OTDK-DOLGOZAT

2025

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM

Gazdaság- és Társadalomtudományi Kar

A gépi tanulás lehetőségeinek vizsgálata a fogyasztói preferenciák és döntések függvényében

Examining the possibilities of machine learning in relation to consumer preferences and decisions

Szerző: PÁL BERTOLD

KONZULENS: Dr. Iványi Tamás

Kézirat lezárta: 2023.11.02

TARTALOMJEGYZÉK

1. BEVEZETÉS 1

2. A MARKETING ÉS A MESTERSÉGES INTELLIGENCIA TÖBBRÉTŰ KAPCSOLATA 3

2.1 MARKETINGKUTATÁS ÉS A MESTERSÉGES INTELLIGENCIA KAPCSOLATA 5

2.2 STP ÉS A MESTERSÉGES INTELLIGENCIA KAPCSOLATA 7

2.3 A 4P, VALAMINT A 4E KAPCSOLATA A MESTERSÉGES INTELLIGENCIÁVAL 7

2.3.1. Termékpolitikai megoldások (Product – Experience) 8

2.3.2 Árpolitika (Price - Exchange) 9

2.3.3 Értékesítési politika (Place- Everywhere/Everyplace) 11

2.3.4 Reklámpolitika (Promotion- Engagement/Evangelism) 13

2.4 HOGYAN ÉPÜL BE A GÉPI TANULÁS A VÁLLALATI DÖNTÉSBE? 14

3. GÉPI TANULÁSI TECHNIKÁK 17

3.1 TRADICIONÁLIS GÉPI TANULÁS 18

3.1.1 Supervised learning 18

3.1.1.1 Lineáris regresszió a marketing vonatkozásában 19

3.1.1.2. Klasszifikáció a marketingben 19

3.1.2 Unsupervised learning 20

3.1.2.1 Klaszterezés a marketingben 21

3.2 MÉLY TANULÁS 21

4. PRIMER KUTATÁS 22

4.1 MÓDSZERTANI BEMUTATÁS 22

4.1.1 Vezetői döntési probléma bemutatása 22

4.1.2 Iparág kiválasztásának indoklása 22

4.1.3 Szűken vett kutatási kérdések bemutatása 24

4.1.4 Tradicionális vs. modern marketingkutatás 26

4.1.4.1 Tradicionális marketingkutatás 26

4.1.4.2 Modern marketingkutatás 27

4.1.5 Adatgyűjtés 27

4.1.6 Modellépítési folyamat 28

4.1.7 KNIME analytics rövid bemutatása 30

4.2 ADATTISZTÍTÁSI ÉS ELEMZÉSI FOLYAMAT 31

4.2.1 Tényleges kitöltők kiszűrése 31

4.2.2 Hiányzó értékek korrigálása 32

4.2.3 Klasszifikációs célváltozó előállítása 33
4.2.4 Változó képzés 35
4.2.5 Kiugró értékek eltávolítása 37
4.2.6 Célváltozó értékének megfigyelése 37
4.3 PREDIKCIÓS MODELLEK ELEMZÉSE ÉS ÉRTÉKELÉSE 38
4.3.1 A predikciós modell építése és optimalizálása 38
4.3.1.1 Logisztikus regressziós modell 38
4.3.1.2 Random forest modell 53
4.3.2 Első kutatási kérdés 56
4.3.2.1 Logisztikus regresszió esetében optimális modell 57
4.3.2.2 Random forest esetében optimális modell: 62
4.3.2.3 Konklúzió az első dimenzióról 68
4.3.3 Második kutatási kérdés vizsgálata 68
4.3.3.1 Logisztikus regresszió esetében a minták közötti eltérés 69
4.3.3.2 Random forest esetében a minták közötti eltérés 69
4.3.3.3 Konklúzió a második dimenzióról 70
5. A KUTATÁSI EREDMÉNYEK ÖSSZEFOGLALÁSA 71
6. IRODALOMJEGYZÉK 75
Melléklet 79

ÁBRAJEGYZÉK

1. ábra: M.I. és stratégiaimarketing döntések 4
2. ábra: M.I. alkalmazásához szükséges lépések 15
3. ábra: M.I. és stratégiai marketing döntések 17
4. ábra: Gépi tanulási módszerek osztályozása 18
5. ábra: A konfúziós mátrix és az abból származtatható mérőszámok 25
6. ábra: Gépi tanulási modell alkotási folyamat jellemzése 29
7. ábra: Adattisztítási és elemzési folyamat 31
8. ábra: Tényleges kitöltők kiszűrése 32
9. ábra: Hiányzó értékek kezelése 33
10. ábra: Célváltozót előkészítő kérdés 33
11. ábra: Célváltozót előkészítő kérdés táblázatba átvitt értékei 34
12. ábra: célváltozó előállítása 35
13. ábra: Új változók létrehozása a prediktív elemzés elősegítése végett 36
14. ábra: boardELMEGY változó értékeinek megoszlása, Partitioning és SMOTE előtt 38
15. ábra: Partitioning node beállításai 39
16. ábra: boardELMEGY változó megoszlása a tréning adatrészben SMOTE előtt 40
17. ábra: boardELMEGY változó megoszlása a tréning adatrészen SMOTE után 40
18. ábra: tényezők kiválasztásának első fázisa 41
19. ábra: Logistic Regression Learner node beállításai 42
20. ábra: tényezők kiválasztásának első fázisában használt hiperparaméter beállítások 42
21. ábra: k-fold cross validációhoz szükséges beállítások 43
22. ábra: konfúziós mátrix létrehozása 44
23. ábra: Első fázis optimális tényező összetételének a kiválasztása 45
24. ábra: Második fázisban lévő Logistic Regression Learner hiperparaméterei 46
25. ábra: Második fázis optimális tényező összetétele 47
26. ábra: Harmadik fázisban lévő Logistic Regression Learner hiperparaméterei 48
27. ábra: Harmadik fázis optimális tényező összetétele 49
28. ábra: Negyedik fázis ábrázolása 50
29. ábra: Nem diák adathalmazon történő prediktálás 50
30. ábra: Modell véglegesítése 51
31. ábra: Modell hiperparamétereinek véglegesítése 52
32. ábra: Modell tesztelése a diák adathalmaz teszt részén 53
33. ábra: Random forest modell felépítése 54
34. ábra: Random Forest Learner node beállításai 54
35. ábra: Első fázis optimális tényező összetételének a kiválasztása random forest esetében 55
36. ábra: Második fázis optimális tényező összetételének a kiválasztása random forest esetében 56
37. ábra: Optimális Logisztikus regresszió együtthatói 57
38. ábra: Logisztikus regressziós modell teljesítménye a cross validáció során döntési kritérium módosítása nélkül 57
39. ábra: Optimális döntési kritérium a logisztikus regressziós modellen belül maximális találati arány elérése végett 58
40. ábra: Predikciók újra definiálása új döntési kritérium mellett a tréning adatrészen logisztikus regresszió esetében 59
41. ábra: Logisztikus regressziós modell teljesítménye a tréning adatrészen döntési kritérium módosításával 59
42. ábra: Logisztikus regressziós modell teljesítménye a teszt halmazon döntési kritérium módosítása nélkül 60

43. ábra: Predikciók újra definiálása új döntési kritérium mellett a teszt adatrészen logisztikus regressziónál 61
44. ábra: Logisztikus regressziós modell teljesítménye a teszt adatrészen döntési kritérium módosításával 62
45. ábra: Random forest modell teljesítménye a cross validáció során döntési kritérium módosítása nélkül 63
46. ábra: Optimális döntési kritérium a random forest modellen belül maximális találati arány elérése végett 64
47. ábra: Predikciók újra definiálása új döntési kritérium mellett a tréning adatrészen random forest esetében 65
48. ábra: Random forest modell teljesítménye a tréning adatrészen döntési kritérium módosításával 65
49. ábra: Random forest modell teljesítménye a teszt halmazon döntési kritérium módosítása nélkül 66
50. ábra: Predikciók újra definiálása új döntési kritérium mellett a teszt adatrészen random forest modellnél 67
51. ábra: Random forest modell teljesítménye a teszt adatrészen döntési kritérium módosításával 67
52. ábra: Logisztikus regressziós modell teljesítménye a nem diák adathalmazon 69
53. ábra: Random forest modell teljesítménye a nem diák adathalmazon 70

1. BEVEZETÉS

A digitalizációnak köszönhetően új fogyasztói modell jelent meg, a vásárlók szokásai és igényei is megváltoztak. A mai fogyasztó aktívan részt vesz a vásárlás folyamatában, közösen alkot (co-creation) a cégekkel, élményekkel veszi körül magát, és jó színben szeretne feltűnni a digitális társadalomban. Azonban a digitális térben jelentős a kommunikációs zaj és nagyon sok az impulzus, ezért nehéz megszerezni ezt az újfajta fogyasztót, a megtartásuk, aktív vásárlóként tartásuk pedig különösen nagy kihívást jelent a vállalatok számára. Ezért is szükséges a mesterséges intelligencia alkalmazása a vállalkozások számára, amely a klasszikus 4P, valamint az újabban megjelent 4E (élmény alapú) marketing-mix eszközök minden területét jelentősen alakítja. A személyre-szabott tartalom segíti a cégeket abban, hogy ki tudjanak tűnni a zajból és meg tudják szólítani közvetlenül is ezt az új fogyasztót, élményekkel, releváns tartalommal töltsék meg az értékcsere folyamat során átadott termékeket, szolgáltatásokat. Ez a fajta személyre-szabás azonban nehezen kivitelezhető, hiszen nagyon jól kell ismerni a fogyasztót és gyors reakciót is igényel az ilyen szintű kapcsolatépítés. Mindebben tud segíteni a mesterséges intelligencia alkalmazása. Ez segíti azt, hogy a vállalatok hatékonyabban tudják az értékteremtési folyamatot menedzselni, például jobb aránnyal tudják megtalálni a termékekre nyitott fogyasztókat az adatbázisaikban, vagy pontosabb termékösszeállítást tudjanak nyújtani a fizikai és a virtuális üzleteikben is. A kutatás során célom, a témában elérhető tudományos cikkek, korábbi kutatások feldolgozásával, több olyan modell azonosítása, amely a különböző vezetői döntési helyzetekben megmutatja, hogy bizonyos fogyasztói adatok birtokában milyen viselkedés várható egy új fogyasztótól, illetve, azt, hogy milyen modellezési lehetőségek vannak a különféle fogyasztói szegmensek jellemzésére, és hogyan lehet őket alkalmazni a stratégiai marketing különböző területén, például a szegmentációban, a targetálás során, vagy az értékesítési, valamint kommunikációs csatornák menedzselésében. Célom továbbá az is, hogy egy példa rendszert hozzak létre, amely egy általam kialakított adatbázis alapján segíti a minták elemzését, a döntés támogatást. A mesterséges intelligencia is igen széleskörű spektrumban értelmezhető még a marketing területén is, így elsődleges célomnak azt tartom, hogy ennek a jelenségnek az általános vizsgálatát megvalósítsam, de emellett az is fontos számomra, hogy az említett konkrét példát is kialakítsam, amelyet kisebb vállalkozások is tudnak hasznosítani. Mindezzel rá tudok arra is világítani, hogy a mesterséges intelligencia, a tanuló algoritmusok használata lehetséges és nem feltétlenül bonyolult egy kisvállalkozás számára is. Mindezekhez kapcsolódva az is a céljaim közé tartozik a TDK kutatás során, hogy egy olyan szoftver, a KNIME analitikai platform működését is bemutassam, amelyek komolyabb programozási tudás nélkül is tudja segíteni ezeknek az algoritmusoknak a használatát. A dolgozatban bemutatott módszereket kisebb átalakításokkal számos

KKV is hasznosítani tudja, sőt akár új innovatív vállalkozások is rá tudnak a kutatásaim eredményeire épülni.

A dolgozatom ennek megfelelően két fő részből épül fel. Az első felében a mesterséges intelligencia lehetséges alkalmazásait vizsgálom elméleti modellek és példák segítségével a marketing különböző területein, míg a második felében egy konkrét iparágra, a társasjátékozásra fókuszálok egy primer kutatás keretében, amely esettanulmányként és követhető példaként tud szolgálni a KKV-k számára. Ebben a részben az eredmények értékelése mellett fontos volt számomra az is, hogy a kutatás, a modellalkotás és a modellek elemzésének az egyes lépéseit részletesen is bemutassam, hogy az mások számára is részben, vagy egészben reprodukálható legyen.

2. A MARKETING ÉS A MESTERSÉGES INTELLIGENCIA TÖBBRÉTŰ KAPCSOLATA

A technológia fejlődésével mind a számítási kapacitás terén, mind pedig a mesterséges intelligencia (MI) eszközeivel új lehetőségek nyíltak a marketinges szakemberek számára. Ezek olyan kapukat nyitottak meg, amelyek régebben még elképzelhetetlenek voltak. Az új eszközök és lehetőségek magukba foglalják az olyan megoldásokat, amelyek segítségével egy sokkal hatékonyabb és fogyasztó centrikusabb marketingkampányt tudunk felállítani, valamint a marketingstratégiánk részévé is tudnak ezek válni. Mindez segíthet a vállalatoknak abban, hogy jobban ki tudnak tűnni az internet zajos közegéből.

A mai marketingkörnyezet rendkívül komplex, számos elemet lehet vizsgálni, számos befolyásoló tényezővel találkozhatunk és a fogyasztót nagy mennyiségben érik a marketing impulzusok. Ebben az összetett környezetben nehéz elérni a fogyasztót és nagy szükség van azokra a megoldások a marketingszakemberek számára, amelyeken keresztül a termék vagy szolgáltatás ki tud emelkedni, el tudja érni a fogyasztó ingerküszöbét, valamint a fogyasztóhoz közel álló értéket tud bemutatni. Találhatunk még olyan iparágakat, amelyek ez egyelőre még nem olyan mérvadó, de felismerhető egy olyan tendencia is, ahol az érzelmi elköteleződés, a fogyasztói involvement növelése sokkal nagyobb szerepet kapott az elmúlt években. Az ilyen, magasabb elköteleződést kiváltó, vagy akár szeretetmárkákat kialakító értékteremtés során fontos szerepet kap a mesterséges intelligencia. Ebben a fejezetben bemutatom, hogy az MI meglátásom szerint egy olyan új perspektívát építhet bele a cégek marketingtevékenységébe, ami segítségével az érzelmek komplex struktúráját át tudjuk vezetni egy erős értéket generáló folyamattá, mindezt pedig úgy teszi, hogy a fogyasztót, mint egy érző és gondolkodó komplex lényt helyezi a középpontba, követve ezzel a marketing filozófiáját.

A mesterséges intelligencia eszközeiről és kapcsolódó gyakorlati megoldásokról mutat be egy összefoglalást Huang és Rust (2021). A szerzőpáros az eszközöket a marketingtevékenység szempontjából csoportosítja és egy ciklikus folyamatként írja le a marketingben való alkalmazását a mesterséges intelligenciának. A ciklust az 1. ábrán láthatjuk, amelynek három fő elemét tudjuk beazonosítani: marketingkutatás, marketing stratégia (amely alapvetően az alapstratégiák közül az STP folyamatra fókuszál), valamint a marketing akciót (amely a mix-stratégiákat vizsgálja). Ami a ciklus mind a három elemében közös az a mechanikus, gondolkodó és érző mesterséges intelligencia jelenléte, ám ezek jelenléte a három elemben még nem jelenti azt, hogy a folyamat összes részében azonos feladatokat látnak el, ahogy ez az 1. ábrából is látható. A következő alfejezetekben is ezen modellezés mentén fogjuk vizsgálni az MI jelentőségét a marketing egyes területein!

1. ábra: M.I. és stratégiaimarketing döntések
Forrás: Huang és Rust (2021) alapján saját szerkesztés

Huang és Rust (2021) általánosan a ***mechanikus mesterséges intelligenciát*** úgy írják le, hogy annak feladata a repetitív cselekedetek automatizálása, ilyenre jó példa a standardizáció, az adatgyűjtés, vagy a fogyasztók különböző szegmensekbe sorolása, illetve például az osztályozási (klasszifikációs) feladatok lebonyolítása. Lényegében ide tartoznak az olyan mesterséges intelligenciával megvalósítható események, amelyek nem igényelnek különösebb intelligens emberi tevékenységet, a monoton munkavégzést helyettesítik, azt gyorsítják.

