	0	0	0	0	0	0	8889.7	8900	10.3	0.17
Dreamland	315.9	382.3	0	1680.6	2079.9	1472.4	0	41.4	357.9	0	607.4	1197.9	0	648.9	0	0	8784.6	8800	15.4	0.18
First Person Shooter	0	765.2	1172.9	0	1647.1	1472.4	0	41.4	357.9	0	0	1780.4	0	648.9	0	0	7886.2	7900	13.8	0.17
Teso Des Biches	499.1	1172.9	0	0	1672.1	1472.4	0	41.4	357.9	0	640.4	1780.4	0	648.9	0	0	8285.5	8300	14.5	0.17
Darkness Falls	0	1980	0	0	2079.9	1472.4	0	41.4	357.9	0	607.4	1197.9	0	648.9	0	0	8385.8	8400	14.2	0.17
Lazarus	0	765.2	0	0	1672.1	1472.4	0	41.4	0	0	640.4	1780.4	0	648.9	0	0	7020.7	7000	-20.7	-0.3
Leonard Betts	315.9	765.2	1172.9	0	1672.1	1472.4	0	41.4	357.9	0	607.4	1780.4	0	648.9	0	0	8834.5	9100	265.5	2.92
Invocation	0	765.2	0	0	1913.6	1472.4	0	41.4	357.9	0	607.4	1780.4	0	648.9	0	0	7587.2	7600	12.8	0.17
Anasazi	0	765.2	0	0	1647.1	1472.4	0	41.4	0	0	0	1780.4	0	648.9	0	0	6355.4	6300	-55.4	-0.88
Essence	0	1655.6	0	0	2079.9	1472.4	0	41.4	0	0	607.4	1780.4	0	648.9	0	0	8286	8300	14	0.17
3	315.9	765.2	0	0	1647.1	1472.4	0	41.4	0	0	0	1197.9	0	648.9	0	0	6088.8	6100	11.2	0.18
Ny Struggle IV	0	382.3	1172.9	0	1647.1	1472.4	0	0	0	83.4	0	1780.4	0	648.9	0	0	7187.4	7200	12.6	0.18
Babylon	0	765.2	0	0	2079.9	1472.4	0	0	357.9	0	0	1780.4	0	0	0	0	6455.7	6400	-55.7	-0.87
Trust no1	0	765.2	0	0	2079.9	1472.4	0	0	0	0	640.4	1780.4	0	648.9	0	0	7387	7400	13	0.18
Space	0	765.2	0	0	1647.1	1472.4	0	41.4	0	0	0	1197.9	0	1095.1	0	0	6189.1	6200	10.9	0.18
Fight Club	0	0	0	0	1647.1	1655.6	0	0	357.9	0	0	1780.4	0	648.9	0	0	6889.8	6900	-189.8	-3.22
Ny Struggle III	0	765.2	0	0	1647.1	1472.4	0	0	0	0	723.7	1780.4	0	0	0	0	6388.8	6400	11.2	0.18
El Hundo Gino	315.9	765.2	0	0	2079.9	1472.4	0	41.4	357.9	0	0	1780.4	0	648.9	0	0	7461.9	8000	538.1	6.73
Rm9sbG93ZXZz	0	1980	0	0	2079.9	1472.4	0	0	357.9	0	0	0	0	648.9	0	0	6539	6700	161	2.4
Nothing Lasts Forever	0	0	0	0	2079.9	1472.4	0	0	0	0	607.4	1780.4	0	648.9	0	0	6588.9	6600	11.1	0.17
Familiar	0	765.2	0	0	2079.9	1472.4	0	41.4	357.9	0	640.4	1780.4	0	648.9	0	0	7786.3	8000	213.7	2.67
Mulder and Scully Meet the Were-Monster	0	0	0	0	1680.6	2828.5	1655.6	0	0	0	0	0	0	0	0	0	7945.1	8000	54.9	0.69

pott delta/tény érték romlott a korábbi állapotokhoz képest (Vö.8. és 14. ábra).

Az immáron bővült, de romló mintázatú modellel végzett becslési kísérlet eredménye így természetesen romlott (Vö 15. ábra). Ennek elhárítása érdekében újabb módszerek kerültek alkalmazásra az optimalizálási kísérletek során.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R				
1	type	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	y				
2	direction	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	n/a				
3	correlation	0.47	0.32	0.45	0.23	-0.40	0.08	0.38	0.06	0.38	0.05	0.39	-0.07	0.33	0.10	0.42	0.59	1.00				
4	unit of measurement	million					second	second		second		second	second	second	second	second	1/10 score	1/10 score				
5	Episode name	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation	Reality				
41	The Truth	17	10	2	17	28	14	1	22	8	1	2	1	3	1	1	1	8200				
42	Detour	4	1	4	4	29	7	27	5	22	13	19	27	1	17	2	1	8100				
43	Redux	2	14	7	22	8	16	2	4	1	2	18	1	11	10	3	3	8600				
44	Redux II	3	9	1	16	1	16	5	28	4	5	4	1	6	3	4	4	9000				
45	Clyde Bruckman's Final Repose	13	2	13	2	3	4	19	10	11	11	15	25	9	9	5	4	9200				
46	Paper Clip	9	21	10	23	3	16	15	1	15	24	3	1	6	23	6	4	9000				
47	Dreamland II	11	23	3	5	6	8	9	20	6	16	6	23	5	6	7	7	8900				
48	Bad Blood	6	4	26	1	10	5	16	17	22	25	25	28	9	28	8	8	9200				
49	Musings of a Cigarette Smoking Man	10	12	5	19	7	6	3	26	16	28	1	21	8	25	9	9	9000				
50	Ghoul	8	5	19	10	14	16	20	2	22	10	21	26	11	27	10	10	5900				
51	Dreamland	7	26	14	3	2	2	13	14	2	26	13	19	2	7	11	11	8800				
52	First Person Shooter	27	11	6	14	23	16	17	9	7	8	27	1	11	10	12	11	7900				
53	Teso Dos Bichos	1	8	18	12	19	16	21	7	3	23	7	1	11	18	13	13	8300				
54	Darkness Falls	19	3	20	21	12	11	8	12	13	15	14	20	11	24	14	14	8400				
55	Lazarus	20	16	21	19	18	14	6	6	18	19	9	1	11	22	15	15	7000				
56	Leonard Betts	12	18	9	29	20	16	7	18	9	20	16	1	11	4	16	15	9100				
57	Invocation	16	13	24	27	17	16	4	13	12	17	10	1	11	13	17	15	7600				
58	Anasazi	15	19	17	11	21	16	10	3	22	27	17	1	4	4	18	18	6300				
59	Essence	18	7	23	23	14	12	14	15	19	6	11	1	11	8	19	19	8300				
60	3	14	20	28	8	26	10	25	10	22	14	20	11	16	20	20	20	6100				
61	My Struggle IV	29	27	8	26	27	16	18	24	19	3	23	1	11	20	21	21	7200				
62	Babylon	25	22	11	7	5	3	23	23	10	7	22	1	11	26	22	22	6400				
63	Trust no1	23	16	12	25	13	16	11	25	22	4	9	1	11	15	23	23	7400				
64	Space	22	15	16	15	22	16	24	8	22	22	24	22	11	2	24	24	6200				
65	Fight Club	21	28	22	6	24	1	25	21	13	9	28	1	11	14	25	25	5900				
66	My Struggle III	26	25	29	27	25	16	12	26	22	21	5	1	11	29	26	25	4400				
67	El Mundo Gira	5	24	14	12	16	16	22	16	17	18	28	1	11	12	27	27	8000				
68	Rm9sbG93Dz	30	6	24	9	11	12	28	29	5	28	26	29	11	19	28	28	6700				
69	Nothing Lasts Forever	31	28	27	18	9	9	28	19	19	12	11	1	11	20	29	29	6600				
70	Familiar	28	14	22	26	11	16	30	18	16	30	9	1	11	22	30	30	8000				
71	Jider and Scully Meet the Were-Mon	24	31	24	3	1	1	30	30	23	100	25	0	29	1	3330	11	27	91	30	8000	
72	Estimate:	0	1041	0	0	1276	1	349	27	0	100	100	25	0	29	1	3330	11	27	91	30	8000