A ***gondolkodó mesterséges intelligencia*** célja az, hogy új mélyebb betekintést kapjunk a fogyasztóinkról és új következtetéseket alkothassunk velük kapcsolatban. Ezek a következtetések egy adatbázis komplex elemzésével kaphatóak, a hagyományos statisztikai módszereken túlmutatnak. Látható az 1. ábrán, hogy az ide besorolt tipikus marketingtevékenységek (targetálás, piacelemzés, személyre-szabás) sokkal komplexebbek és nehezebben automatizálhatóak, és itt nem csak a mesterséges intelligencia jelenléte szükséges, hanem egy külső szakértő is fontos szerepet tud betölteni. Ide sorolhatók azok a mesterséges intelligenciával lebonyolított folyamatok, amelyek strukturálatlan adatok, nagy adatbázisok (big data) feldolgozását és értelmezését segítik, valamint az olyan folyamatok is, amelyek a fogyasztókra vonatkozóan fognak előrejelzéseket adni (Huff & Bonde, 2022).

Az ***érző mesterséges intelligencia*** feladata viszont az érzelmi kapcsolat elmélyítése a fogyasztókkal, ez is egy rendkívül komplex folyamat, nehezen automatizálható, azaz itt is nagy szükség van egy a mesterséges intelligencia mellett lévő szakértői átlátásra. A továbbiakban is látható lesz, hogy a marketing folyamatok fontos elemévé fog válni, az érzelmi elmélyítés. Ez azért is lesz majd fontos, hiszen mint ahogyan már említettem a mai vállalati környezet sokkal komplexebb, nagyon sok impulzus éri a fogyasztókat, és az egyszerű mindennapi szimbólumokkal nehéz hosszútávú benyomást kelteni, meg kell ismerni a fogyasztókat és mint egy jó barát támogatni őket és értéket kell teremteni számukra. Ehhez számos esetben az érzelmekre kell hatni és olyan tartalmat kell létrehozni, amely érzelmi szinten is kapcsolódik a fogyasztóhoz.

A következő alfejezetekben az 1. ábrán látható ciklus elemeit fogjuk részletesen is tárgyalni.

2.1 MARKETINGKUTATÁS ÉS A MESTERSÉGES INTELLIGENCIA KAPCSOLATA

A marketingkutatás során fel szeretnék térképezni a piacot, amely során egyik fontos célunk az, hogy modellezni tudjuk a fogyasztókat, a várható magatartásukat és megpróbáljuk őket megérteni (Gyulavári et al.,2017). A fogyasztók modellezése nem egy egyszerű folyamat, a hagyományos statisztikai eszközök ezt limitáltan tudják csak segíteni, azonban a mesterséges intelligencia itt is számos területen tud biztosítani számunkra mélyebb elemzési lehetőségeket, ezen eszköz segítségével egy sokkal pontosabb és fogyasztó-orientáltabb képet tudunk alkotni, ami által a vállalat marketing stratégiája és majd a későbbiekben annak megvalósítása (ami a marketing mix alkalmazásában fejeződik ki) sokkal hatásosabb lehet.

A mesterséges intelligencia egyik alapköve a pontos és jó minőségű adat, vannak olyan technikák, amelyek akár kisebb mennyiségű adattal is jól működnek ilyenek a tradicionális gépi tanulás eszköze, de legtöbb esetben hüvelykujj szabályként azt szokták mondani, hogy minél több adat van a fogyasztókról annál jobb (Qi & Luo, 2020). Az adatgyűjtésben segíthet a ***mechanikus mesterséges intelligencia*** azzal, hogy az adatgyűjtési folyamatokat automatizálja és ezáltal mindig naprakész és pontos adataink tudnak lenni a fogyasztókkal kapcsolatban, ami elkerülhetetlen lehet egy sikeres marketing kampány megtervezésében és levezénylésében. Az automatizált adatgyűjtésnek köszönhetően Ezáltal most már nemcsak egy múltbéli képet tudunk alkotni a fogyasztókról, hanem akár valós időben tudjuk megvizsgálni a fogyasztó viselkedését, attitűdjét és érzelmeit, és így valós időben tudunk reagálni is ezekre a komplex folyamatokra. Erről a folyamatról számol be például Andrejevic (2019), aki szerint egy olyan társadalom felé haladunk, ahol a kényelem nagyon fontos és ezért a vállalatok számára egyre számottevőbb a sokoldalú adatgyűjtés, hogy a

fogyasztók számára a lehető leginkább optimalizált élményt lehessen biztosítani. Ezt a jelenséget Andrejevic (2019) egy úgynevezett megfigyelő gazdaságként jellemzi, ami magába foglalja a folyamatos és sokoldalú, számos csatornára kiterjedő adatgyűjtést.

De nem elég az adatokat csak gyűjteni, azokat elemezni és értelmezni is kell. Ebben tud segíteni a ***gondolkodó mesterséges intelligencia***, például klaszterezés keretein belül képet tudunk alkotni, arról, hogy milyen szegmentációs felosztással tudnánk legjobban lefedni a piacot és hogyan tudnánk növelni a személyre szabhatóságot (Ricotta & Costabile, 2007). Ezeket az algoritmusokat a 3. fejezetben részletesen is fogom majd tárgyalni.

Az ***érző mesterséges intelligencia*** segítségével értelmezhetünk olyan folyamatokat, amik a fogyasztókon belül játszódnak le és nehezebben számszerűsíthetők, vagy egyszerűen nem számszerűsíthetők. Ilyen kvalitatív elemzési feladat lehet a Natural Language Processing (NLP), amely során megpróbáljuk az emberi nyelvet lefordítani a gépek által értelmezhető struktúrára. Ezek után különféle - a mesterséges intelligenciára épülő - algoritmusok segítségével próbálunk következtetéseket adni a fogyasztók érzéseire azáltal, hogy az általuk megfogalmazott gondolatokat (például internetes hozzászólásokat) elemezzük, és egy képet próbálunk kapni arról, hogy milyen belső folyamatok játszódnak le a fagyasztóinkban (Luisa, Mariangela, & Pierluigi, 2004). Tegyük fel például, hogy van egy termékünk, amit az amazon.com-on keresztül forgalmazunk. Egy ideje már jelen van a piacon és az Amazon platformján számos értékelés érkezett be a termékkel kapcsolatban. Ezeket az értékeléseket különböző csoportokba lehet sorolni (1-5 csillag, ahol 5 csillag a legjobb értékelés, 1 pedig a legrosszabb, illetve az értékelés mellé megjegyzést lehet fűzni). Azt szeretnénk megtudni, hogy például az 1-es értékelésre milyen szavak jellemzőek, milyen érzések alakultak ki a fogyasztókban. Ezt gépi tanulás segítségével el lehet érni és elemezhetővé lehet tenni a szöveges adatot. Az NLP-s módszereken keresztül szófelhőket tudunk alkotni a különböző értékelési kategóriákhoz. Például képesek leszünk alkotni egy szófelhőt, ami jellemzi az 1 csillagos értékeléseket, ezáltal tudunk kapni egy képet arról, hogy milyen problémába ütközhettek a fogyasztók, és később nagyobb fókuszt adhatunk ezeknek a problémáknak. Ám nem csak ez az egy terület, ahol kiemelkedő lehetőségeket biztosít az úgynevezett social listening. Ha jobban meg szeretnénk ismerni fogyasztóinkat, esetlegesen azt, hogy milyen problémákkal küzdenek, miket szeretnek, nem szeretnek, mi az, amit értéknek találnak, abban mind segíteni tudnak számunkra az NLP alapú algoritmusok: rámutatnak arra, hogy milyen kontextusban említik a fogyasztóink a márkánkat, termékeinket, vagy milyen érzelmek jellemzik az értékeléseiket (Piris & Gay, 2021).

2.2 STP ÉS A MESTERSÉGES INTELLIGENCIA KAPCSOLATA

A mai világban a szegmensmarketing már önmagában nem elegendő, a klasszikus szegmentációs folyamatot szükséges finomhangolni, ezért is emeli ki Hunag és Rust (2021) azt, hogy a fent említett mechanikus, gondolkodó és érző mesterséges intelligenciát a stratégiaalkotásba is szükséges beépíteni a már jól bevált STP folyamat új szintre emelésével. Az STP három kulcsfontosságú döntést hordoz magában a szegmentálást, a célcsoport meghatározását, és a pozicionálást. Ez egy jó keretrendszert tud adni a különböző mesterséges intelligenciára alapozott algoritmusok besorolására.

A szegmentálás során a célunk az, hogy különböző fogyasztói csoportokat tudjunk elkülöníteni egymástól, hogy a későbbiekben ezen fogyasztói csoportok némelyikére különböző stratégiát tudjunk kidolgozni. A ***mechanikus mesterséges intelligencia*** megfelelő tud lenni erre a feladatra, ugyanis képes arra, hogy mintázatok és szegmenseket, csoportokat beazonosítson. Megfelelő adatmennyiséggel akár olyan kis méretű halmazokat is létre tudunk hozni, hogy bizonyos fogyasztók számára az egyedi és személyre szabott élmény látszatát tudjuk biztosítani, amely segítségével egy sokkal erősebb érzelmi reakciót tud kiváltani a fogyasztóból a vállalat, amely például hosszútávú elköteleződéshez és lojalitáshoz is tud vezetni. Erre jó példa lehet a biztosítási piac, ahol a finom szegmentáció hatására sokkal jobb ajánlatokat tudunk biztosítani a fogyasztó igényeire összpontosítva (Yum, Yoo, & Lee, 2022). Például ilyen lehet az utas biztosítás személyre szabása, olyan szegmentált lehetőségek mentén, amin keresztül a fogyasztó meg tudja találni számára az optimális utas biztosítást.

A pozicionálás során el akarjuk helyezni a termékünket a fogyasztó fejében. A mesterséges intelligencia ebben is tud segíteni. Itt akár Hunag és Rust (2021) által felállított modell alapján egy olyan fajta megközelítést is meg lehet említeni az egyik ilyen megközelítés a ***gondolkodó mesterséges intelligencia***, ennek segítségével elemezni tudjuk a fogyasztói szegmenseket mélyebben, például és egy diverzebb pozícionálási stratégiát is tudunk alkotni (Hassan & Tabassum, 2018). Ezt a folyamatot az ***érző mesterséges intelligencia*** segítségével fel lehet erősíteni, azzal, hogy ezekhez a szegmensekhez érzelmi elemeket is társítunk ezt esetleg szentiment analízissel lehet elérni.

2.3 A 4P, VALAMINT A 4E KAPCSOLATA A MESTERSÉGES INTELLIGENCIÁVAL

Az STP stratégiák mellett Hunag és Rust (2021) a 4P területére, vagyis a marketing mixbe is belehelyezi a már fent említett három mesterséges intelligencia típust. A marketing mix mind a négy - a termék (product), ár (price), értékesítési csatorna (place), és a promócióból (promotion) - területére számos példát mutatnak, hogy a mesterséges intelligencia hogyan tudja segíteni a különböző

marketing eszközök használatát. Ne felejtsük el azonban, hogy a digitális tér létrejöttével a fogyasztó egy új módon jelenik meg, egy úgynevezett mesterséges imázst hoz létre önmagáról, és sokkal jobban képes szabályozni a körülötte mozgó digitális világot. Tulajdonképpen egy új egyén, új személy, új énképpel válik belőle, aki szabadon formálja azt a világot, amit ő alkot magának az online térben. Élményekkel veszi körül magát, és jó színben szeretne feltűnni a digitális társadalomban, ezért is formálja a világát (Freedman, 2019). A digitális térben jobban ki tudja fejezni önmagát, mint a fizikális térben, és szabadabbá válik. Ám ennek az új fogyasztónak ára is van, mivel a digitális térben hatalmas a zaj és nagyon sok az impulzus, ezért nehéz megszerezni ezt az újfajta fogyasztót. A digitális térben az egyszerűbb, klasszikus 4P marketing mix felfogás már kevésnek bizonyul (product, price, place, promotion), és egy úgy mix stratégia veszi át a helyét, hogy jobban biztosítva legyen a zajból való kiemelkedés. Ez a marketing mix modell a 4E (experience, exchange, everywhere/everyplace, engagement/evangelism). A modellt Brian Featherstonehaugh, az Ogilvy One Worldwide elnöke és vezérigazgatója mutatta be, és célja az újkori, digitális fogyasztó jobb megértése és hatékonyabb kiszolgálása (Aditya, 2020). Mindezek miatt a következőekben kitérek arra is, hogy hogyan lehet alkalmazni a mechanikus, a gondolkodó és az érző mesterséges intelligenciát a 4P mixben, majd elemzem azt is, hogy esetlegesen ezt a keretrendszert hogyan lehet áthelyezni az a modern fogyasztóra jobban illő 4E marketing mixbe.

2.3.1. Termékpolitikai megoldások (Product – Experience)

Product:

A terméket Hunag és Rust (2021) keretrendszere alapján a mesterséges intelligencia több módon is tudja fejleszteni. A ***mechanikus mesterséges intelligencia*** tud segíteni a terméktervezési folyamatokban, ilyen lehet az, hogy egy-egy cég esetlegesen a fogyasztói igények alapján automatizálva készíti el a termék logóját vagy a csomagolás dizájnját. Erre említhetjük példának a Nutella által alkalmazott termékpolitikai döntést, amin keresztül a termékük csomagolását mesterséges intelligencia segítségével a fogyasztói igényeknek megfelelően színes és kreatív csomagolásba öltöztették (Aouf, 2017).

A ***gondolkodó mesterséges intelligencia*** segítségével különböző elemzéseket és piaci trendeket értelmezhetünk, amiket bele tudunk építeni termékstratégiánkban. Ilyen lehet például az az eset, amikor egy autógyártó vállalat gondolkodó mesterséges intelligenciát használ a termék tervezés során, aminek következményeként be tudja építeni azokat a célpiaci trendeket és igényeket, amiket a fogyasztók támasztanak a termékfejlesztések során ezáltal jobban kiszolgálva és előzetesen megtippelve a fogyasztók igényeit (Grühn et al., 2019). Az ***érző mesterséges intelligencia*** egy hidat

tud képezni az ügyfelek és a termék között. Ez a híd úgy jelenhet meg, hogy idővel beleépítjük a fogyasztói bázis érzelmi profilját a termékünkbe és így a termékünk asszociációba tud kerülni azokkal az értékekkel, amiket a fogyasztó preferál. Erre nagyon sok jó példa van, egy ilyen példa tud lenni az IKEA által indított kampány, ami a nők elleni erőszakra (domestic violence) próbálta felhívni a figyelmet, mivel vásárlóik nagy részét családok teszik ki ezért is zseniális egy kampány, ami képes értéket adni és érzelmeket megszólítani. Az ***érző mesterséges intelligencia*** segítségével ehhez, hasonló érzelmi dimenziókat könnyebb felfedezni megkönnyítve ezzel az építkezési folyamatot.