15.ábra: A, megnövelt objektumszámmal megbecsült referencia epizód (sárgával jelölt) és annak a szakértői rendszer számításai szerint várható IMDB index értéke *1000 (R oszlop) A differencia már 1,6 pontot meghaladó (forrás: saját ábrázolás, [IMDb szimulátor.xlsx](#), Familiar becslés munkalap A-R 41-72 tartomány)

4.3.3 Út az optimalizálás felé (a Renitens objektumok kiszűrésével)

Mivel az objektumok számának növelése nem növelte a becslési pontosságot (Vö 13. és 15. ábra), szükségessé vált a már meglévő objektumok alapos vizsgálata. Meg kellett találni azokat az eseteket, amelyek szélsőséges értékeket eredményezhetnek az adatvagyonban, így torzítva a becslési potenciált. A COCO Y0 modellt használatával. (https://miau.my-x.hu/myx-free/coco/beker_y0.php) megvizsgálható, hogy lehet-e minden objektum másként egyforma.

COCO:Y0	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	0-0 Value	Naive estimation	Becslés	Tény-s	Delta	Delta/Tény
The Truth	14	21	29	36.5	16182.5	16385.9	30	28	23	44	90	157.5	39.5	188.5	30	30	33229.3	33333	3.7	0.01
Detour	27	205	27	99	16181.5	16392.9	4	45	9	18	12	4	213	33	29	30	33329.3	33333	3.7	0.01
Redux	29	17	24	31.5	16547.9	16347.5	29	46	30	43	13	157.5	20	40	28	28	33431.3	33333	-98.3	-0.29
Redux II	28	22	30	37.5	16554.9	16347.5	26	3	27	26	88	157.5	25	47	27	27	33473.3	33333	-140.3	-0.42
Clyde Bruckman's Final Repose	18	97.5	10	101	16552.9	16395.9	12	40	20	20	16	6	22	41	26	27	33413.3	33333	-80.3	-0.24
Paper Clip	22	10	21	30.5	16552.9	16347.5	16	65	16	7	89	157.5	25	8	25	27	33419.3	33333	-86.3	-0.26
Dreamland II	20	8	28	98	16549.9	16391.9	22	30	25	15	43.5	68	26	44	24	24	33417.3	33333	-84.3	-0.25
Bad Blood	25	95.5	5	143.5	16545.9	16394.9	15	33	9	6	6	3	22	3	23	23	33352.8	33333	-19.8	-0.06
Musings of a Cigarette Smoking Man	21	19	26	34.5	16548.9	16393.9	28	24	15	3	134.5	70	23	6	22	22	33390.8	33333	-57.8	-0.17
Ghouli	23	94.5	12	83	16524.4	16347.5	11	64	9	21	10	5	20	4	21	21	33270.3	33333	62.7	0.19
Dreamland	24	5	17	100	16553.9	16397.9	18	36	29	5	24	72	40.5	43	20	20	33405.3	33333	-72.3	-0.22
First Person Shooter	4	20	25	61.5	16515.4	16347.5	14	41	24	23	4	157.5	20	40	19	20	33335.8	33333	-2.8	-0.01
Teso Dos Bichos	30	23	13	63.5	16519.4	16347.5	10	43	28	8	42.5	157.5	20	32	18	18	33373.3	33333	-40.3	-0.12
Darkness Falls	12	96.5	11	32.5	16543.9	16388.9	23	38	18	16	17	71	20	7	17	17	33328.8	33333	4.2	0.01
Lazarus	11	15	10	34.5	16520.4	16385.9	25	44	13	12	41.5	157.5	20	28	16	16	33349.8	33333	-16.8	-0.05
Leonard Betts	19	13	22	2	16518.4	16347.5	24	32	22	11	15	157.5	20	46	15	16	33280.3	33333	52.7	0.16
Invocation	15	18	7	26.5	16517.4	16347.5	27	37	19	14	33	157.5	20	37	14	16	33309.8	33333	23.2	0.07
Anaxazi	16	12	14	64.5	16517.4	16347.5	21	63	9	4	14	157.5	38.5	46	13	13	33350.3	33333	-17.3	-0.05
Essence	13	24	8	30.5	16524.4	16387.9	17	35	12	25	32	157.5	20	42	12	12	33352.3	33333	-19.3	-0.06
3	17	11	3	95	16512.4	16389.9	6	40	9	17	11	67	20	34	11	11	33254.3	33333	78.7	0.24
My Struggle IV	2	4	23	27.5	16511.4	16347.5	13	26	12	42	8	157.5	20	30	10	10	33243.8	33333	89.2	0.27
Babylon	6	9	20	96	16550.9	16396.9	8	27	21	24	9	157.5	20	5	9	9	33368.3	33333	-35.3	-0.11
Trust no1	8	15	19	28.5	16525.4	16347.5	20	25	9	27	40.5	157.5	20	35	8	8	33293.3	33333	39.7	0.12
Space	9	16	15	60.5	16516.4	16347.5	7	42	9	9	7	69	20	187.5	7	7	33328.8	33333	4.2	0.01
Fight Club	10	3	9	97	16514.4	16415.9	6	29	18	22	3	157.5	20	36	6	6	33352.8	33333	-19.8	-0.06
My Struggle III	5	6	2	26.5	16513.4	16347.5	19	24	9	10	87	157.5	20	2	5	6	33239.8	33333	93.2	0.28
El Mundo Gira	26	7	17	63.5	16522.4	16347.5	9	34	14	13	3	157.5	20	38	4	4	33279.8	33333	53.2	0.16
Rm9sbG93ZXJz	1	93.5	7	94	16544.9	16387.9	3	2	26	3	5	2	20	31	3	3	33226.3	33333	106.7	0.32
Nothing Lasts Forever	0	3	4	35.5	16546.9	16390.9	3	31	12	19	32	157.5	20	30	2	2	33288.8	33333	44.2	0.13
Familiar	3	17	9	27.5	16544.9	16347.5	1	32	15	1	40.5	157.5	20	28	1	1	33245.8	33333	87.2	0.26
Naider and Scully Meet the Were-Monster	7	0	7	100	16554.9	16415.9	4	1	8	6	2	157.5	20	4	0	1	33288.3	33333	44.7	0.13