Experience:

A 4E modellben a termék (product) átváltozik élményre, a fogyasztók egyre kevésbé különböztetik meg a terméket és a szolgáltatást. Az egyszerű határ a termékek és a szolgáltatások között lebomlott és a fogyasztó komplex élményekre vágyik, hogy jobban ki tudjon teljesülni, önjutalmazni, önkifejezni. Egy terméknek valamilyen szolgáltatást kell nyújtania, és minden terméknek valamilyen kulcsfontosságú észlelhető előnnyel kell rendelkeznie, ilyen például Johnson's Baby, amely ugyan fizikai termékeket kínál, de mellette támogatást is nyújt az anyának arra, hogy hogyan fejlessze a gyermek érzékeit (Aditya, 2020). Látható, hogy így a Johnson’s Baby nem csak egyszerűen egy terméket ad a piacra, hanem a gyermekneveléssel kapcsolatos élményt, és biztonságot is hozzá építi a termékéhez így megfelelő élményt és értéket biztosítva az anyukák számára. Az ilyen élményeket relatív nehéz megteremteni, ezért is fontos a mesterséges intelligencia, és azon belül a fogyasztói preferenciákkal kapcsolatos adatok gyűjtése, elemzése, hogy egyedülálló és a fogyasztóhoz közelálló élményt tudjunk biztosítani. MI-t használó megoldás lehet például egy repjegyeket árusító weboldal szempontjából, ha a jegy mellé útiterveket és programokat generálni az adott úticéllal kapcsolatban ezzel javítva a fogyasztó élményét. ilyen például a KAYAK, ami segít a felhasználónak repjegyek vásárlásában, hotelek foglalásában, teljes, komplex útitervek elkészítésében (Manhas, 2023). A ***gondolkodó mesterséges intelligencia*** tud segíteni a különféle repjegy csomagok összeállításában a fogyasztói igényekre reflektálva. Az ***érző mesterséges intelligencia*** pedig abban tud segíteni például, hogy egy online vásárlási folyamat során megfigyeli a fogyasztó viselkedését és hogyha úgy érzékeli, hogy a fogyasztó frusztrálttá vált vagy elakadt a vásárlási folyamat során, akkor egy chatbotot küld a vásárló segítségére, hogy tovább léptesse a fogyasztót a vásárlási folyamatban.

2.3.2 Árpolitika (Price - Exchange)

Price:

Az árazásban a mesterséges intelligencia is különböző módon is tud segíteni a korábban említett hármas struktúrát is figyelembe véve. A ***mechanikus mesterséges intelligencia*** az árképzési feladatok ellátásában kap szerepet azáltal, hogy segít a különböző fizetési módszerek automatizálásában. Ilyen lehet például a TESLA által használt árképzési stratégia, amiben a mechanikus mesterséges intelligencia segít reagálni a piaci változásokra és a versenytársak árazási stratégiájához viszonyítva alakítja úgy az árakat, hogy optimális helyzetet alakítson ki a Tesla és a vásárlók számára is (Candelon et al., 2022). Ez a fajta árképzés a dinamikus árazás, amely folyamatosan függ a kereslettől és a kínálattól, időben változik és egy mechanikus algoritmus alakítja ki ezeket a változásokat (Danyi, Rekettye, & Veres, 2020). A ***gondolkodó mesterséges intelligencia*** az árak dinamikus beállításán túl a személyre-szabásában is tud segíteni, ilyen megoldás lehet például egy repülőjegy foglalási platformon lévő dinamikus árazás, ahol a mesterséges intelligencia elemzi a repülőjegy árak változásait, valamint a felhasználók preferenciáit és ezáltal egy a vásárló számára optimális árat tud szabni. Ez azt is jelentheti például, hogy az algoritmus előrejelzi, hogy kik lehetnek a potenciális ügyfelek, akik többet is hajlandóak fizetni egy jegyért, és kik lehetnek azok, akik számára érdemes inkább kedvezményesebb árakon nyújtani a szolgáltatást. Az ***érző mesterséges intelligencia*** a különféle áralkuk kezelésére lehet optimális megoldás, ahol nagyon jelentős szerepet tud játszani az érzelem, a fogyasztók érzései és a megfelelő kommunikáció is. Ezen esetekben a vásárlási folyamatot erősen befolyásoló tényezőként tekinthetünk arra, hogy a fogyasztó is megelégedve legyen az árral és ne érje őt érzelmi csalódás mert ez hosszútávon a fogyasztó lojalitását is képes befolyásolni a márkákkal és termékünkkel kapcsolatban.

Exchange:

A 4E modellben a termék ára (price) átalakul és egy cserefolyamatként (exchange) jelenik meg. Sok márka ma már nem csak az ár alapján kell, hogy értéket nyújtson a fogyasztóinak. Ki kell emelkedniük az erős zajból és olyan, akár társadalmi értékeket kell képviselniük, amik a fogyasztói számára elengedhetetlenek. Ezért is sokkal komplexebb ez a tényező, mint egy egyszerű ár, amivel felhívjuk a figyelmet. Sokkal kétoldalúbbá válik ez a folyamat és a fogyasztó még inkább a középpontba kerül, azzal, hogy az ő figyelme, érzelmei, értékei is beépülnek ebbe a cserefolyamatba. Számos esetben megfigyelhető az, hogy a nagyobb vállalatok aktív szerepet vesznek fel a társadalomban, és közösségi értékeket képviselnek, például a globális felmelegedéssel és hasonló fenntarthatósági kérdésekkel foglalkoznak. A klasszikus árpolitika helyetti értékcserek-folyamatra jó példa lehet a Zomato szolgáltatása, amely fogyasztó számára nem kerül semmibe, csak a személyes ételpreferenciákkal (kedvelések és nemkedvelések) kapcsolatos információkat kéri el. Ebben az esetben a fogyasztók nagy értékű személyes információikat cserélik el olyan szolgáltatásokért,

amelyek számukra teremtenek értéket, a cég számára pedig a mesterséges intelligencia bemenetét tudják nyújtani (Bye bye 4Ps, hello 4Es, 2015). Az érték csere- folyamatnak a megtervezésében is több ponton tud segíteni a mesterséges intelligencia. A ***mechanikus mesterséges intelligencia*** a rutinszerű tranzakciók és az ügyfélközi interakciók felgyorsítására és automatizálására alkalmas. A chatbotok használata támogathatja a rendelés feldolgozást és az ügyfelek egyéni kéréseinek kezelését, árazását, a nyújtott árajánlat személyre-szabását, ami javíthatja az értékcsere-folyamat hatékonyságát. A ***gondolkodó mesterséges intelligenciát*** fel tudjuk használni az ügyfél adatok és viselkedés elemzésére, ezzel akár a kommunikációs folyamatunkat személyre tudjuk szabni és egyértelműbbé tenni az értékcsere folyamatot. Az ***érző mesterséges intelligencia*** ugyancsak több lehetőséget ad az érzelmek beépítésére az értékcsere-folyamatba.

2.3.3 Értékesítési politika (Place- Everywhere/Everyplace)

Place:

A harmadik P az a place, vagyis az értékesítési csatorna. Ez is egy nagyon fontos és kritikus pont, hiszen a 4P-nek ez az a pontja, ahol döntés születik arról, hogy a termék hogyan és milyen módon jut el a fogyasztóhoz. Ennek nem optimális megválasztása, a későbbiekben olyan problémákat is tud okozni, hogy a fogyasztó nem képes hozzáférni a termékhez, csak olyan módokon tud hozzájutni, ami kényelmetlen számára. Ezek alapján hosszú távon akár el is pártolhat termékünkről/szolgáltatásoktól, és olyan más márkákat választhat, amiknél a hozzáférés sokkal kényelmesebb és kevésbé leterhelő számára. Ezért is nagyon fontos az optimális döntések meghozatala az értékesítési csatornákat illetően. Ebben is nagyon sokat tud segíteni a mesterséges intelligencia. A ***mechanikus mesterséges intelligencia*** például segíthet egy e-commerce cégnek a rendelések és szállítások optimalizálásában, megfelelő útvonaltervezésben ezáltal a szállítások hatékonyabb véghezvitelében, erre jó példa az Amazon, aki mesterséges intelligencia segítségével újra gondolta az értékesítési csatorna jelentőségét ezzel elnyerve magának egy nagy piaci részesedést (Anh, 2019). A ***gondolkodó mesterséges intelligencia*** sok módon hozzájárulhat a vásárlási procedúra kényelmessé tételében, esetlegesen ilyen megoldás lehet az AmazonGO által alkalmazott teljesen automatizált boltok, amelyek például képesek segíteni a fogyasztóknak a termékek megtalálásával ezáltal gördülőképesebbé tudják tenni a fogyasztási folyamatot (Wankhede, Wukkadada, & Nadar, 2018).

Everywhere/Everyplace:

Az értékesítési csatorna (place) átalakul, és a vállalatnak, nem csak bizonyos értékesítési csatornákon kell jelen lennie, hanem fel kell térképeznie az összes értékesítési csatornát és minden a

fogyasztói számára jelentős helyen ott kell lennie, hogy el tudja érni őket ebben a zajos környezetben. Ez maga egy nehéz folyamat hiszen mint mindennapi emberek mi is tudjuk, hogy az internet az emberi érzékeket sok esetben túltölti, hiszen mindenhonnan jön az impulzus és az új információ. *"A vásárlók vásárlási útja már nem diktálható, és több tényező, mint a kényelem és a személyes preferencia alapján választják meg"* (Aditya, 2020). Ezért nagyon fontos az, hogy maga a stratégia és a kommunikáció is optimális legyen. Ebben is segítséget nyújthat a mesterséges intelligencia, hiszen különböző technikákkal, mint például a szentiment analízis, sokkal pontosabb képet tudunk adni arról, ami a fogyasztó fejében járhat, vagy még ilyen technika a reach (elérés) és reakciók figyelése tartalmainkkal kapcsolatban. Ezek alapján próbálunk kialakítani egy képet arról, hogy hol és mikor, mivel, és hogy tudjuk elérni a célközönségünket. Fontos még azt is kiemelni, hogy ez az újszerű fogyasztó nem feltétlenül a megszokott módon fogyasztja az interneten fellelhető információt. Megjelenik egy úgynevezett olyan jelenség, aminek a neve az, hogy multi-screening. Ez a jelenség azt foglalja magába, hogy ez az újszerű fogyasztó egyszerre több képernyőn is fogyaszt tartalmat legyen az akár 2-3 is egyszerre. Erre egy remek példa az, hogy például a fogyasztó telefonon néz hirdetéseket, és ha talál valamit, ami a számára tetszetős, akkor átvált egy másik eszközre és ott folytatja tovább az információ gyűjtést, vagy kiegészíti az információ gyűjtési folyamatot egy másik eszköz bevonásával. Ezért kiemelkedő szerepet kap ebben az újszerű marketing megközelítésben az omni channel marketing. Az omni channel marketing lényegében online és offline marketing technikákat ötvöz egybe, azért, hogy a fogyasztó számára egy konzisztens élményt és jelenlétet tudjon biztosítani. Fontos kiemelni azt is, hogy egy jó omni channel stratégia lényege az, hogy egyszerű és könnyű átváltást biztosítson a online és offline tér között. Egy kiemelkedő stratégia ezen a területen képes könnyedén és invitálóan végig vezetni a fogyasztót a fogyasztás különböző fázisain, úgy, hogy az élmény folyamatos marad számára. Az Asian Paints nagyon jó példa erre. *"A vállalat tudta, hogy csak a kiskereskedelmi szegmensben való jelenlétével nem tud jelentős növekedést elérni. Különböző sikeres marketing technikákat vezettek be, és eljutottak a vásárlókhoz olyan csatornákon keresztül, mint hirdetések, mobilalkalmazások, TV reklámok és YouTube hírességek videói. Ennek eredményeként az Asian Paints most az indiai piaci arénában vezeti a versenyt"* (Aditya, 2020). A ***gondolkodó mesterséges intelligenciát*** itt úgy tudjuk felhasználni, hogy ezen keresztül megpróbáljuk megérteni a vásárlók viselkedését és preferenciáit különböző érintési pontokkal kapcsolatban, így akár jobban személyre tudjuk szabni a vállalattal kapcsolatos interakció folyamatát és akár a fogyasztók egy sokkal személyre szabottabb úton vehetnek részt a fogyasztás folyamatában (Walsh, 2018). Ez azért lehet fontos mert így a vállalat egy sokkal személyre szólóbb élményt tud biztosítani a fogyasztónak. Ilyen megoldás lehet például egy weboldal tartalmának optimalizálása, mobilapplikáció funkcióinak módosítása vagy esetleg a vállalat közösségi média tartalmainak a fogyasztói igényekre való átformálása (Abi, Zakraoui, & Benahmed, 2021). Az ***érző mesterséges***

intelligencia képes lehet arra, hogy fokozza az érzelmi kapcsolatot különböző érintési pontokon. Ilyen lehet egy olyan mesterséges intelligencia alkalmazása, ami képes igazodni a fogyasztó érzelmi állapotához, és ahhoz megfelelő kommunikációs stílust ölt magára, így kiszolgálva a fogyasztó érzelmi dimenzióját.

2.3.4 Reklámpolitika (Promotion- Engagement/Evangelism)

Promotion:

A 4P legutolsó eleme a promóció, ami azokat a tettleges folyamatokat foglalja magába amik során a piaci jelenlétünket sugározzuk a fogyasztó felé. A modern világban a promóciónak kiemelkedő szerepe van, hiszen ezen keresztül tudják megismerni a fogyasztók termékünket/szolgáltatásunkat. A nem megfelelő promóció következménye, akár az is lehet, hogy a fogyasztó számára nem elég személyre szóló, így nem tudja megmutatni a termék/szolgáltatás értékét ezért nem marad meg benne, ilyen probléma lehet még az hogyha nincsenek meg azok az érzelmi elemek és az a komplexitás, amit a fogyasztó megigényel, hogy értékként kezelje termékünket/szolgáltatásunkat. Ezeknek következménye akár más terméket/szolgáltatást választása lehet. Emellett a promóciós tevékenység nagyon magas költségekkel is tud járni ezért nem elég véletlenszerűen promotálni termékünket/szolgáltatásunkat, hanem a promotálásnak egy jól kitervelt összecsengő tevékenységnek kell lennie, ami képes értéket kommunikálni a fogyasztó felé, megszólítani őt érzelmi szinten és optimális választásként kitűnni a tömegből. Mint ahogy látható ez egy nagyon összetett folyamat, aminek optimális véghezviteléhez mély ismeret kell a fogyasztókról. A mesterséges intelligencia ebben is számos lehetőséget tud adni. A ***mechanikus mesterséges intelligencia***, akár olyan módon is tud segíteni, hogy a marketing kampányon belül bizonyos paraméterek alapján a tartalom generálási folyamatok automatizálhatók vele, ami segítségével rendszeresen optimális tartalmat tudunk biztosítani fogyasztóink számára ezzel hosszú távon lekötve figyelmüket (Reisenbichler, Reutterer, Schweidel, & Dan, 2022). Ezen releváns tartalmakon keresztül meg tudjuk szólítani az érzelmi dimenziójukat is. A ***gondolkodó mesterséges intelligencia*** segítségével esetlegesen elemezhetünk piaci trendeket és ezeket beépítve az előzőben leírt megközelítési módba folyamatosan generálhatunk tartalmakat, amik így dinamikusak maradnak a fogyasztóink igényei számára. Az ***érző mesterséges intelligencia*** figyelemmel tarthatja a fogyasztó reakcióit és viselkedését ez alapján módosíthatja a tartalmakat.

Engagement/Evangelism:

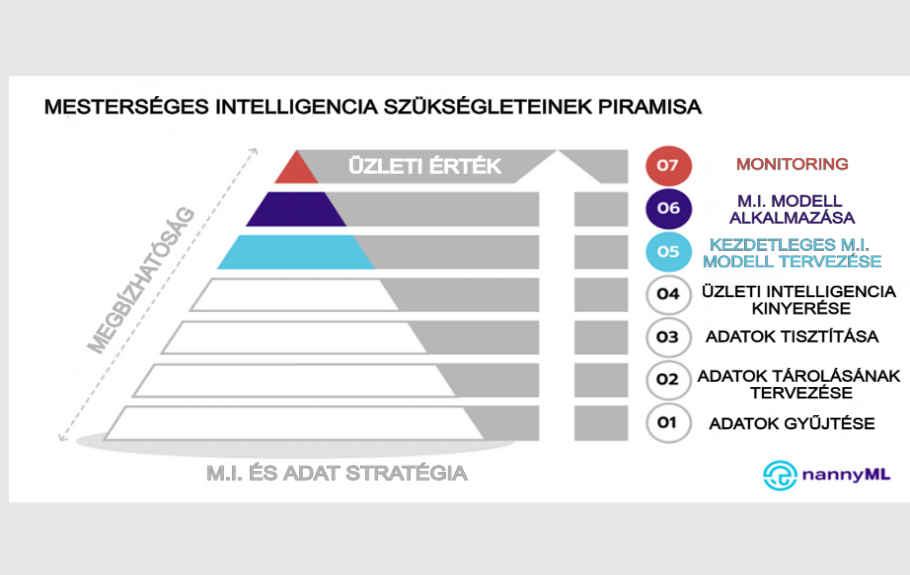
Ez nagyrészt a márka iránti szeretet elősegítéséről és a relevanciáról szól. "*Egy olyan fogyasztó, aki szereti a márkádat, sokkal értékesebb, mint egy teljes kampány, amely promótálja a*

márkádat, ez a fogyasztó beszélni fog a márkádról sok más, hasonló gondolkodású potenciális fogyasztónak, és még meg is győzheti őket arról, hogy a te fogyasztóid legyenek, olyan módon, amit te sosem tudnál megtenni" (Bye bye 4Ps, hello 4Es, 2015). Itt is nagyon fontos felhozni azt, hogy ebben mit is tud segíteni a Big Data. Social listeningen, és különféle tartalmak teljesítményének a monitorozásán keresztül (like számok, reakciók, kommentek, megosztások figyelése), optimalizálni tudjuk a tartalmakat, és olyanra szabni őket, amilyennek a fogyasztóink szeretik, ezzel nagyobb elköteleződést elérve, folyamatosan relevánsak tudunk maradni számukra és kapcsolatot tudunk építeni velük. Itt is nagyon sok lehetőséget tud biztosítani a mesterséges intelligencia mind a 3 dimenzión keresztül. Ilyen megoldás lehet egy ***mechanikus mesterséges intelligencia*** szempontjából az a megoldás, hogy bizonyos tartalom orientált tevékenységeket automatizálunk. Ilyen megoldás lehet különféle e-mail marketing kampányok automatizált lebonyolítása (Babet, 2020). Erre egy jó példa a tipikus hírlevél, ami folyamatosan tájékoztatja a fogyasztót az újdonságokkal kapcsolatban és így nem marad le és ezáltal nem érzi magát lemaradva és vagy kihagyva az újdonságokból. A ***gondolkodó mesterséges intelligencia*** itt esetlegesen olyan segítséget tud nyújtani, ami segít a fogyasztói preferenciák mélyebb megértésében ezáltal a vállalat képes lesz tartalmát jobban személyre szabni és ezáltal folyamatosan releváns maradni a fogyasztó számára (Khan, 2023). Az ***érző mesterséges intelligencia*** itt olyan segítséget tud nyújtani, hogy fokozza a fogyasztó érzelmi elkötelezettségét a vállalattal kapcsolatban. Ilyen megoldás lehet a különböző közösségi média platformokon keresztül való fogyasztó-vállalat interakciók figyelése.

2.4 HOGYAN ÉPÜL BE A GÉPI TANULÁS A VÁLLALATI DÖNTÉSBE?

A mesterséges intelligencián belül fontos kiemelni a gépi tanulást, mint intelligens megoldást a marketing feladatok ellátására. Maga a gépi tanulás egy eszköztárat takar, aminek a célja az, hogy adatok modellezésével, belátást tudjon biztosítani komplexebb folyamatokba, és képesek legyünk ez által predikciókat készíteni a jövővel kapcsolatban. A gépi tanulás a statisztika és a programozás eszköztárából épül fel, és feladata az, hogy ne kelljen lépésről lépésre utasítani magát a programot, hanem képes legyen felismerni mintákat az elérhető adatokból, és konklúziót tudjon levonni belőlük, ami alapján egy új adathalmazon következtetéseket tudjon formálni és predikciók alapján segíteni magát az elemzést. Amikor gépi tanulásról beszélünk akkor modellekre gondolunk, amik a valós világban történő folyamatokat, próbálják reprezentálni a modelleken keresztül, nagy mennyiségű adat segítségével. A gépi tanulás segítségével, olyan komplex Big Data-val kapcsolatos feladatokat is el tudunk látni, amit tradicionális elemzési metódusokkal képtelenek lennénk, maga az adathalmaz komplexitása és mérete végett. Gépi tanulást körülölelő döntési folyamat az üzleti életben lényegében 3 fő lépésből áll, az üzleti problémából, az üzleti lehetőség felismeréséből a megfogalmazódott

helyzetben, és maga a tettleges akcióból, amit a vállalat produkál az előző 2 pont ismeretében. Ennek implementálásában a mesterséges intelligencia szükségleteinek piramisa segít a piramis maga a maslowi szükséglet piramis alapján épül fel, ahol a felsőbb szintek nem tudnak megjelenni anélkül, hogy az alsóbb szintek ki lennének szolgáltatva maximálisan, tehát a vállalat csak úgy tud mesterséges intelligenciát alkalmazni effektíven, ha minden szintet ki tud szolgálni. Ezt a folyamatot fogom bemutatni az 2. ábrán keresztül.



2. ábra: M.I. alkalmazásához szükséges lépések
Forrás: Cools (2021) alapján saját szerkesztés

A legalsó szinten az adatok gyűjtése található. Lényegében ez az 1. szint a Big Data vállalati döntésekbe való beépítésében. Ez a szint azzal foglalkozik, hogy hogyan tudunk különböző forrásokból kinyerni adatokat, amiket majd később a vállalat fel tud használni. Mivel nem elég csak összegyűjteni az adatokat, tárolni is kell őket, ezért a 2. szint azzal foglalkozik, hogy az összegyűjtött adatokat hogyan lehet megfelelő és megbízható formában tárolni a későbbi felhasználás végett. Egy példa elejéig tegyük fel, hogy egy zenemegosztó szolgáltatásunk van. Ez egy freemium (a használat ingyenes, de reklámokkal van tele, a prémium verzió mentesíti a felhasználót a reklámoktól, és más előnyökkel is ellátja őt például korai hozzáférés új albumokhoz) alapon működő vállalkozás, aminek van applikációs és webes verziója is. Ebből a két forrásból gyűjtünk adatokat a fogyasztókkal kapcsolatban, mint például előfizetés kategóriája (éves, havi, ingyenes tag), profil regisztrációjának ideje, előfizetés kezdete, újra vásárlások száma, használat intenzitása, lemondta-e az előfizetését,

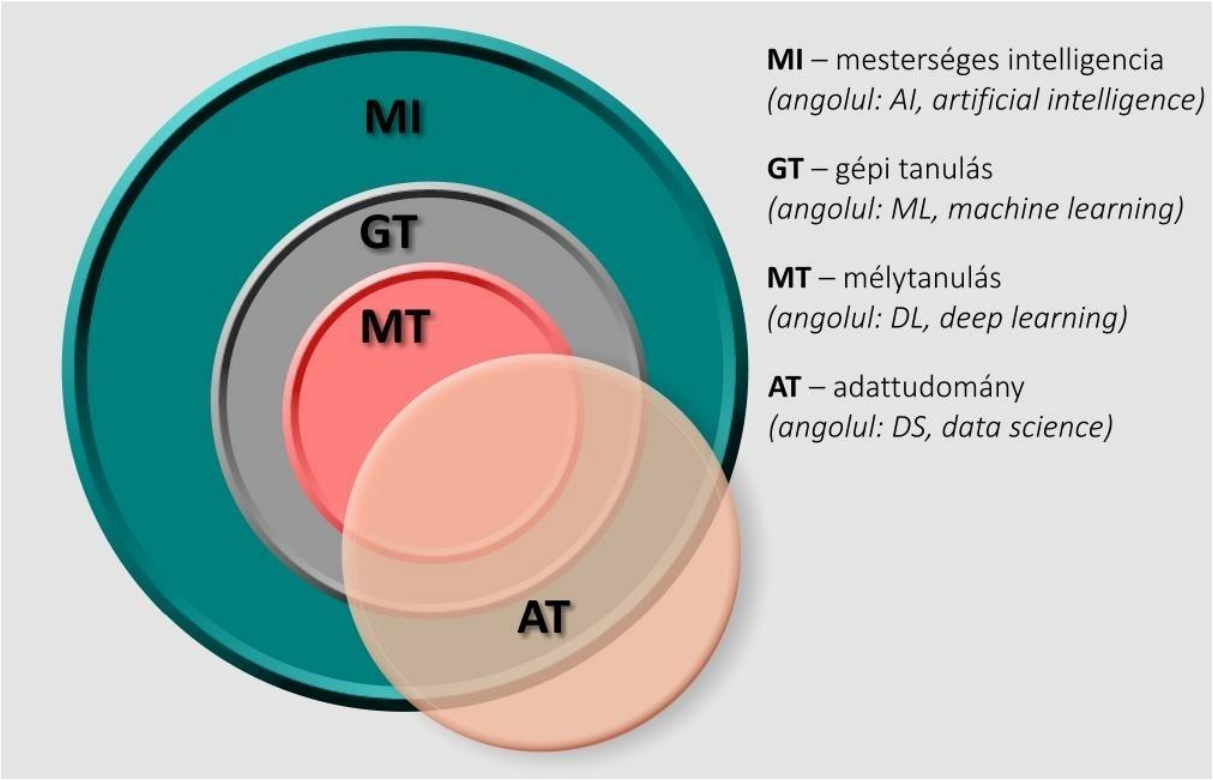
platformon eltöltött idő, kedvelt zenék, követett zenei előadók stb. Ezeket a gyűjtött adatokat kell majd tárolnunk és rendszereznünk egy adatbázisban.

A piramisnak a 3. szintje magával az adatok előkészítésével foglalkozik, mivel az adatok jelen pillanatban csak tárolt és nem tisztított formában vannak, ezért ennek a szintnek az a felelőssége, hogy előkészítse az adatokat a 4. szint szükségleteinek megfelelően. Itt a példánkkal élve vegyük azt a feladatot, hogy a gyűjtött adatokat kell az elemzésnek megfelelően előkészítenünk, hogy majd a későbbi elemzés során értékes a vállalat számára hasznos információt kinyerjünk. A 4. szint maga az adatok elemzésével foglalkozik és azzal, hogy az adatok vizsgálatával értékes trendeket egyéb problémákat, és üzleti lehetőségeket tárjon fel, amiket majd később gépi tanulással értékesíteni lehet. Ez az a pont, ahol maga az üzleti probléma felfedi magát és az, ahol az üzleti lehetőséget is fel lehet mérni. Példánkat folytatva tegyük fel, hogy az adatokból megállapítottuk, hogy egyre többen lemondják a feliratkozásukat a zenemegosztó szolgáltatásunkon. Ez lenne maga az üzleti probléma. Az üzleti lehetőség viszont az lenne, hogyha csökkentjük a leiratkozások számát X%-al, akkor a vállalatunknak Y%-al nő a bevétele. Nem elég csak ennek a két ténynek az ismerete, kell egy megoldást is találni, amivel magát az üzleti lehetőséget értékesíteni tudjuk. A példa vetületében most ez a tettlegesség (megoldás), vagyis az üzleti lehetőség értékesítése, hipotetikusan egy olyan modell kidolgozását igényelné, ami bizonyos inputok alapján képes beazonosítani azokat a fogyasztókat, akik valószínűleg le fognak iratkozni. Ezeknek a fogyasztóknak ismeretében majd a későbbiekben egy a fogyasztók megtartására irányuló stratégia keretében képesek lennénk ösztönzőket küldeni, kedvezményt a következő havi díjra, vagy e-mailen keresztül biztatni őket a további használatra.

Az 5. szint az, ahol maga a gépi tanulás először szóba jön. Ez a szint az, ahol maga az adattudomány igazán megmutatja mire is képes, hiszen ez az a szint, ahol maga a vállalat elindul a tettleges megoldás kidolgozása és későbbiekben a modell implementálása felé. Itt jön szóba maga a fent említett modell prototípus megalkotása, a prototípussal való A/B tesztek. A példával élve ezen a szinten elkezdődik az a prediktív modell megtervezése, ami képes a fogyasztók leiratkozási valószínűségének meghatározására, majd a modell prototípus megalkotása és tesztelése A/B tesztek keretében, végül a konklúzió alkotás. Ha a modell sikeresen kiállta a próbát, akkor jön szóba csak a 6. szint. Ez a szint maga a modell alkalmazása. Először kis fogyasztói környezetben, és ha élesben is jól működik akkor majd a későbbiekben való felskálázása. A 7. szint a modell folyamatos megfigyelése, kontrollja, és csiszolása.

3. GÉPI TANULÁSI TECHNIKÁK

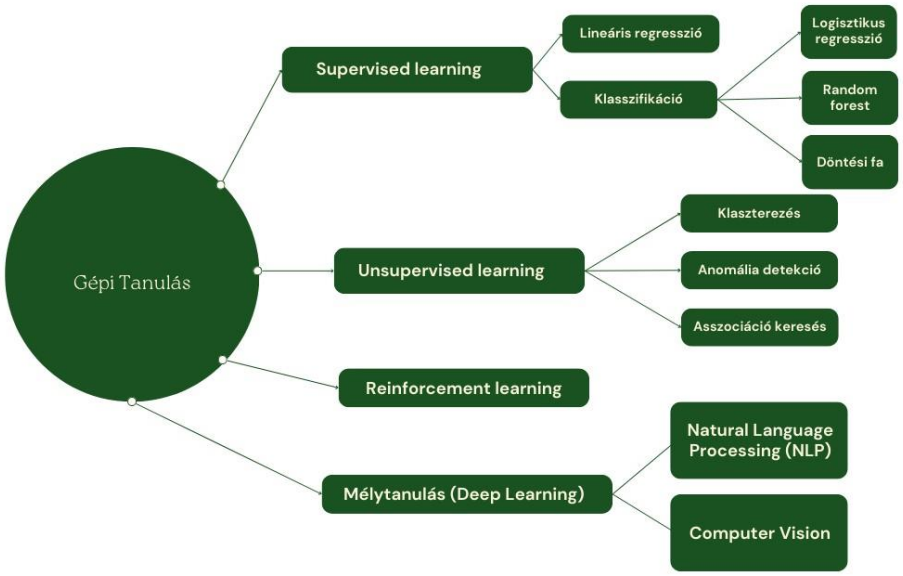
A mesterséges intelligencia egy gyűjtő fogalom, ami olyan programokat és gépeket takar, amik képesek az ember által definiált intelligens viselkedésre. A mesterséges intelligencia egyik válfaja az a gépi tanulás. A lent 3. ábrából is jól látszik az az állítás, hogy minden gépi tanulás mesterséges intelligencia, de nem minden mesterséges intelligencia gépi tanulás, és ez az kifejezés átvihető a mélytanulásra is.



3. ábra: M.I. és stratégiai marketing döntések
Forrás: Kenczler (2020)

A gépi tanuláson belül elkülöníthetünk tradicionális gépi tanulási megoldásokat, illetve mély tanulási megoldásokat. A tradicionális gépi tanulási problémák legtöbbször egyszerűbb algoritmusokat használnak és akkor optimálisabbak hogyha kevesebb adat áll rendelkezésünkre, illetve, ha nagy hangsúlyt akarunk helyezni a modell értelmezhetőségére ennek az árnyoldala pont az, hogy modelljeink kevésbé lehetnek pontosak, mint egy mély tanulási modell. A tradicionális gépi tanulási módszereket a legtöbb esetben strukturált adatra lehet alkalmazni. Ezzel ellentétben a mély tanulás akkor lehet optimális hogyha nagy mennyiségű adat áll rendelkezésre, az adat strukturálatlan és nagyon szeretnénk arra fókuszálni, hogy a modellünk minél pontosabb predikciót alkosson. A mély

tanulási módok sokkal komplexebbek és interpretálásuk sokkal nehezebb, mint a tradicionális gépi tanulási modelleknek ám mint ahogy említettem sokkal pontosabbak tudnak lenni kellő mennyiségű adat mellett.



4. ábra: Gépi tanulási módszerek osztályozása
Forrás: saját szerkesztés

A következőekben célom az, hogy bemutassam a legnépszerűbb gépi tanulási technikákat, röviden jellemezzem azokat és a marketing vonatkozásában történő alkalmazásuknak lehetőségeit bemutassam hipotetikus példákon keresztül, viszont a matematikai jellemzéstől és a reinforcement learning-től eltekintek, illetve az unsupervised learning-en belül csak a klaszterezésről ejtek szót.