16. ábra: Az antidiszkriminatív COCO Y0-ás hasonlóságelemző modell számításai alapján nem áll fent azonosság az objektumok között, az eltérések abszolút értékei a delta/tény oszlopban (forrás: <https://miau.my-x.hu/myx-free/coco/engine3.php>)

Amennyiben nem, úgy a mintázattól vagy akár a referenciafilmtől eltérő renitens elemek szűrhetővé válnak arányaiban és abszolútértékben is. (Vö 16. ábra).

4.3.4 Út az optimalizálás felé (származtatott adatok bevonásával)

A módszer során származtatott adatok (pl. vicces/szomorú, ijesztő/szomorú, vicces/ijesztő) injektálása történt az Objektum-Attribútum Mátrixba. Az osztás nullával való elkerülése érdekében az attribútumokat egy plusz 1 másodperces sablonnal lettek ellátva, amely hatékonyan küszöbölte ki a problémát. Az így elvégzett szimulációs kísérlet ugyan a genetikai potenciál tekintetében nem hozott javulást, azonban minden korábbinál pontosabb becslési eredményeket biztosított (Vö 13.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	type	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	y
2	direction	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	n/a
3	correlation	0.47	0.32	-0.22	0.45	0.29	0.23	0.00	-0.40	0.08	0.38	0.06	0.38	0.05	0.39	-0.07	0.33	0.10	0.44	0.59	1.00
4	unit of measurement	million	second	second	second	second	second	second	second	second	second	second	second	second	second	second	second	second	second	second	1/10 score
5	Episode name	viewership	Scary	Scary/Sad	Sad	Sad/funny	Funny	Funny/sary	Action	Character introduction humorously	Surprising moments	Detectives scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value?	Naive estimation based on the Q column	Reality
71	The Truth	17	10	5	2	6	18	19	30	15	1	23	8	1	2	1	3	1	1	1	8200
72	Detour	4	1	19	4	20	5	13	31	8	28	5	23	13	20	29	1	17	2	1	8100
73	Redux	2	15	9	7	7	23	24	9	17	2	4	1	2	19	1	12	10	3	3	8600
74	Redux II	3	9	3	1	5	17	18	1	17	5	29	4	5	4	1	6	3	4	4	9000
75	Clyde Bruckman's Final Repose	13	2	23	13	26	2	8	4	5	19	10	11	11	16	27	9	9	5	4	9200
76	Paper Clip	9	22	7	10	4	24	23	4	17	15	1	15	24	3	1	6	24	6	4	9000
77	Dreamland II	11	24	2	3	17	6	5	7	9	9	21	6	16	6	25	5	6	7	7	8900
78	Bad Blood	6	4	30	28	31	1	4	11	6	16	17	23	26	26	30	9	30	8	8	9200
79	Missing a Cigarette Smoking Man	10	12	13	5	8	20	22	8	7	3	27	17	29	1	23	8	26	9	9	9000
80	Ghoul	8	5	22	19	21	11	16	16	17	10	1	23	10	22	28	12	29	10	10	5900
81	Dreamland	7	17	6	14	15	4	3	3	3	13	14	1	27	14	11	1	7	12	11	8600
82	First Person Shooter	27	11	14	6	12	15	17	25	17	17	9	7	8	28	1	12	10	13	11	7900
83	Tesla Dies Blokes	1	8	20	18	18	13	14	21	17	21	7	3	23	7	1	12	18	14	13	8300
84	Darkness Falls	19	3	25	20	11	22	27	14	12	8	12	13	15	15	22	12	25	15	14	8400
85	Laarus	20	17	21	21	14	20	21	20	15	6	6	19	19	8	1	12	23	16	15	7000
86	Leonard Betts	12	19	8	9	1	31	31	22	17	7	19	9	20	17	1	12	4	11	15	9100
87	Invocation	16	13	26	26	10	29	30	19	17	4	13	12	17	11	1	12	13	17	15	7500
88	Anasazi	15	20	16	17	19	12	10	23	17	10	3	23	28	18	1	4	4	18	18	6300
89	Essence	18	7	28	25	13	24	28	16	13	14	15	20	6	12	1	12	8	19	19	8300
90	3	14	21	27	30	29	9	7	28	11	25	10	23	14	21	26	12	16	20	20	6100
91	MyStruggle IV	29	28	1	8	2	27	20	29	17	18	15	20	3	24	1	12	20	21	21	7200
92	Babylon	25	23	10	11	23	8	6	6	4	23	24	10	7	23	1	12	28	22	22	6400
93	Trust no1	23	17	15	12	3	26	26	15	17	11	26	23	4	10	1	12	15	23	23	7400
94	Space	12	16	17	16	15	16	15	24	17	24	8	23	22	25	24	12	2	24	24	6200
95	Fight Club	11	29	12	23	28	7	2	26	2	25	22	13	9	20	1	12	14	25	25	5900
96	MyStruggle III	26	26	31	31	24	29	25	27	17	12	27	23	21	5	1	12	31	26	25	6400
97	El Mundo Gira	5	25	11	14	16	13	9	18	17	22	16	18	18	30	1	12	12	27	27	8000
98	RmBxG9320a	30	6	29	26	27	10	12	13	13	29	31	5	29	27	31	12	19	28	28	6700
99	Nothing Lasts Forever	31	29	18	29	22	19	11	10	10	29	20	20	12	12	1	12	20	29	29	6600
100	Familiar	28	14	24	22	9	27	29	12	17	31	18	16	31	9	1	12	22	30	30	8000
101	Usher and Sully Meet the Were-Mom!	24	31	4	24	30	3	1	1	1	27	30	23	25	29	1	11	27	31	30	8000
102	Err	0	323	0	0	1145	3144	0	299	0	0	646	254	0	1443	0	0	0	0	0	7254