3.1 TRADICIONÁLIS GÉPI TANULÁS

A tradicionális gépi tanulás segítségével a már a 2. fejezetben említett mechanikus mesterséges intelligencia feladatai tudjuk elvégezni ezek az algoritmusok segítenek az olyan problémák megoldásában, mint például a klasszifikáció a klaszterezés vagy a regressziós feladatok ellátása. Ezek a megoldások könnyen automatizálhatók és nem igényelnek komplexebb gondolkodó, vagy érző mesterséges intelligencia feladatot.

3.1.1 Supervised learning

A supervised learning a gépi tanulás egyik módszere, lényegében a modelleknek a célja az, hogy a modellekhez használt adathalmaz jellemei alapján segítsenek megbecsülni egy célváltozót.

Ezeket a modelleket sok esetben prediktív modelleknek is szokás hívni, céljuk az, hogy minél pontosabban meg tudják becsülni a célváltozó értékét. Két népszerű fajtája van az egyik a klasszifikáció, a másik a lineáris regresszió.

3.1.1.1 Lineáris regresszió a marketing vonatkozásában

A lineáris regresszió egy olyan gyakran használt technika, amelynek segítségével egy folytonos célváltozó értékének előrejelzése történik egy input (jellem) kombináció alapján. Ebben a módszerben a célváltozó, függ az inputoktól, míg az input változók független változók. A folyamat során egy olyan egyenes vagy felület illesztése történik, amely minimalizálja az előrejelzett és a tényleges kimeneti értékek közötti különbségeket, az egyszerű lineáris regresszió kalkulátorok "legkisebb négyzetek" módszerét alkalmazzák a legjobban illeszkedő vonal vagy felület meghatározásához (IBM, *What is linear regression?,* 2023).

Vegyünk egy példát a marketing vonatkozásában, mondjuk legyen egy adathalmazunk, aminek segítségével, majd azt szeretnénk később megbecsülni, hogy egy új fogyasztó mennyit fog költeni az ittléte során, tehát becslést szeretnénk adni a CLV-re (customer lifetime value), ami az adott fogyasztóhoz illik. Ebben az esetben a célváltozónk a fogyasztó CLV értéke lesz. Ahhoz, hogy ezt az értéket meg tudjuk becsüli a modell felhasználja a mi input tényező szelekciónkat az adathalmazból és azokat inkorporálja a lineáris regresszió modelljébe, tehát súlyokat ad nekik és felállít egy regressziós egyenest. A regressziós egyenes alapján, ha van új input csomagunk, tehát új fogyasztó, meg tudjuk becsülni, hogy mennyi lesz annak az új fogyasztónak a CLV-je, akihez az az adatsor illik. Ez azért nagyon fontos, mert így tudjuk látni, hogy bizonyos fogyasztók körülbelül mennyit fognak költeni, és ez alapján optimalizálni tudjuk a kommunikációt, esetlegesen ösztönözni a fogyasztókat, arra, hogy magasabb CLV-t érjenek el, vagy kiemelkedőbb szolgáltatást biztosítani a magas CLV-vel rendelkező tehát prémium fogyasztók számára.

3.1.1.2. Klasszifikáció a marketingben

A klasszifikációs modellek az adat jellemek alapján segítenek nekünk megbecsülni egy diszkrét változót, tehát megbecsülni egy kategóriát. Ilyen klasszifikációs modellt igényel, például a marketing nézőpontjából a churn modellezés, ami alapján az a célunk, hogy megbecsüljük, vajon egy fogyasztó tovább fogja-e használni szolgáltatásunk, vagy elpártol tőlünk. Az ehhez hasonló feladatok bináris klasszifikációt igényelnek tehát a célváltozó, amit megszeretnénk becsülni az két érték lehet, legtöbbször, 0 vagy 1. Vannak olyan klasszifikációs modellek is, amik képesek több osztályba sorolni, tehát akár a célváltozó több értéket is felvehet ezeket multinomiális klasszifikációnak hívjuk.

Logisztikus regresszió

Ez a supervised learning modell típus (amelyet logit modellnek is neveznek) gyakran használatos klasszifikációra. A logisztikus regresszió becslést ad egy esemény előfordulásának valószínűségéről, például szavazott vagy nem szavazott, egy adott input változókra vonatkozó adatkészlet alapján, mivel az eredmény egy valószínűség, a célváltozó 0 és 1 közötti értéket vehet fel (IBM, *What is logistic regression?,* 2023).

Döntési fa

A döntési fa egy hierarchikus, fa szerkezetű algoritmus, amely klasszifikációs és regressziós feladatokra használható. Az algoritmus egy gyökér csomópontból indul, és az ágak segítségével eljutunk a belső döntési csomópontokhoz, ezek a belső csomópontok kiértékelik a rendelkezésre álló jellemzőket, hogy homogén részhalmazokat határozzanak meg, ezeket a részhalmazokat levél csomópontoknak vagy terminális csomópontoknak nevezzük, amelyek a végső kimeneteket reprezentálják az adathalmazban. (IBM, *What is a decision tree?,* 2023).

Random forest

A random algoritmus több véletlenszerű döntési fát kombinál azok végkimeneteleit átlagolja. Ez klasszifikáció esetében úgy valósul meg hogy a random forest algoritmus végső kimenetele az az osztály lesz amely a random forest algoritmus véletlenszerű döntési fáinak végkimenetele között a legtöbbször szerepel (Biau & Scornet, 2016). A random forest algoritmus jó lehet olyan adathalmazokon való prediktálásra, ahol a változók száma esetlegesen nagyobb mint a megfigyelések száma.

3.1.2 Unsupervised learning

A supervised learning mellett egy másik nagyon népszerű gépi tanulási módszer az unsupervied learning.

"Az unsupervised learning, gépi tanulási algoritmusokat használ fel a címkézetlen adathalmazok (nem tartalmaz megbecsülendő célváltozót) *elemzéséhez és csoportosításához. Ezek az algoritmusok felfedezik a rejtett mintázatokat vagy adatcsoportosításokat. Az információ hasonlóságainak és különbségeinek felfedezése miatt ideális megoldást jelent az adatelemzési kutatásokhoz, keresztezett értékesítési stratégiákhoz, ügyfél szegmentáláshoz és képfelismeréshez"* (IBM, What is unsupervied learning?, 2023).

3.1.2.1 Klaszterezés a marketingben

"Klaszterezés során csoportokba, osztályokba soroljuk az elemeket, tehát osztályozást végzünk. Az eredeti osztályozási feladattól [klasszifikáció] *(lásd előző fejezet) az különbözteti meg a klaszterezést, hogy nincs megadva, hogy melyik elem melyik osztályba tartozik (tehát nincs egy tanító, aki helyes példákkal segíti a tanulásunkat), ezt nekünk kell meghatároznunk"*

(Bodon, 2009, 159).

Marketing szempontból olyan téren releváns a klaszterezés, hogy ezzel a metódussal, nagy mennyiségű input szám alapján tudunk fogyasztói szegmenseket képezni. Ezeket a fogyasztói szegmenseket fel tudjuk használni, arra hogy sokkal jobban megértsük márkánkat, és azt a folyamatot, ahogyan a vállalat képes értéket teremteni különböző embercsoportok számára. A klaszterezést fő 3 csoportba lehet sorolni. Vannak a hierarchikus módszerek, a nem hierarchikus módszerek, és az egyéb módszerek, ide tartozik például a 2 lépcsős klaszterezés (Naresh K. Malhotra & Simon Judit, 2017).

3.2 MÉLY TANULÁS

A mély tanulás egy olyan gépi tanulási technika, amely lehetővé teszi, hogy komplex strukturálatlan adatokat is tudjunk értelmezni, ezt a feladatot a mély tanulás neurális hálók segítségével próbálja véghez vinni (Vargas, Mosavi, & Ruiz, 2017). A neurális hálók az emberi agy működését próbálják replikálni és próbálnak segíteni a számítógépnek abban, hogy olyan komplex feladatokat el tudjon látni, mint a nyelvi feldolgozás vagy a látás.

Esetlegesen mély tanulással ellátható feladatok közé tartozik:

1. Képfelismerés: Arc és objektum felismerés
2. Hangfeldolgozás: Beszédfelismerés
3. Nyelvfeldolgozás: Automatikus fordítás, szöveg generálás, nyelvi értelmezés

A fent említett példák mind a gondolkodó és mind az érző mesterséges intelligencia feladatait jól el tudják látni ezzel élénkítve a fogyasztó élményét. Esetlegesen ilyen lehet egy olyan megoldás, ami képes egy vállalat weboldalát különböző nyelvekre fordítani az olvasó igénye szerint (Perkovitz & Etzioni, 1997). Ezenkívül esetlegesen még ilyen megoldás lehet a fogyasztói élmény fokozására a marketing szempontjából a különböző külső tényezők beépítése az élmény folyamatba, ebben tud segíteni a kép felismerés vagy a hang feldolgozás. Ilyen megoldás lehet például egy applikáció, ami hangvezérlést alkalmaz (Zhu & Wu, 2020).

4. PRIMER KUTATÁS

A primer kutatás olyan tevékenység, amely során közvetlenül gyűjt a kutató új és az adott problémához kapcsolódó eredeti adatokat egy adott témával vagy kérdéssel kapcsolatban (Naresh K. Malhotra & Simon Judit, 2017). A következő alfejezetekben a dolgozatom primer kutatását, mint egy esettanulmányt fogom bemutatni, amely során egy iparágra fókuszálva vizsgálom a gépi tanulási lehetőségeket a KKV-k számára.

4.1 MÓDSZERTANI BEMUTATÁS

4.1.1 Vezetői döntési probléma bemutatása

A kutatás vezetői döntési problémája az, hogy megéri-e egy KKV szempontjából alkalmazni mesterséges intelligenciát és ha igen, akkor hogyan tudják ezt megvalósítani? Ez párhuzamba hozható három vállalati szemponttal is: Képes-e a vállalat felismerni különböző üzleti problémákat? Ezekre képes-e a vállalt realizálni egy üzleti lehetőséget M.I. alkalmazása során? Valamint milyen válasz reakciót képes kidolgozni? Ennek megválaszolása végett számos olyan rész kérdést vizsgálok, amelyek egy esettanulmány keretében igyekeznek támpontokat adni. A részkérdések vizsgálata során kitérek a modellépítés részelemeire és hogy mi miért fontos annak megválaszolásában, hogy milyen körülmények kellenek egy KKV számára ahhoz, hogy predikciós modelleket tudjon felállítani.

Már a primer kutatás elején azonban ki kell hangsúlyozni, hogy egy prediktív modell megalkotása nem feltétlenül egyszerű feladat, a vállalatnak szükséges a 2. ábrában említett M.I. szükséglet piramisban jól teljesíteni ahhoz, hogy effektíven tudja alkalmazni a mesterséges intelligenciát. Ehhez nem feltétlenül szükséges programozói tudás, ám fontos lehet egy adatorientált gondolkodás, illetve szükséges hozzá egy mélyebb statisztikai tudás is. Ha azonban képes a vállalat teljesíteni ezen kritériumokat, akkor hosszútávon a jobb predikciók következtében egy erősebb piaci pozíciót tud betölteni, jobb fejlődési lehetőségekkel. A következő oldalakon az előzőekben felsoroltak meglétét feltételezve igyekszem végigvezetni az olvasót a programozási tudás nélküli modellezés lehetőségein és néhány optimalizálási kérdést is feszegetve, amely egyúttal segíthet egy KKV-nak például abban is, hogy mire érdemes odafigyelni és melyek a kevésbé releváns tényezők.

4.1.2 Iparág kiválasztásának indoklása

A vezetői döntési probléma megválaszolásához az esettanulmányomban egy iparágra fogok fókuszálni, amely a magyar társasjátékok piaca lesz, ezen belül is a gépi tanulási előrejelzéseket a társasjáték kávézókra fogom majd elkészíteni. Mindez az iparági szűkítés a kutatás feltáró jellege

miatt elfogadható, azonban fontos felhívni a figyelmet arra, hogy a kapott eredmények kevésbé általánosíthatóak. Ez a kutatás ezen fáziságan azért nem okoz különösebb problémát, mert a cél inkább a módszer használatának a bemutatása és a használat jelentőségének az alátámasztása, nem pedig a konkrét eredményekből való javaslatok megfogalmazása, ezáltal a kapott eredmények általánosíthatóságára kevesebb hangsúly helyeződik. Bár az iparág kevésbé jelentős, de fontosnak érzem megindokolni, hogy miért érdemes például a hazai társasjátékok élménytérrel foglalkozni.

A magyar társasjáték piac rendkívül élénk és fejlődő terület, ahol a szórakozás és közösségépítés egyre népszerűbb. A piac növekedésének okai között találhatjuk a következőket:

A magyar társasjáték piac az utóbbi években jelentős növekedést mutatott. Azt is lehet mondani, hogy a covid utáni időszaknak köszönhetően, egy új reneszánszba lépett. A lockdown általi bezártság által egyre több ember fedezett fel örömet és közösséget a társasjátékokban, és a fiatalabb generációk is egyre inkább vonzódnak hozzájuk. A piacon sok hazai társasjátékgyártó működik, akik saját tervezésű játékokat kínálnak a vásárlóknak, ilyen például a GémKlub vagy a Piatnik (Balla, 2018). Emellett a nemzetközi, külföldi gyártók is széles választékot kínálnak a magyar vásárlók számára, így a játékosok számára hatalmas választék áll rendelkezésre. A másodlagos (használt társasjáték) piac is egész sokoldalú. Ez a paic legfőbbképpen Facebook csoportokból és facebook marketpalce hirdetésekből tevődik össze. A magyar társasjáték piac rendkívül sokoldalú. Az elérhető játékok között megtalálhatóak a különböző játékstílusok, beleértve a stratégiai, családi, kvíz, kártya és kooperatív játékokat is. Ez lehetővé teszi, hogy az összes korosztály és játékstílus számára találjunk megfelelő játékot. A társasjátékok beszerzése számos üzletben lehetséges, ugyanakkor az online vásárlás is népszerű. Az internetes piacterek széles választékot kínálnak, és lehetővé teszik a vásárlók számára, hogy könnyen hozzáférjenek a legújabb társasjátékokhoz. A magyar társasjáték közösség rendkívül aktív és élénk. Számos rendezvény és találkozó szerveződik, ahol játékosok találkozhatnak, új játékokat próbálhatnak ki, és barátságokat köthetnek a szenvedélyüket osztó emberekkel. Ezen kívül az online térben számos társasjátékkal foglalkozó blog és Facebook csoport található, ahol a felhasználók meg tudják osztani társasjátékos élményeiket, és társasjátékokkal kapcsolatos tapasztalataikat, ezen kívül tudnak segíteni kezdő társasjátékozóknak.

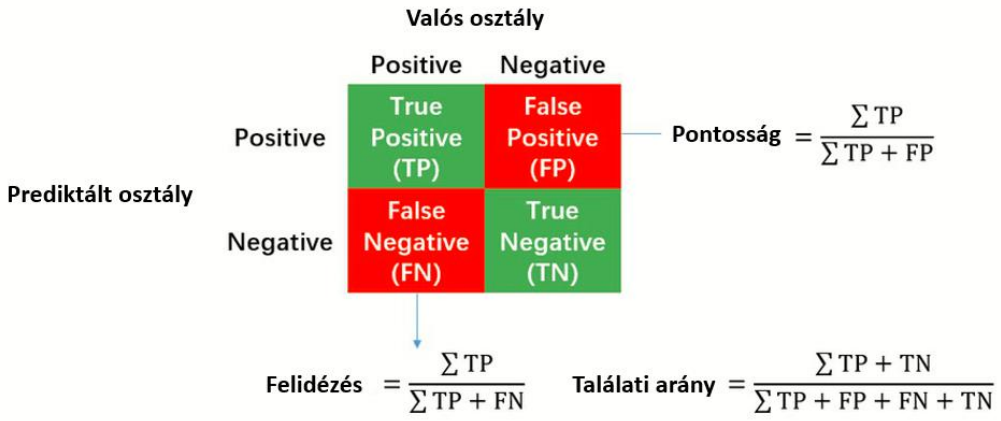
Ezen szereplők és tényezők együttesen formálják a magyar társasjáték piacot, és lehetővé teszik a játékosok számára, hogy együtt szórakozzanak, tanuljanak és élvezhessék az egyre szélesedő játékkínálat előnyeit. Az interakció és közösségteremtés hatalmas szerepet játszik ezen a dinamikusan fejlődő piacon.