és 17. ábra)

17. Ábra: A Származtatott adatokkal és +1 másodperces idősablonnal ellátott és sárgával jelölt predikciós kísérlet javuló IMDb index eredménye *1000 (forrás: saját ábrázolás \IMDb_szimulator.xlsx, +1sec +% +Y0 munkalap A-U 71-102 tartomány)

4.3.5 Út az optimalizálás felé (Származtatott adatok bevonásával és a renitens objektumok ignorálásával)

A korábban ismertetett (4.3.3 és 4.3.4) módszerek alkalmazásával egy koordináta-rendszer alapú megközelítés vált lehetővé, amelyben az objektumok referenciafilmhez viszonyított arányos eltérései lekérdezhetővé váltak. Ezen adatok alapján a leginkább és legkevésbé hasonló objektumok szisztematikus eltávolítása került fókuszba. A kísérletek eredményei szerint a legkevésbé hasonló objektum likvidálása jelentős mértékű romlást okozott a becslési pontosságban az alapállapothoz képest. Hasonló negatív hatás volt megfigyelhető a referenciafilmhez leginkább hasonló objektum eltávolításakor is. Mindkét objektum együttes eltávolítása esetén a becslési pontosság további romlását sikerült azonosítani. A három leginkább hasonló objektum likvidálása további romlást eredményezett, amely tendencia négy, illetve öt objektum likvidálása esetén is fennmaradt. Ugyanakkor a három legeltérőbb film eltávolítása fordított hatást váltott ki: a szimulátor a valós IMDB értékhez képest mindössze 0,5 pontos eltéréssel, az eddigi legpontosabb becslést nyújtotta. Ez az eredmény azonban nem mutatott további javulást több eltérő objektum likvidálása esetén. Az optimális becslési pontosság tehát e módszertan alkalmazása során a három legkevésbé ha-

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	type	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	y
2	direction	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	n/s
3	correlation	0.47	0.32	-0.22	0.45	0.29	0.23	0.00	-0.40	0.08	0.38	0.06	0.38	0.05	0.39	-0.07	0.33	0.10	0.44	0.59	1.00
4	unit of measurement	million	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	sec	1/10 score	1/10 score
5	Episode name	viewership	Scary	Scary/Sad	Sad	Sad/funny	Funny	Funny/scary	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation	Reality*1000
71	Redux	2	13	7	4	6	20	21	8	14	1	4	1	1	17	1	9	8	1	1	8600
72	Redux II	3	8	2	1	5	15	16	1	14	4	26	4	4	3	1	3	2	2	2	9000
73	Clyde Bruckman's Final Repose	11	1	20	10	23	2	7	4	5	17	9	9	10	14	25	6	7	3	2	9200
74	Paper Clip	8	20	5	7	4	21	20	4	14	13	1	13	21	2	1	3	21	4	2	9000
75	Bad blood	5	3	27	25	28	1	4	10	6	14	16	21	23	23	27	6	27	5	5	9200
76	Musings of a Cigarette Smoking Man	9	10	11	2	7	17	19	7	7	2	24	15	26	1	22	5	23	6	6	9000
77	Ghoul	7	4	19	16	18	9	14	15	14	18	2	21	9	19	26	9	26	7	7	5900
78	Dreamland	6	24	4	11	22	4	3	3	3	11	13	2	24	12	20	1	5	9	8	8800
79	First Person Shooter	24	9	12	3	11	13	15	24	14	15	8	6	7	25	1	9	8	10	8	7900
80	Teso Dos Bichos	1	7	17	15	16	11	12	20	14	19	6	3	20	5	1	9	15	11	10	8300
81	Darkness Falls	16	2	22	17	10	19	24	13	10	7	11	11	13	13	21	9	22	12	11	8400
82	Lazarus	17	15	18	18	13	17	18	19	13	5	5	17	16	6	1	9	20	13	12	7000
83	Leonard Betts	10	17	6	6	1	28	28	21	14	6	18	7	17	15	1	9	3	8	12	9100
84	Invocation	14	11	23	23	9	26	27	18	14	3	12	10	14	9	1	9	11	14	12	7600
85	Anasazi	13	18	14	14	17	10	9	22	14	8	3	21	25	16	1	2	3	15	15	6900
86	Essence	15	6	25	22	12	21	25	15	11	12	14	18	5	10	1	9	6	16	16	8300
87	3	12	19	24	27	26	7	6	27	9	23	9	21	12	18	24	9	14	17	17	6100
88	My Struggle IV	26	25	1	5	2	24	17	28	14	16	22	18	2	21	1	9	17	18	18	7200
89	Babylon	22	21	8	8	20	6	5	6	4	21	21	8	6	20	1	9	25	19	19	6400
90	Trust no1	20	15	13	9	3	23	23	14	14	9	23	21	3	8	1	9	13	20	20	7400
91	Space	19	14	15	13	14	14	13	23	14	22	7	21	19	22	23	9	1	21	21	6200
92	Fight Club	18	26	10	20	25	5	2	25	2	23	20	11	8	27	1	9	12	22	22	5900
93	My Struggle III	23	23	28	28	21	26	22	26	14	10	24	21	18	4	1	9	28	23	22	6400
94	El Mundo Gira	4	22	9	11	15	11	8	17	14	20	15	16	15	27	1	9	10	24	24	8000
95	Rm9sbG93Z0Uz	27	5	26	23	24	8	11	12	11	26	28	5	26	24	28	9	16	25	25	6700
96	Nothing Lasts Forever	28	26	16	26	19	16	10	9	8	26	19	18	11	10	1	9	17	26	26	6600
97	Familiar	25	12	21	19	8	24	26	11	14	28	17	14	28	7	1	9	19	27	27	8000
98	Wilder and Scully Meet the Were-Mon	21	28	3	21	27	3	1	1	1	25	27	21	22	26	1	8	24	28	27	8000
99	Red's kidnapping	0	264	0	0	1569	1015	0	0	0	0	949	142	0	664	0	0	0	0	0	7500

sonló objektum eltávolításával érhető el. (Vö. 18. ábra)

18. ábra: A szimulátor legpontosabb becslése (U 99-es pirossal jelölt cella). A referenciafilm 8-as értékelését az attribútumok birtokában 7,5-re (*1000) becsülte (forrás: saját ábrázolás \IMDb_szimulator.xlsx, Piros renitek nélkül munkalap A-U 71-99 tartomány)

4.4 Becslés eredményessége A szimulátor által generált becslések az alkalmazott módosításokat követően ígéretes pontosságot mutattak, azonban a tanulási modell önellenőrzési folyamatának eredményei túltanulás jelenlétére hívták fel a figyelmet. A túltanulás jelensége különösen akkor vált egyértelművé, amikor a referenciafilm Y-értékét szándékosan radikálisan módosítottuk. (Vö. 19. ábra)

COCO:STD	viewership	Scary	Sad	Funny	Action	Character introduction humorously	Surprising moments	Detective scenes	Interference	Character introduction	Introduction of antagonist	Scary but funny	Sad but funny	Total time	D-O Value	Naive estimation	Becslés	Tény=0	Delta	Delta/Tény	
The Truth	0	235.5	0	64	959	1717.5	200	0	1986	0	352.5	600	0	1603	178	0	244.5	0	0	8200.1	8200
Detour	555.5	6794.1	0	64	87	2385.5	272.5	0	2376.5	0	1111	0	0	0	0	6353.6	0	0	0	19999.8	5000
Redux	555.5	276	0	64	959	1717.5	200	53	1986	0	1111	1256	0	0	178	0	244.5	0	0	8600.6	8600
Redux II	555.5	295.5	0	64	959	1717.5	200	53	1986	0	0	1145	0	1603	178	0	244.5	0	0	9001.1	9000
Clyde Bruckman's Final Repose	0	1532	0	64	0	2544	695	53	2376.5	0	1111	580	0	0	0	0	244.5	0	0	9200.1	9200
Paper Clip	0	276	0	64	959	1717.5	200	53	1986	0	1363.5	580	0	1603	178	0	0	0	0	9000.1	9000
Dreamland II	0	276	458	64	100	1717.5	695	53	2300	0	786	1145	0	1061.5	0	0	244.5	0	0	8900.6	8900
Bad Blood	186	798	0	0	0	3980.1	695	53	2376.5	0	1111	0	0	0	0	0	0	0	0	9199.6	9200
Maniacs of a Cigarette Smoking Man	0	276	0	64	959	1717.5	200	53	2376.5	0	352.5	580	0	2421.5	0	0	0	0	0	9000.1	9000
Ghoul!	0	798	0	0	87	1717.5	200	0	1986	0	1111	0	0	0	0	0	0	0	0	5899.6	5900
Dreamland	0	0	0	64	0	2385.5	695	53	2376.5	0	1111	1145	0	0	0	726	244.5	0	0	8800.6	8800
First Person Shooter	0	295.5	0	64	959	1717.5	200	0	1986	0	1111	1145	0	0	178	0	244.5	0	0	7900.6	7900
Tesé Dos Bichos	555.5	295.5	0	0	100	1717.5	272.5	0	1986	0	1111	1145	0	939	178	0	0	0	0	8300.1	8300
Darkness Falls	0	1532	0	0	959	1717.5	200	0	2300	0	1111	580	0	0	0	0	0	0	0	8399.6	8400
Lazarus	0	276	0	0	592	1717.5	200	0	1986	0	1111	0	0	939	178	0	0	0	0	6999.6	7000
Leonard Betts	0	276	0	64	4965.6	0	0	0	1986	0	786	600	0	0	178	0	244.5	0	0	9100.1	9100
Invocation	0	276	0	0	959	1717.5	0	0	1986	0	1111	580	0	792	178	0	0	0	0	7599.6	7600
Anasazi	0	276	0	0	100	1717.5	686.5	0	1986	0	1111	0	0	0	178	0	244.5	0	0	6299.6	6300
Essence	0	798	0	0	959	1717.5	200	0	2300	0	1111	0	0	792	178	0	244.5	0	0	8300.1	8300
3	0	276	0	0	0	1717.5	695	0	2300	0	1111	0	0	0	0	0	0	0	0	6099.6	6100
My Struggle IV	0	0	1015.5	64	1686.5	1717.5	200	0	1986	0	352.5	0	0	0	178	0	0	0	0	7200.1	7200
Babylon	0	276	0	64	87	1717.5	695	53	2376.5	0	352.5	600	0	0	178	0	0	0	0	6399.6	6400
Trust no1	0	276	0	64	1686.5	1717.5	200	0	1986	0	352.5	0	0	939	178	0	0	0	0	7399.6	7400
Space	0	276	0	0	592	1717.5	272.5	0	1986	0	1111	0	0	0	0	0	244.5	0	0	6199.6	6200
Fight Club	0	0	0	0	0	1717.5	695	0	2376.5	0	352.5	580	0	0	178	0	0	0	0	5899.6	5900
My Struggle III	0	276	0	0	87	1717.5	200	0	1986	0	352.5	0	0	1603	178	0	0	0	0	6600.1	6600
El Mundo Gira	555.5	276	0	64	592	1717.5	695	0	1986	0	1111	580	0	0	178	0	244.5	0	0	7999.6	8000
Amfibi9323232	0	798	0	0	0	1717.5	686.5	53	2300	0	0	1145	0	0	0	0	0	0	0	6700.1	6700
Nothing Lasts Forever	0	0	0	0	87	1717.5	686.5	53	2300	0	786	0	0	792	178	0	0	0	0	6600.1	6600
Familiar	0	276	0	0	959	1717.5	200	53	1986	0	1111	580	0	939	178	0	0	0	0	7999.6	8000
Wulder and Scully meet the Were-Hunter	0	0	0	0	0	2385.5	3007.5	53	2376.5	0	0	0	0	0	178	0	0	0	0	8000.6	8000

19. ábra: A Detour mint referenciafilm IMDB indexét szándékosan 20 pontra (*1000) módosítottuk, azonban a szakértői rendszer nem azonosította hibaként

Az elvárásokkal ellentétben a modell nem azonosította hibaként a referenciafilm megváltoztatott állapotát, hanem azt strukturálisan összhangban lévőnek ítélte a többi objektummal, ami arra utal, hogy a referenciafilm nem volt kellően megbízható az ilyen eltérések detektálására.

A probléma orvoslása érdekében kísérletet tettünk új referenciafilm(ek) kiválasztására. Az újonnan bevont referenciafilmek ugyan nem tekinthetők teljesen hibátlan etalonnak, azonban jelentősen megbízhatóbbnak bizonyultak az előzőleg alkalmazott filmnél. (Familiar) Ezáltal a becslési modell érzékenysége és pontossága javult, és a referenciaértékek módosításaira adott reakciók jobban tükrözték a modell valós struktúraelemzési képességeit.