A népszerűsödés hatására Magyarországon főleg Budapesten megjelentek a társasjáték kávézók is. A társasjátékos kávézók olyan kávézók/bárok, ahol a rendes bárokban és kávézókban megtalálható szolgáltatások (alkoholos italok, rágcsálni valók, kávék, üdítők) fogyasztása mellett szabadon kipróbálhatod a kávézó készletében megtalálható társasjátékokat. Általában két felmerülő kiadás van az asztalfoglalás díja és maga az ételfogyasztás értéke. Számos ilyen kávézó található Budapesten, ilyen például a Board Game Café, a Game Up PUB és a game UP 2, A játszma társasjáték kávézó, a Playstation Budapest, illetve a BarCraft (Költő, 2023). Ezek inkább egy élmény orientáltabb módot adnak a társasjátékok fogyasztására, és mint, ahogy látszik keverednek benne a kávézói szolgáltatások és a társasjáték kultúra elemei is.

4.1.3 Szűken vett kutatási kérdések bemutatása

A vezetői döntési probléma értelmezéséhez és mélyebb megvizsgálásához a primer kutatáson belül kettő dimenziót vizsgáltam. A következőkben szeretném bemutatni ezen dimenziókat, valamint azt, hogy milyen kérdésekre kerestem a válaszokat. Vizsgálataim fókuszában a kiválasztható adathalmaz (minta és hozzákapcsolódóan a mintavételi keret), a lehetséges modellek választása, valamint a modellek teljesítményét értékelő mérőszám meghatározása állt. Ezutóbbinál többféle módszer is lehetséges (a pontosság, a felidézés és a találati arány vizsgálata) , azonban jelen kutatásban én csak a találati arányra optimalizált verziót fogom bemutatni, mert a vizsgálataim és az elemzések során az derült ki a számomra, hogy a másik kettő esetében a modellekbe bevont dimenziók számán és összetételén is szükséges lehet változtatni ahhoz, hogy konzisztens eredményt kapjunk, így az egyes mutatószámok összehasonlítását egy későbbi kutatási fázisra halasztottam, amely során több változóval és több adattal is tudok majd remélhetőleg dolgozni (Farkas, 2023).

A modellek értékelése a gépi tanulás esetén a sokszor használt és jól bevált, úgynevezett konfúziós, vagy tévesztési mátrix segítségével megy végbe. A mérőszám, amit a dolgozatban vizsgálok az a ***találati arány*** (accuracy), ez ugyancsak ebből a mátrixból vezethető le. Ezt mutatja az 5. ábra is.



5. ábra: A konfúziós mátrix és az abból származtatható mérőszámok
Forrás: Ma et al (2020) alapján saját szerkesztés

Az 5. ábrán láthatjuk, hogy a tévesztési mátrixból mind a három említett mérőszámot meg lehet határozni, az alapján, hogy a valóságban mit tudunk az egyes esetekről és ez hogyan viszonyul az előrejelzett, prediktált értékhez. Például TP, azaz true positive esetről akkor beszélhetünk, ha a valóságban is pozitív az adott változó és az előrejelzés is pozitív eredményt adott számunkra. A későbbiekben ilyen lesz a vizsgálataimban például az, amikor egy fogyasztóról tudjuk, hogy elmenne egy társasjátékos kávézóba és az algoritmus is azt jelzi előre, hogy el fog menni a fogyasztó. A TN ehhez hasonlóan ugyancsak azt mutatja, hogy jól tippelt az algoritmus, azt az értéket kaptuk, amit el is vártunk. A három mérőszám tulajdonképpen az egyes cellákba kerülő értékek viszonyát, arányát mutatja meg. Amíg a pontosság és a felidézés tulajdonképpen csak a pozitív jó tippelésekkel foglalkozik és azokat arányosítja az összes valóban pozitív esethez vagy az összes pozitívnak jelölt esethez, addig a találati arány azt nézi meg, hogy az algoritmus hány esetben tippelt jól az összes esethez képest, ezutóbbi tehát egy sokkal átfogóbb képet ad. Ezért is választottam ezt a mutatót a vizsgálataimra. Természetesen, ahogy említettem az optimalizálást a későbbi kutatásaimban érdemes lesz a többi mutatószámra fókuszálva is elkészíteni és az összevetést ilyen irányból is megtenni.

Az első vizsgált dimenzió - választás a lehetséges modellek közül

Az első dimenzió célja, hogy a már fent említett mérőszámot különböző modell fajtában hasonlítsam össze. Ebben a dimenzióban 2 modellt hasonlítok össze a ***logisztikus regressziót, illetve a random forest algoritmust.*** (A két módszer rövid leírását a 3.1. fejezetben olvashatjuk részletesen.) Ezt a 2 algoritmust vizsgálom a találati arány vetületében és próbálom összehasonlítani, hogy az adott optimalizált mérőszám mellett melyik lehet az ideális választás egy KKV számára, melyik tud jobb eredményt hozni az előrejelzésekben.

A kutatási kérdés:

Adott mérőszám és adathalmaz mellett melyik lehet az optimális választás 2 modell közül a predikciók megvalósítására?

Második dimenzió - a különböző adathalmazok vegyes használatára vonatkozóan

A második dimenzióban arra fókuszálok, hogy a fixen meghatározott modell és a mérőszám mellett milyen eltéréseket tudunk tapasztalni, ha különböző adatgyűjtési módszereket alkalmazunk. A vizsgálat lényege az lesz, hogy a tanuló algoritmusokat egyik verzióban ugyanazon az adatbázison (annak egy részén) tanítom be, mint amin tesztelem is, a másik verzióban pedig eltérő tanuló és tesztelő adatbázist használok (ez különböző mintavételi módszert és mintavételi keretet jelent az én esetemben). Az adatgyűjtés folyamatát és a kétféle adatbázis kialakítását a következő fejezetekben fogom részletesen is bemutatni.

A kutatási kérdés:

Tapasztalhatunk-e számottevő eltérést a különböző adatbázisok használata esetén? Mennyire fontos egy KKV számára, hogy a megfelelő mintavételi módszert és mintavételi keretet alkalmazza a predikciók megvalósításához?

Kutatómunkám során ezekre a kérdésekre próbálok reflektálni a társasjáték piacon belül egy predikciós helyzetre vonatkozóan, amely során a társasjátékos kávézókba való elmeneteli hajlandóságot próbálom előre jelezni.

4.1.4 Tradicionális vs. modern marketingkutatás

A kutatásom során a trandicionális és modern marketingkutatási eszközöket vegyesen alkalmazom. Az adatgyűjtéshez már régóta használt, klasszikus módszerekkel dolgozom, míg az adatfeldolgozás során fogom a mesterséges intelligenciát, a tanuló algoritmusokat hasznosítani. Természetesen az adatgyűjtésben is lehetséges a mesterséges intelligencia alkalmazása például weboldalakról, közösségi oldalakról való szisztematikus adatgyűjtés keretében. A jelen kutatásban azonban az volt a célom, hogy egy könnyen elkészíthető adatbázison tudjak példát mutatni a modern adatfeldolgozás használatára.

4.1.4.1 Tradicionális marketingkutatás

A kvalitatív kutatás egy strukturálatlan feltáró módszer, amelynek fő célja egy probléma megértése, ezzel párhuzamosan a kvalitatív kutatások nem nagyon foglalkoznak hipotézisek vagy előzetes feltételezések megfogalmazásával és tesztelésével, hanem a problémát a kutatási feltevések

szintjén próbálják megérteni (Kiss, 2016). A kvalitatív kutatások nem a számszerűsíthető kérdésekre fókuszálnak, hanem inkább a miért, hogyan és hányféle típusú kérdésekre keresnek választ. A kvalitatív kutatás fő célja a mélyebb okok feltárása viselkedési mintázatok kimenetek és motivációk megértése. Dolgozatomban a társasjáték piacának ilyen irányú felmérésével nem foglalkoztam, de természetesen lehetséges ennek is a beépítése a folyamatba. Az adatbázis kialakításához, a modellbe bevonni kíván változók kiválasztásához az ilyen jellegű kutatások is fontos inputként tudnak szolgálni.

Ezzel ellentétben a kvantitatív kutatások során számszerűsíteni próbáljuk a különböző viselkedési, motivációs, és más attitűdöt jellemző tényezőket, aminek végső célja az, hogy statisztikai következtetéseket tudjunk levonni. Hagyományosan ebben az esetben leíró statisztikai elemeket és ok-okozati összefüggéseket tudunk vizsgálni (Gyulavári et al., 2017).

4.1.4.2 Modern marketingkutatás

Az elmúlt évtized technológiai fejlődése számos formában hatással volt a fogyasztói viselkedésre, illetve a piackutatás fejlődésére is. Ilyen változás például az IoT megjelenése. Az okos eszközök hálózata lehetőséget ad a tömeges adatgyűjtésre a felhasználókról, amiket később az adatelemzés során prediktív analitika keretében is fel tudunk használni, ami képes segíteni különféle marketing folyamatok automatizálásában (Szűcs et al., 2023).

Egy másik erős motorja a marketingkutatás rohamos átalakulásának az a mesterséges intelligencia. A mesterséges intelligencia számos marketing folyamatot képes automatizálni, így felgyorsítva a marketing tevékenységet és a pontos és releváns információk kinyerését. Az öntanuló algoritmusok lehetőséget biztosítanak a személyre szabására is. A belső adatforrások szerepe is ezáltal megnőtt a kutatásokban, amelyek a vállalaton belül összegyűlt jelentős mennyiségű adatot tartalmazzák a fogyasztókról. Az adatbányászat kulcsfontosságú szerepet ölt magára ezen nagy mennyiségű adatok elemzésében és értelmezésében. A mesterséges intelligencia, bár kvantitatív alapokon nyugszik, de számos kvalitatív módszert is képes támogatni. A gondolkodó mesterséges intelligencia segíthet a kvalitatív és strukturálatlan információk számszerűsítésében, így egy átmenetet és köztes állapotot biztosítva a kvalitatív és kvantitatív kutatások között (Wulff et al., 2022).

4.1.5 Adatgyűjtés

Kutatásomban a tradicionális és a modern marketingkutatást vegyítettem. A tradicionális marketingkutatás folyamata az az adatgyűjtésben nyilvánult meg, míg a modern marketingkutatás az adatfeldolgozásban, a prediktív modell készítésében. Az adatbázis összeállítása kettő módon zajlott.

Az A adatbázist úgy állítottuk föl, hogy megkérdeztünk számos hallgatót a BME-n belül, akik önkényesen tudták kitölteni a kérdőívet, tehát saját maguk tudták eldönteni, hogy kitöltik-e vagy sem (Gyulavári et al., 2017). A B adatbázis felállítása úgy zajlott, hogy a kérdőív meg lett osztva több közösségi média platformon (Facebook, Instagram), ezen belül társasjátékkal foglalkozó csoportokban, illetve baráti közösségekben tehát egy hólabda elv alapján építettem fel a B adatbázist (Malhotra - Simon, 2017).

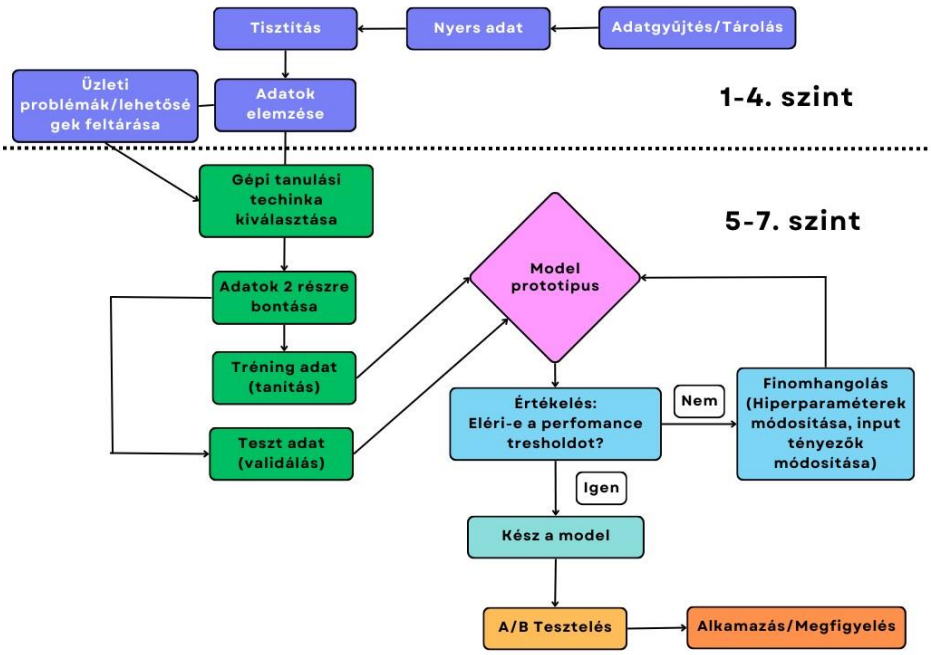
Az adatbázisban lévő változókat a Törőcsik, M. (2016). „*FOGYASZTÓI MAGATARTÁS (Insight, trendek, vásárlók)*" által leírt szcéna modell alapján állítottam össze. A szcénák olyan közösségek vagy csoportok, amelyek különböző élményekben vagy értékekben találnak közös alapot (Törőcsik, 2016). Így számos olyan társasjátékkal kapcsolatos kérdést tettem föl, ami bele élhet a társasjátékos szcénába, ilyen lehet például az egyén társasjátékozás közegére irányuló kérdések, mint például az, hogy kivel szokott társasozni, milyen gyakran társasozik, vagy hogyan és milyen módon követi társasjátékhoz köthető tartalmakat a közösségi média platformjain. A másik nagyobb tömb, ami a társasjátékos szcénára irányult az a társasjátékokkal kapcsolatos preferenciák, az erre irányuló kérdések magukba foglalták például azt, hogy mi az, ami fontos az egyénnek egy társasjátékkal kapcsolatban, milyen társasjáték zsánereket szeret, milyen társasjáték stílusokat szeret és ezekhez hasonló további kérdések. A kérdőív során a társasjátékozással kapcsolatos kérdéseken kívül feltettem az egyén közösségi média használatára, illetve általános társas közegben való viselkedésére irányuló kérdéseket. A közösségi média használatra irányuló kérdések magukba foglalták azt, hogy az egyén milyen gyakran használ milyen közösségi média szolgáltatásokat, hol osztja meg élményeit és hogyan, illetve milyen közösségi média forrásokat használ fel programjai megtervezésére. A társas közegben való viselkedésre irányuló kérdésre magukba foglalták azt, hogy az egyén hogyan viselkedik a baráti közösségen belül lesz az erre irányuló kérdések magukba foglalták azt, hogy Az egyén hogyan is oszt meg számára fontos dolgokat a baráti közegben, milyen gyakran teszi ezt, hogyan érzi magát abban baráti közegében, mennyire befogadó a baráti közege és ehhez hasonló kérdések.

A kérdőívekkel felállított adatbázis összeállításához a limesurvey felületét használtam.

4.1.6 Modellépítési folyamat

Ebben az alfejezetben szeretnék kitérni arra, hogy gépi tanulás szempontjából pontosan hogyan is közelíthető meg a modellalkotási folyamat a már 2.4-es alfejezetben bemutatott szükséglet piramis segítségével. A 6. ábra segítségével szeretném bemutatni a modellalkotási folyamatot, illetve

megmutatni azt, hogy milyen párhuzamok húzhatók a modell alkotási folyamat és az adatszükséglet piramis között.