A referenciafilmek cseréje nemcsak a túltanulás mérséklésében hozott előrelépést, hanem rámutatott arra is, hogy a referenciafilm(ek) kiválasztása alapvető fontosságú a becslési modell hatékonyságának és általánosíthatóságának bizonyítása érdekében.

5. Vita

Állítás: Noha a szimulátor célja az objektívítás alapú döntéstámogatás, a filmek attribútumainak rögzítése manuálisan, szubjektív módon történt, ami megkérdőjelezheti a becslési modell eredményeinek megbízhatóságát.	Válasz: Elismerem, hogy a filmek attribútumainak rögzítése során bizonyos szubjektívítás elkerülhetetlen volt, még akkor is, ha az adatok feldolgozását a lehető legnagyobb körültekintéssel és objektivitásra való törekvéssel végeztem. Azonban ez a korlát nem csupán a jelenlegi kutatás sajátossága, hanem az emberi értékelésből fakadó természetes jelenség, amely szinte minden hasonló jellegű tanulmányban fennáll. A jövőben azonban a technológiai fejlődés ígéretes lehetőségeket kínál ezen kihívás áthidalására. Olyan mesterséges intelligencia rendszerek és robotok fejlesztése zajlik, amelyek képesek lesznek automatikusan dekódolni a filmekben megjelenő attribútumokat – például az érzelmi tónusokat, narratív szerkezeteket vagy vizuális stílusjegyeket –, ezáltal a manuális adatrögzítésből eredő szubjektívítást minimálisra csökkenteni. Az ilyen rendszerek várhatóan képesek lesznek kvantifikálni és objektíven elemezni a vásznon megjelenő érzelmeket, interakciókat és vizuális hatásokat. Ezen fejlesztések előrehaladtával reális lehetőségként merül fel, hogy az attribútumok teljes mértékben automatizált és objektív úton kerüljenek rögzítésre. Ennek köszönhetően a hasonló szimulátorok nemcsak pontosabbá,
---	--

	hanem megbízhatóbbá is válhatnak, hozzájárulva a filmes sikeresség előrejelzésének tudományos alapokon nyugvó fejlődéséhez. Ez a kutatási irány nemcsak az adatfeldolgozás terén, hanem az emberi szubjektivitás kiküszöbölésében is áttörést jelenthet.
--	---

Állítás 2: Az attribútumok között nem szerepel olyan mutató, amely a filmek látványvilágára összpontosulna. Ennek következtében a szimulátor nem méri a vizuális szépséget, és így nem képes nagy pontossággal előrejelezni egy látványfilm várható fogadtatását.	Válasz: Valóban igaz, hogy a szimulátor jelenlegi verziójában nem szerepel kifejezetten a látványvilágot célzó attribútum. Ennek oka, hogy az emberi szem által érzékelt vizuális szépség mérésére olyan objektív, kvantifikálható mutatókat létrehozni, amelyek széleskörűen alkalmazhatók és elfogadottak, jelenleg rendkívül nehéz. A szépség szubjektív természetéből adódóan a különböző közönségrétegek eltérően ítélik meg a látvány minőségét és hatását, ami jelentős kihívást jelent a modellezés szempontjából. Azonban a technológia fejlődése ígéretes lehetőségeket kínál ezen probléma kezelésére. A modern képfelismerő algoritmusok és mesterséges intelligencia rendszerek már képesek bizonyos esztétikai paraméterek, például színhasználat, kompozíció vagy vizuális harmónia elemzésére. Ezek a rendszerek a jövőben még tovább fejlődhetnek, lehetővé téve, hogy a vizuális szépség objektív módon is mérhetővé váljon.
---	---

	Amint ezek a technológiák elérik azt a szintet, hogy megbízhatóan képesek legyenek a filmek látványvilágának kvantitatív elemzésére, ezek az eszközök integrálhatók lesznek a szimulátorba. Így a szimulátor egy újabb modulal bővíthet, amely a látványvilág hatását is figyelembe veszi a film sikerességének előrejelzésekor. Ez a fejlődés nemcsak a látványfilmek fogadtatásának pontosabb előrejelzését tenné lehetővé, hanem hozzájárulna a szimulátor általános pontosságának további növeléséhez is.
--	--

Állítás 3: Még ha az IMDb szimulátor képes is nagy pontossággal előrejelezni egy közeljövőben készülő film fogadtatását, a nézők preferenciái és a piaci trendek az idő múlásával változnak. Ennek következtében a szimulátor elavulttá válhat, hiszen csak a jelenlegi trendek mintázatait képes azonosítani.	Válasz: A szimulátor tervezése során egyik fő szempont volt annak rugalmassága és adaptálhatósága, így az objektumok – azaz a vizsgált filmek és azok attribútumai – könnyedén bővíthetők vagy akár teljesen lecserélhetők egy esetleges piaci trendváltozás esetén. Ezáltal a szimulátor nem csupán a jelenlegi trendekre korlátozódik, hanem képes dinamikusan alkalmazkodni a nézői preferenciák és piaci igények átalakulásához. Ezen túlmenően a technológiai fejlődés jelentős előrelépéseket kínál a szimulátor hosszú távú relevanciájának biztosításához. Az attribútumok automatikus rögzítésének és elemzésének technológiai lehetőségei folyamatosan bővülnek, például a mesterséges intelligencia alapú képfelismerés vagy szövegelemzés ré-
--	--

	vén. Ezek az új eszközök lehetővé teszik, hogy a szimulátor automatikusan adaptálódjon az új objektumokhoz és azok attribútumaihoz, minimalizálva az emberi beavatkozás szükségességét. Mivel a számítási módszertan nem a konkrét adatokra, hanem azok struktúrájára és kapcsolati mintázataira épül, az új objektumok integrálása esetén is megbízható eredményeket nyújt. Ennek köszönhetően a szimulátor képes lesz tartósan fenntartani relevanciáját, és elavulásának veszélye gyakorlatilag kizárható. Így a szimulátor hosszú távon is hatékony döntéstámogató eszközként szolgálhat a filmipar számára.
--	--

Állítás 4: A szimulátor ígéretes eredményei ellenére a gyakorlati alkalmazás további validációt igényel. Különösen fontos lenne a szimulátor ipari környezetben történő tesztelése, amely során valódi filmprojektek sikerességét próbálnák előre jelezni. Ez azonban költséges és kockázatos folyamat lehet a befektetők számára.	Válasz: Az állítás valós problémát vet fel, hiszen egy ilyen szimulátor gyakorlati alkalmazhatóságának ipari környezetben történő validációja kétségtelenül jelentős pénzbeli és reputációs kockázatot hordozhat magában. Azonban léteznek olyan alternatív megközelítések, amelyek lehetővé tennék a szimulátor biztonságos és alacsony költségű tesztelését. Az egyik ilyen lehetőség, az amatőr filmesek bevonása, akik kisebb költségvetésből és alacsonyabb kockázattal dolgoznak, a szimulátort használhatnák filmjeik sikerpotenciáljának vizsgálatára. Az ilyen kísérletek nemcsak
--	---

	értékes gyakorlati visszajelzést nyújthatnának a szimulátor teljesítményéről, hanem lehetőséget kínálnának a technológia szélesebb körű elterjesztésére is. Egy másik, még előremutatóbb megközelítés a mesterséges intelligencia bevonása a tesztelési folyamatba. Elképzelhető például, hogy egy MI-alapú <i>robotíró</i> a szimulátor által generált eredmények alapján létrehoz egy forgatókönyvet, amely szimulált filmként szolgálhat a tesztelés során. Egy másik MI-rendszer pedig vizualizálná ezt a forgatókönyvet, hiszen már ma is léteznek olyan mesterséges intelligencia alapú programok, amelyek képesek néhány utasítás avagy <i>prompt</i> alapján vizuális tartalmat létrehozni. Ez az eljárás nemcsak költséghatékony és kockázatmentes lenne, hanem lehetőséget nyújtana a szimulátor tesztelésére a valósághoz közeli körülmények között. Ezzel a megközelítéssel a szimulátor validációja elvégezhető lenne anélkül, hogy az ipari szereplők jelentős pénzügyi vagy reputációs kockázatot vállalnának, miközben az MI-alapú megoldások hozzájárulnának a szimulátor technológiai adaptációjához és folyamatos fejlesztéséhez.
--	---

6.Következtetések és jövőkép

A szimulátor eredményei ígéretesek, különösen a becslési pontosság terén, azonban a referenciafilmek és az attribútumok kiválasztásának jelentős hatása van a szimulátor teljesítményére. Ez rámutatott arra, hogy a szimulátor optimalizálása során a bemeneti jelek precíz definiálása alapvető fontosságú.

- A túltanulás jelensége, amelyet a referenciafilmek hibás kezelése eredményezett, kiemeli a modell robusztusságának fontosságát. Az új referenciafilmek kiválasztása részben megoldotta ezt a problémát, de további kutatások szükségesek az algoritmus általánosíthatóságának javítása érdekében.

- A származtatott attribútumok bevonása hozzájárult a becslések pontosságának növeléséhez. Ez arra utal, hogy a komplex, többdimenziós jellemzők integrálása kulcsfontosságú a filmes sikeresség előrejelzésében.

- A renitens objektumok kiszűrése és azok hatásainak vizsgálata megerősítette, hogy a szélsőséges értékek jelentős torzító hatással vannak a modell teljesítményére. A módszertan igazolta, hogy az ilyen objektumok kezelése javíthatja a szimulátor megbízhatóságát.

- A jelenlegi vizsgálat még nem használta ki a limitációs fiktív adatok lehetőségét.

- Dupla attribútumos készletek bár készültek, de terjedelmi okokból nem kerültek bele a dokumentumba.

7.Összefoglalás

A dolgozat során bemutatásra került a kutatásom előrehaladása, amely számos matematikai, filozófiai és technológiai kihívás leküzdésén keresztül valósult meg. Az olvasó betekintést nyerhetett abba, hogy milyen célokat tűztem ki, miért nem sikerült ezeket teljes mértékben megvalósítani, valamint, hogy a jövőben milyen új megközelítések és megoldások válhatnak elérhetővé. A dolgozat számos különböző megközelítést tárgyal, amelyek mindegyike önálló szakdolgozati témát is képezhetne. A terjedelem és egyéb korlátok miatt ezen megközelítések részletes kifejtése nem szerepelhetett a dolgozatban, azonban a csatolt Excel-fájl számos, a dolgozatban nem teljesen kifejtett kísérletet tartalmaz, amelyeket a téma iránt továbbra is érdeklődő olvasók számára elérhetővé teszek.

8. Mellékletek

..\IMDb_szimulator.xlsx

9. Irodalomjegyzék

(https://miau.my-x.hu/miau2009/index_tki.php3?_filterText0=*knuth) -2024

Attribútum dokumentációs példa: (<https://www.youtube.com/watch?v=QSLm-IqB0oE>) -2024

COCO: (<https://miau.my-x.hu/myx-free/coco/index.html>) -2024

Donald E. Knuth 2001: Things a Computer Scientist Rarely Talks About
(<https://zeta.forum4.org/txt/1575863278.pdf>) -2024

Epizódok rangsora: (<https://www.imdb.com/title/tt0106179/episodes/?season=1>) -2024

Hasonlóságelemzés: Dr. Pitlik László, 2009 <https://miau.my-x.hu/mediawiki/index.php/Hasonl%C3%B3s%C3%A1gelemz%C3%A9s#Defin.C3.ADci.C3.B3-2024>

<https://www.kaggle.com/code/saurav9786/imdb-score-prediction-for-movies>) -2024

https://www.youtube.com/results?search_query=dalfut%C3%A1r-2024

IMDb score prediction: Victory Onomaku 2023
(https://medium.com/@Onumaku_chibuike/imdb-movie-ratings-prediction-with-machine-learning-7bdaf843c268; -2024

Nézettségi adatok: (https://x-files.fandom.com/wiki/The_X_Files_Ratings) -2024

Standard COCO felhasználói felület: (https://miau.my-x.hu/myx-free/coco/beker_std.php) -2024