6. ábra: Gépi tanulási modell alkotási folyamat jellemzése
Forrás: saját szerkesztés

Mint ahogy az 6. ábrából is látható 2 különböző részre tudjuk osztani a modell alkotási folyamatot. Az 1. rész, ami az adatszükséglet piramis 1-4. szintig tartó részeit foglalja magába ez még nem teljesen a modell alkotással foglalkozik. Ebben a folyamatrészben kell kialakítani azokat a változókat és teljesítménymutatókat, amik az adatelemzéséhez szükségesek. Az adatelemzés során, mint ahogy már az adatszükséglet piramisban is említettem, fontos feltárni az adott üzleti problémákat és azokhoz üzleti lehetőséget társítani. Ilyen üzleti lehetőség lehet akár egy általános klasszifikációs probléma, ahol el szeretnénk különíteni a számunkra értékes fogyasztókat az olyan fogyasztóktól, akik nem hoznak akkora értéket a vállalat számára. Az előbbi egy mechanikus mesterséges intelligencia folyamat lenne, de akár olyan gondolkodó és érző mesterséges intelligencia folyamatok integrálása is szóba kerülhet, amik mély tanulást igényelnek.

Így jutunk át az 5. szintre, ahol elindulhat a modell kiválasztása és majd az üzleti lehetőség vállalati mechanizmusokba való integrálása. Fontos az adathalmaz 2 részre választása. Az 1. adatrész a 6. ábrából is látszik az a modell taníttatására van, ebből az adatrészből fogja megtanulni a modell az adathalmazban lévő logikai összefüggéseket. Miután a modell megtanulta az adathalmazban lévő logikai összefüggéseket azután használjuk a kettéválasztott adathalmaz másik részét, vagyis a teszt adatrészt, ami a modell teljesítményének értékelésére szolgál. A teszt adatrész azért fontos mert ez

egy olyan adatrész, amit a modell még nem látott és így egy szimulált, de a valósághoz hasonló környezetben tudjuk tesztelni a modell teljesítményét. Miután a modell tesztelése megtörtént, azután jön a modell teljesítményének értékelése, ha modell eléri az általunk kiszabott teljesítmény határt akkor elfogadjuk a modellt, mint egy valid és alkalmas modell, ha pedig nem, mint ahogy az ábra is mutatja különböző technikákkal módosítjuk modellünket, hogy jobb teljesítményt tudjon elérni. Miután elfogadtuk modellünket elkezdhetjük először kis környezetben, de valós fogyasztói csoporton alkalmazni. A modell teljesítményét ebben az új valós környezetben A/B teszteken keresztül értékeljük. Ha modell tényleg jelentős különbséget tud elérni abban a feladatban, amit az üzleti lehetőség formájában realizáltunk akkor lép életbe a modell folyamatos alkalmazásra. Viszont hiába lép életbe a modell folyamatos alkalmazásra, folyamatosan meg kell figyelnünk azt és korrigálni hibáit hosszútávon.

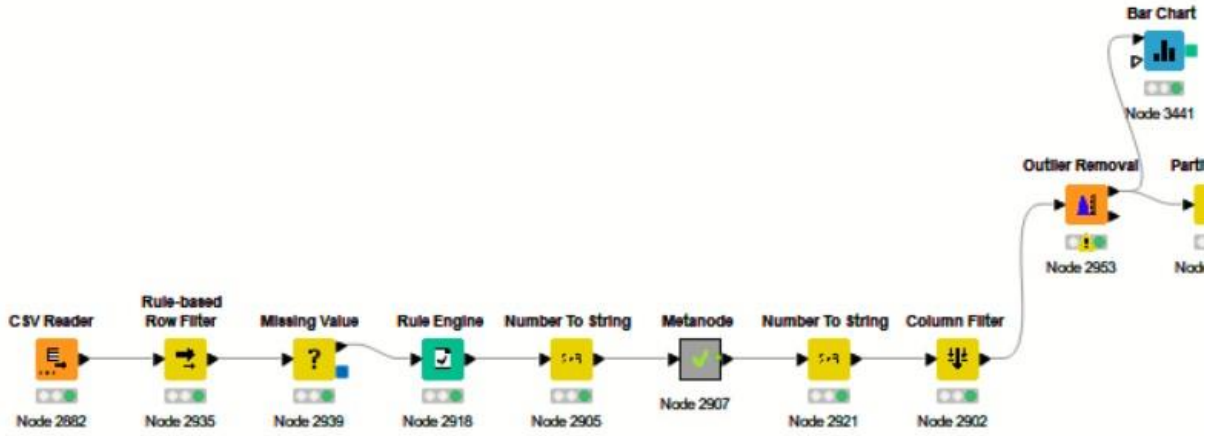
4.1.7 KNIME analytics rövid bemutatása

A kutatáshoz a KNIME Analytics Platformot használtam. A KNIME Analytics Platform egy ingyenes, teljeskörű adatelemzési és gépi tanulási eszköz, amely egy könnyedén használható vizuális munkafelülettel rendelkezik (Gulyás & Illés, 2019). A platform könnyen értelmezhető és jól átlátható vizuális felülete végett jó megoldás lehet marketingkutatási feladatok ellátására és akár gépi tanulási, és prediktív elemzési projektek előállítására. A platformon a lépések könnyedén tervezhetőek és egy úgynevezett Drag&Drop módszerrel lehet összeállítani a különféle adattisztítással, gépi tanulással és prediktív elemzéssel kapcsolatos rendszereket. A tervezési folyamatot node-ok vagyis kisebb feladatokat ellátó egységeken keresztül tudjuk irányítani, moderálni. A platform lehetővé teszi kis és nagy méretű adatok könnyed importálását és előkészítését, és eszközei sok segítséget nyújtanak az adatok gyors és precíz formázásában, így ez akár tudja javítani a marketingkutatások teljesítményét és gyorsaságát, ami a mai világban elengedhetetlen, ha versenyben szeretnénk maradni. Nem csak az adatfeldolgozásban nyújt nagy segítséget, hanem az adatok vizualizálásában, statisztikai interpretálásában, jelentések készítésében, és eredmények értelmezésében is számos lehetőséget biztosít. Ezek mellett a KNIME kiváló lehetőséget ad a gépi tanulási modellek összeállításában és ezek prediktív elemzésre való felhasználásában, ilyen lehetőség lehet például egy olyan modell készítése, ami segít a fogyasztói magatartás előrejelzésében vagy különböző szegmentációs megoldások véghezvitele. Ezenfelül a KNIME egy nagy és aktív közösséggel rendelkezik, ahol a felhasználók megosztják tapasztalataikat, akár más felhasználóktól segítséget is tudnak kérni bizonyos problémák megoldásában. A KNIME azért lehet sok marketing kutató számára jobb megoldás, mint a tradicionális programozási nyelvek, amiket a gépi tanulásra használnak, ilyen például a Python vagy az R, mivel ezek a nyelvek a KNIME-hoz képest jelentősen magasabb szintű

programozási tudást igényelnek és a marketing kutatóknak szilárd alapokkal kell rendelkezniük, hogy effektíven fel tudják használni az adott programozási nyelv adta lehetőségeket. Meglátásom szerint a KNIME ideális választás a marketingkutatási célokra, modell építésre és prediktív elemzésre, mivel lehetővé teszi a marketing kutatók számára az adatok hatékony feldolgozását és elemzését egy vizuális környezetben, miközben a programozási igényeket minimalizálja és csökkenti a szaktudással rendelkezők belépési korlátját. Ezáltal időt és erőforrásokat lehet megtakarítani.

4.2 ADATTISZTÍTÁSI ÉS ELEMZÉSI FOLYAMAT

Az adattisztítási és elemzési folyamat egy nagyon kritikus pontja a gépi tanulási folyamatnak. Nem megfelelő adattisztítással, és pontatlan adatelemzéssel rossz konklúziókhoz érhetünk az adathalmazzal kapcsolatban. Ezért kutatásomban én is megpróbáltam kellő odafigyeléssel és szakmaisággal elvégezni ezt a folyamatot. Mivel kutatásomban egy prediktív elemzési folyamatot szerettem volna bemutatni ezért az adatokelemzésétől a későbbiekben eltekintek mivel az üzleti kérdés, és az adott üzleti lehetőség már előre meg van fogalmazva. A későbbiekben az adattisztítási és elemzési folyamatot a 7. ábra segítségével szeretném leírni, illetve a KNIME munkafelületen belül előállított node folyamat segítségével, ami a 8. ábrán látható.

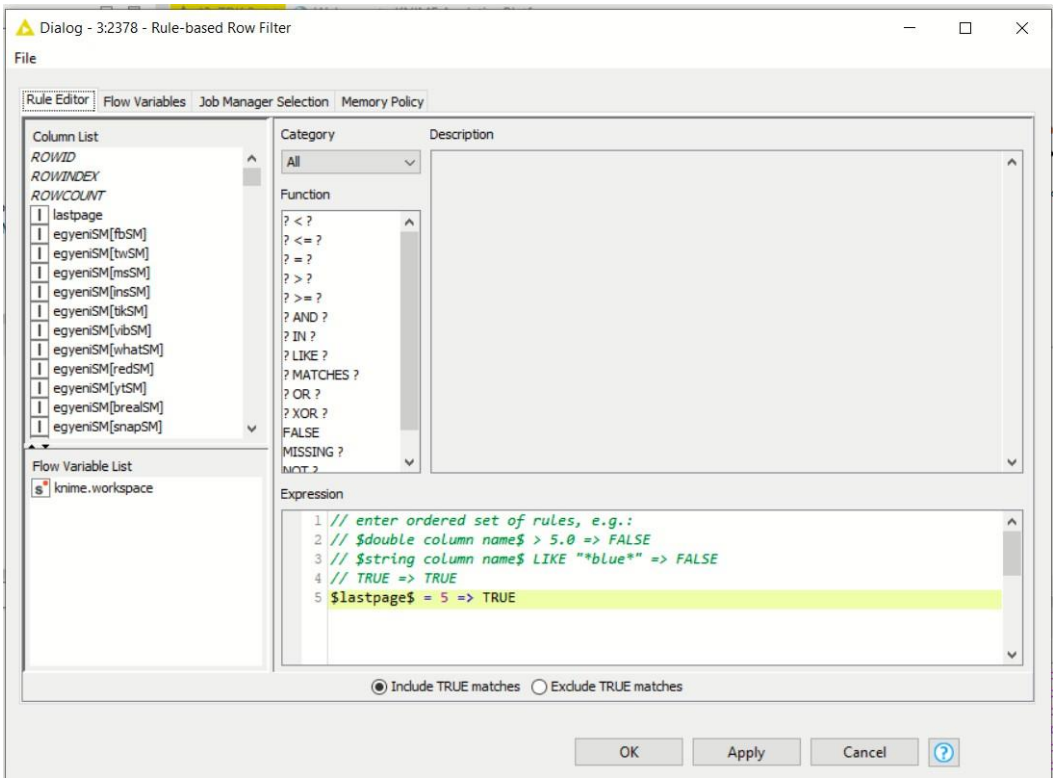


7. ábra: Adattisztítási és elemzési folyamat
Forrás: KNIME szoftver, saját eredmények

Az adattisztítási folyamat minden elkészített modellben hasonló formában jelenik majd meg ezáltal is garantálva a modellek összehasonlíthatóságát. 1. lépésként importáltam az adatokat a kérdőívből, és a kérdőívben szereplő felesleges értékeket eltávolítottam, ilyen volt például az, hogy adott oldalon milyen gyorsan haladt végig a kitöltő.

4.2.1 Tényleges kitöltők kiszűrése

A kérdőív által generált lastpage, tehát utolsó oldal változót azért hagytam bent, hogy a későbbiekben ki tudjam szűrni azont a kérdőív kitöltőit, akik befejezték a kérdőívet. Erre a feladatra szolgált a 8. ábrán látható Rule-Based Filtering.



8. ábra: Tényleges kitöltők kiszűrése
Forrás: KNIME szoftver, saját eredmények

A 8. ábrán látható, hogy próbáltam azokat a személyeket kiszűrni, akik eljutottak a kérdőív 5. oldalára, így eltávolítva azokat, akik elkezdték a kérdőívet, de nem fejezték be. A lastpage értéke, csak akkor lehetett 5 ha a kitöltő befejezte a legutolsó oldal kitöltését is. Ezután, mivel a kérdőív tartalmazott feltételes kérdéseket, ezért néhány kitöltő bizonyos változóval kapcsolatos oszlopában missing value, azaz hiányzó érték volt található. Ez a szűrés csökkentette az adatbázis méretét, hiszen kiestek belőle azok a kitöltők, akik nem fejezték be a kérdőívet, azonban ez sokat segített később a hiányzó adatok pótlásában, amely rendkívül fontos a modell szempontjából. A későbbiekben lehetségesnek tartom azonban azt, hogy a részleges kitöltésekkel együtt is vizsgáljunk modelleket.

4.2.2 Hiányzó értékek korrigálása

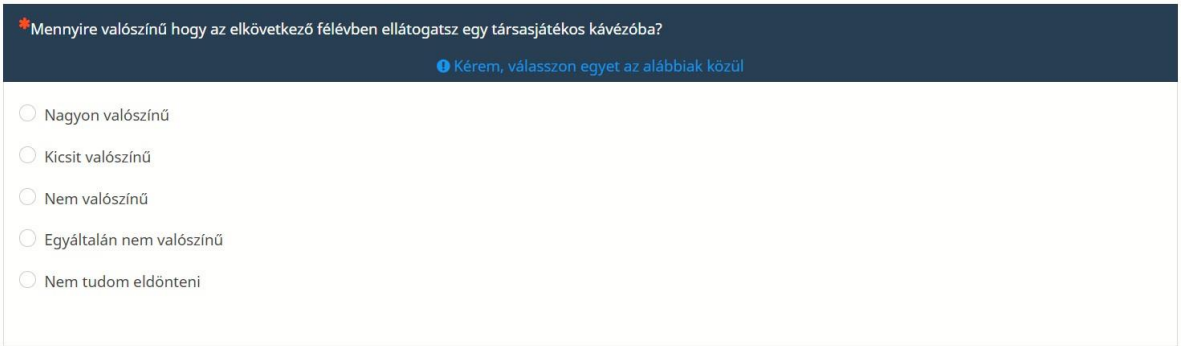
A hiányzó értékek korrigálására használtam a Missing Value node-ot. Mivel feltételes kérdés volt kérdőívekben és azok számára, akik egy bizonyos közösségi média szolgáltatás nem használnak elég gyakran azoknak ez a kérdés nem merült fel a kérdőív során, így a kérdőív a kérdéses oszlopokba bekerülő értéküket missing value, azaz hiányzó értékként tüntette fel. Ennek korrigálására az érintett

oszlopokban a minimális értéket tüntettem föl ezeknél a kitöltőknél. Ez a beállítás a 9. ábrából is látható.

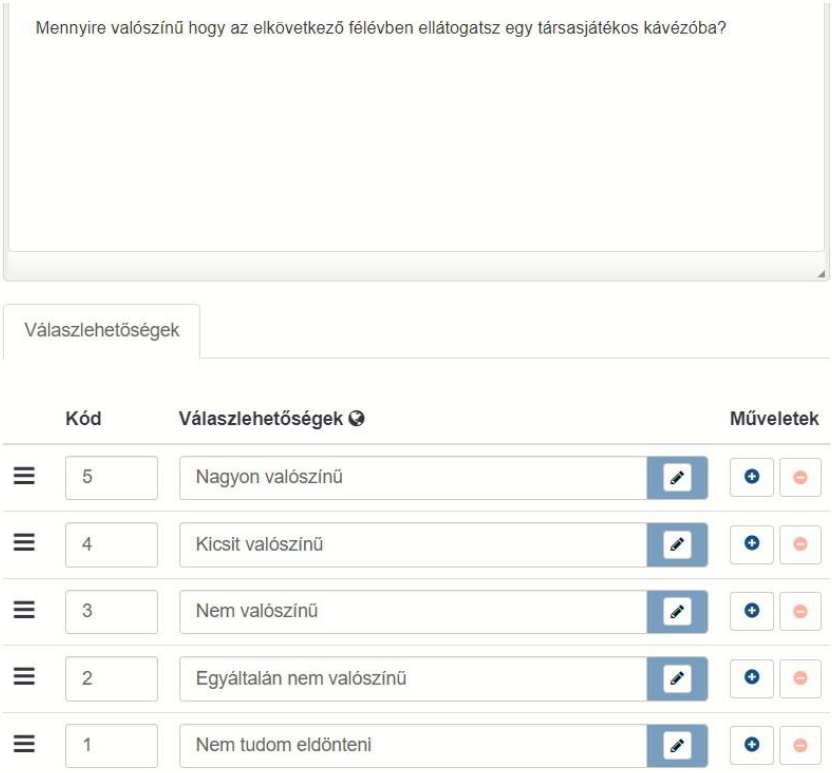
9. ábra: Hiányzó értékek kezelése
Forrás: KNIME szoftver, saját eredmények

4.2.3 Klasszifikációs célváltozó előállítása

A klasszifikációs célváltozót a boardELMEGY változó újrakódolásával hoztam létre. Az eredeti változó, arra a kérdésre adott választ, hogy „Mennyire valószínű, hogy az elkövetkező félévben ellátogatsz egy társasjátékos kávézóba?”. A válaszlehetőségek 5 kategóriát tartalmaztak (Nem tudom eldönteni, Egyáltalán nem valószínű, Nem valószínű, Kicsit valószínű, Nagyon valószínű). A kitöltő számára a kérdés a 10. ábrán látható formában volt bemutatva, illetve a táblázat eredeti (nem átkódolt boardELMEGY) értékei a 11. ábrán láthatóak.