X-akták: (<https://hu.wikipedia.org/wiki/X-akt%C3%A1k>) -2024

10. Rövidítésjegyzék

CBIR: Content-Based Image Retrieval

OAM: Objektum-Attribútum Mátrix

11. Ábra jegyzék

- 4. Ábra: Az IMDB pontszámok és a filmek profitja közötti kapcsolat 6. oldal
- 5. Ábra: Filmek IMDb pontszáma régiós bontásban 7. oldal
- 3. Ábra: Az USA és egyéb országok lakóinak szavazatát összehasonlító mutató 7. oldal
- 4. Ábra: A facebook like-ok és az IMDb pontszám közt növekvő kapcsolat figyelhető meg 8. oldal
- 5. Ábra: A kiválasztott színészek és a filmek IMDB pontszám kapcsolatának mutatója 8. oldal
- 6. Ábra: Az ábra a vizsgált attribútumok közötti korrelációs együtthatókat mutatja. 9. oldal
- 7. Ábra: Egyszerűsített folyamatábra, mely a dolgozat keretein belüli lépéseket kívánja beutatni. 13. oldal
- 8. ábra: A hőtésképpel ellátott és az irányfigyelembevételével sorbarendezt OAM mátrix 14. oldal
- 9. ábra: Az ábrán a rangsorolást követő állapot látható 15.sor
- 10. ábra: A Coco standard modell kezelőfelületén a felhasználó betáplálja a sorszámozott értékeket 16. oldal
- 11. ábra: Coco standard modell futtatást követő állapota 17. oldal
- 12. ábra: Az adathiány generálást követően megbecsülni kívánt +1.-ik objektum berögzítése 18. oldal
- 13. ábra: A 47. sorban megtekinthetők a szakértői rendszer által megbecsült +1. objektum 19. oldal
- 14. Ábra: A megnövelt objektumszámú adatvagyron romló tanulási modellje 20. oldal
- 15. ábra: A, megnövelt objektumszámmal megbecsült referencia epizód 21. oldal
- 16. ábra: Az antidiszkriminatív COCO Y0-ás hasonlóságelemző modell számításai 21. oldal
- 17. Ábra: A Származtatott adatokkal és +1 másodperces idősablonnal ellátott és sárgával jelölt predikciós kísérlet 22. oldal
- 18. ábra: A szimulátor legpontosabb becslése 23. oldal
- 19. ábra: A Detour mint referenciafilm IMDB indexét szándékosan 20 pontra (*1000) módosítottuk 24. oldal

12. Definíciós jegyzék

Algoritmus: „Az algoritmusok fontos részei az informatikának és mindennapi életünknek is. Nem csak a számítástechnikában, hanem napi rutinjaink elvégzésében is algoritmusokat hajtunk végre. Az algoritmus nem más, mint jól meghatározott lépések sorozata egy bizonyos cél elérésének a céljából.” (Adonyi Róbert 2011, Adatstruktúrák és algoritmusok)

Forráskód: „A forráskód egy szöveges dokumentum, amely valamelyik programozási vagy leíró nyelv utasításait tartalmazza.” (Dávid Ádám 2017, Kiszervezett Marketing)

Szimulátor: „Egy szoftver vagy eszköz, amely egy valóságos rendszert, folyamatot vagy környezetet modellez és szimulál. Lehetővé teszi a felhasználók számára, hogy virtuálisan kipróbálják és tanulmányozzák a valós világban történő eseményeket anélkül, hogy közvetlenül beavatkoznának vagy kockázatot vállalnának.” (Jerry Banks 2005, Handbook of Simulation: Principles Methodology, Advances, Applications, and Practice)

OAM: „Egy strukturális modell, amelyben az objektumok és azok attribútumai közötti kapcsolatokat tároljuk. A mátrix általában egy táblázatos formában van megvalósítva, ahol az oszlopok az objektumokat, míg a sorok az attribútumokat reprezentálják.” (Berente, N., Lyytinen, K., Yoo, Y., & King, J. L. 2016, Routines as a source of change in organizational schemata: The role of trial and error learning.)

Objektum: „Az objektum olyan entitás, amely rendelkezik bizonyos tulajdonságokkal vagy attribútumokkal.” (ChatGPT- 3.5)

Attribútum: „Az attribútumok olyan tulajdonságokat foglalnak magukba, amelyekkel egy objektum felruházható.” (ChatGPT- 3.5)

Induktív tanulás: „Az induktív tanulás egy olyan megközelítés a gépi tanulásban, amely során az algoritmus az adatokból próbál általános szabályokat kinyerni. Célja, hogy az AI rendszerek képesek legyenek tanulni az adathalmazból, és ezek alapján tegyenek előrejelzéseket.” (Mitchell, T. M. 1997, Machine Learning. McGraw Hill)

Hasonlóságelemzés: „A hasonlóságelemzés lényege a lépcsős függvény, mely az összehasonlítható objektumok összehasonlítást lehetővé tevő attribútumainak attribútumokénti rangsorszá-

mait optimalizálás keretében olyan csereértékekkel tölti fel, mely csereértékek aggregációja a lépcsős függvény által kialakított modell célját a lehető legjobban közelíti a megadott hibadefiníció keretében.” (Dr. Pitlik László 2009, miau.my-x.hu)

Döntéstámogatás: *„A döntéstámogató rendszer olyan számítógépre alapozott rendszer, mely adatok keresésével, rendszerezésével, mesterséges intelligencia használatával segíti a döntések meghozatalát, de nem képes önálló döntések meghozatalára, a döntéshozatal továbbra is emberi feladat marad.” (Druzdzel, M. J. and R. R. Flynn (1999). Decision Support Systems. Encyclopedia of Library and Information Science. A. Kent, Marcel Dekker, Inc.)*

Genetikai potenciál: *Megmutatja mekkora lehet az adott rendszerben rejlő maximálisan elérhető jószágpontok mértéke a megadott keretfeltételek (X1-X10 stb.) együttállása esetén. (saját értelmezés)*

Robotizáció: *„A robotizált folyamatautomatizálás olyan szoftverekre utal, amelyek mesterséges intelligenciával vagy gépi tanulási képességekkel rendelkeznek, és képesek úgy elvégezni egy cél-faladatot, akár egy ember”. (March 9, 2021 Universal Robots)*

COCO STD: *„A COCO-STD modelljében rejlő lehetőség, hogy valós Y-változót (pl. árat, terméseredmény, gázkoncentrációt, mesterséges intelligencia alapú fogalomalkotás keretében előállt context free-jellegű index-értékeket, stb.) az X-változók lépcsős függvényeként közelítjük, vagyis minden X változóhoz meghatározunk egy-egy lépcsős függvényt pl. ár/érték-arány-elemzés érdekében, ahol az X-tömb csak pozitív egész számokat, azaz rangsor-számokat tartalmaz.” (Pitlik Marcell 2020, Térfogati égés optikai vizsgálata és mesterséges intelligencia alapú elemzése)*

COCO Y0: *„Az anti-diszkriminációs számítások-- más néven az ideálkereső, mesterséges fogalom-alkotó modell jelölése --, ahol minden X esetén az idealitás irányába ható irány megadása után optimalizálás keretében keressük az átlagtól leginkább eltérő objektumot úgy, hogy az optimalizálás célja mindvégig az objektumok azonosságának kikényszeríteni akarása” (Pitlik Marcell 2020, Térfogati égés optikai vizsgálata és mesterséges intelligencia alapú elemzése)*

Benchmark: *„A benchmarking vagy összehasonlító értékelés egy olyan folyamat, amelynek célja, hogy azonosítsa, mérje, összehasonlítsa és megismerje más vállalatok versenyelőnyeinek alapjait függetlenül attól, hogy versenytársak-e vagy sem, és amelyek megmagyarázzák sikereiket” (Tanya Sammut-Bonnici, 2015, o.n.)*