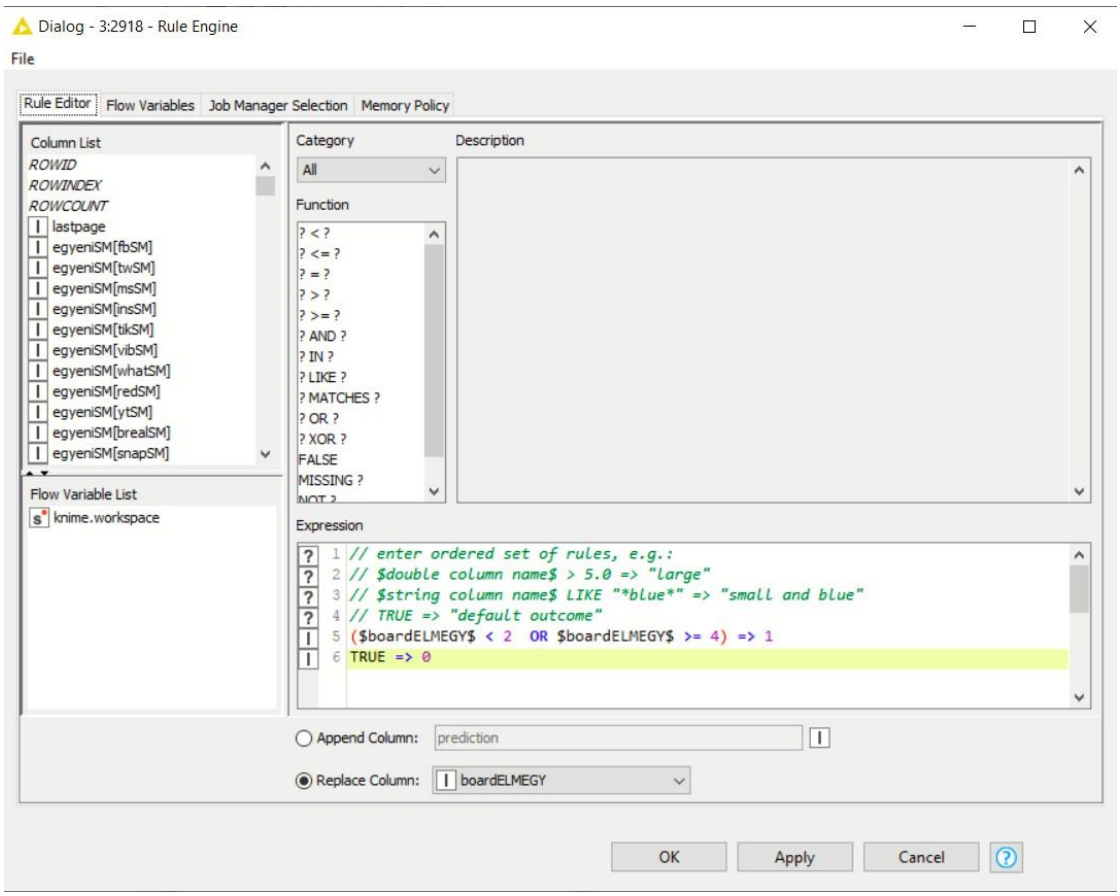


10. ábra: Célváltozót előkészítő kérdés
Forrás: KNIME szoftver, saját eredmények



11. ábra: Célváltozót előkészítő kérdés táblázatba átvitt értékei
Forrás: KNIME szoftver, saját eredmények

A klasszifikációs változó előállításánál, azt a szempontot tartottam szem előtt, hogy egy társasjátékos kávézó tulajdonosának a szemszögéből nézve ki szeretném szűrni a későbbiekben azokat a potenciális fogyasztókat, akikre érdemes erőforrást fordítani, mivel nagy valószínűséggel fogyasztókká válnak. A változó előállítását a 12. ábra mutatja be.

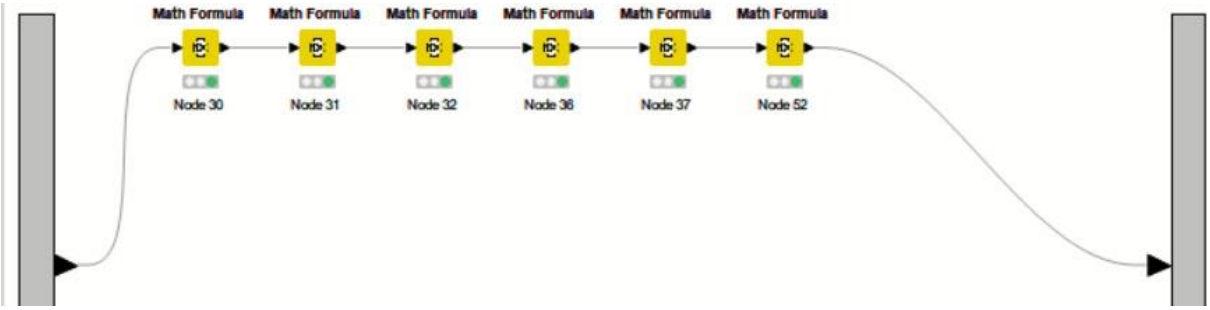


12. ábra: célváltozó előállítása
Forrás: KNIME szoftver, saját eredmények

Az új változó érték (1-es), a 2 nél kisebb tehát a „Nem tudom eldönteni”, illetve a 4 vagy annál nagyobb értékekből lett összeállítva így a „Kicsit valószínű” és a „Nagyon valószínű” is ide tartozik. Vállalati szempontból ezek azok a fogyasztók, akiknek értékük van ezért őket kell kiszűrni. A többi értéket 0-ra kódoltam. Majd az így kapott változót átváltottam string formába, ami a KNIME rendszerén belül a kategorikus változó funkciójával ér fel.

4.2.4 Változó képzés

A meglévő változók mellett én is létrehoztam számos egyéb változót, ami segíteni képes az algoritmusoknak a predikció képzésben. Ahogyan a 7. ábrából is látszik ezt a folyamatot egy úgynevezett meta node segítségével oldottam meg, ami lényegében annyit tesz, hogy számos node összevonásával segít egyszerűsíteni a folyamat vizualizálását. A folyamatot a 13. ábra segítségével fogom bemutatni, amely a meta node kibontásaként jött létre.



13. ábra: Új változók létrehozása a prediktív elemzés elősegítése végett
Forrás: KNIME szoftver, saját eredmények

Az új változók előállításához a KNIME math formula node-jait használtam, ezek a node-ok segítenek abban, hogy különböző változókon elvégzett matematikai műveletek segítségével új változókat tudjunk létrehozni. Az új változókat azonos struktúra alapján állítottam össze, mivel felépítésük hasonló volt, ám motivációm az előállításuk mögött különböző.

Előállított változók:

1. infoCAFE
2. kozosAKTIV
3. nyitottTEMA
4. osszPROGRAMTERV
5. osszFAJTA
6. elmegybMOTIV

Az **infoCAFE** előállítási mögött a motivációm az volt, hogy összegezzem azt, hogy hány fajta információ forrásból hajlandó az adott illető tájékozódni. Ezt a mellékletben található *„Az alábbi opciók közül melyek azok a csatornák, amelyeken keresztül tájékozódnál a társasjátékos kávézóval kapcsolatos újdonságokról?”* bejelölt négyzetek összegzésével tettem meg egyénenként.

A **kozosAKTIV** változó mögött az volt a motivációm, hogy az egyén milyen aktívan osztja meg élményeit. A változó összeállítását a *„Melyik platformokat szoktad használni élmenyeid megosztására?”* bejelölt négyzetek összegzésével tettem meg egyénenként.

A **nyitottTEMA** arra próbál reflektálni, hogy a kitöltő mennyire nyitott új társasjáték tematikák kipróbálására, nyilvánvalóan magasabb érték nagyobb nyitottságot eredményez. A változó összeállítását *„Jelöld be azokat a társasjáték tematikákat, amik érdekesek számodra és szívesen játszanál velük!”* bejelölt négyzetek összegzésével tettem meg egyénenként.

Az **osszPROGRAMTERV** változó előállítása mögött azt vizsgáltam, hogy amikor az egyén valamiféle programot tervez, hány fajta információs forrást használ fel. A változó összeállítását „*Az alábbi információforrások közül, melyek azok, amelyeket rendszeresen használsz szabadidős programok tervezésére, kitalálásra”* bejelölt négyzetek összegzésével tettem meg egyénenként.

Az **osszFAJTA** változó előállítása mögött a motivációm az volt, hogy meg akartam vizsgálni azt, hogy az egyén milyen nyitott új társasjáték fajták kipróbálására. A változó összeállítását „*Az alábbi társasjáték kategóriák közül jelöld be azokat, amelyeket preferálsz!”*

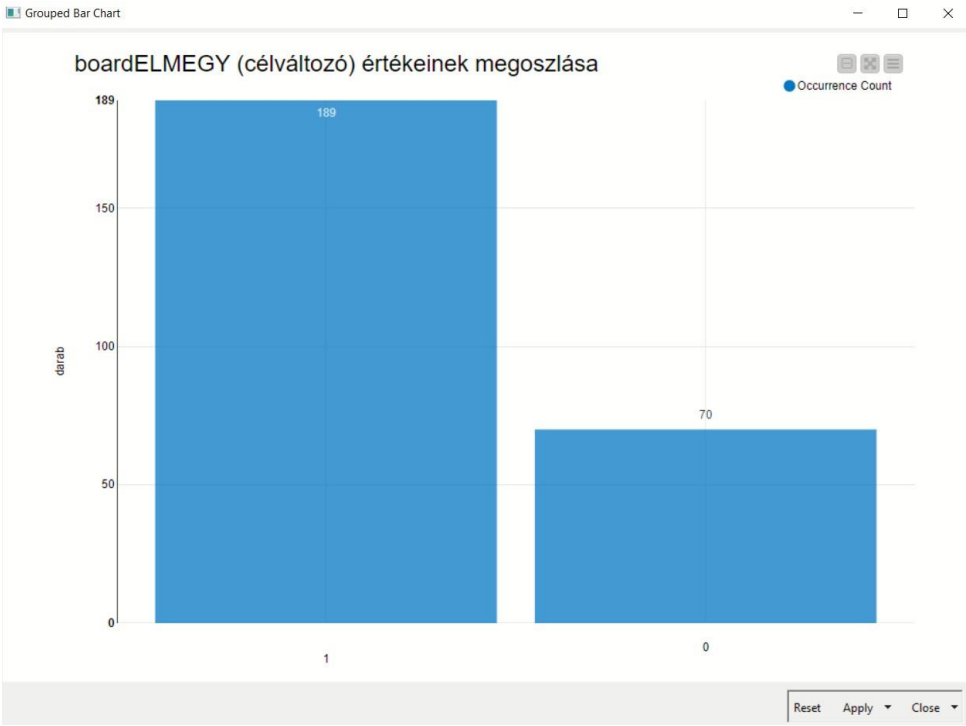
Az **elmegybMOTIV** változó előállításánál az volt a célom, hogy egy motivációs számértéket állítsak elő, az alapján, hogy az egyénnek hány féle motivációja van arra, hogy elmenjen egy társasjátékos kávézóba. Ezt a változót végül nem használtam fel inkább lebontott (eredeti) értékekkel dolgoztam.

4.2.5 Kiugró értékek eltávolítása

Az adattisztítási és elemzési folyamat legutolsó fázisa elemzésemben a kiugró, azaz outlier értékek eltávolítása. Ezen értékek eltávolításához a KNIME-on belül az Outlier Removal node-ot használtam fel. A kiugró értékek elvtávolítása abban segít a modell építési folyamaton belül, hogy a túlzottan eltérő értékek, amik torzítani tudják a modell paramétereinek beállítását, eltávolítódnak, így garantálva egy precízebb modellt.

4.2.6 Célváltozó értékének megfigyelése

A kiugró értékek kiszűrése után, fontos megfigyelnünk azt, hogy a célváltozónk milyen arányban oszlik meg. Az elemzés során ebben fog segíteni az oszlopdiagram és ennek segítségével később diagnosztikát tudunk felállítani az esetleges adateltérésekkel való problémák esetén. A 14. ábrából látható, hogy sokkal több (189) 1-es megfigyelésünk van mint 0-ás (70). Ez később majd a tanítási folyamatban olyan a modellen belüli anomáliákra vezethet, hogy a modell nem lesz képes felismerni a negatív eseteket tehát nem tud majd ténylegesen prediktálni 0-ás eseteket. Ennek orvoslására majd a modell alkotási folyamatban fogok kitérni, hogy hogyan lehet ezt a problémát megoldani, és így a modell tanítási folyamatot javítani. Másik javítási módszer lehet még a mintavétel során az, ha a célváltozó szempontjából diverzebb mintát készítünk, azonban ez jelen esetben nem volt megoldható.



14. ábra: boardELMEGY változó értékeinek megoszlása, Partitioning és SMOTE előtt
Forrás: KNIME szoftver, saját eredmények

4.3 PREDIKCIÓS MODELLEK ELEMZÉSE ÉS ÉRTÉKELÉSE

Ebben az alfejezetben szeretném bemutatni a KNIME-on belüli predikciós modellalkotási folyamatot, és jellemezni az algoritmusok által alkotott modelleket a már említett kettő dimenzió keretein belül. Először bemutatom a tanulási modell kialakításának a folyamatát, majd ezután értelmezem a kapott eredményeket az adott dimenziókon belül.

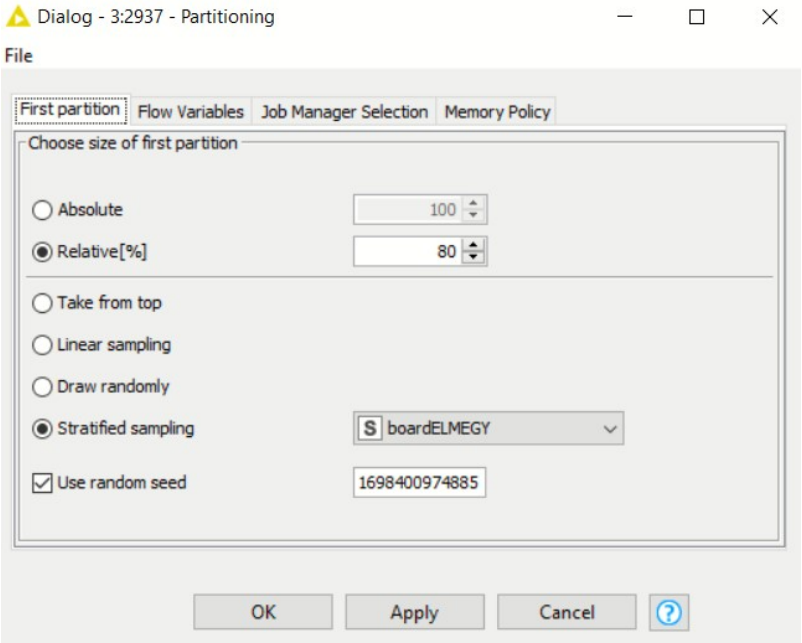
4.3.1 A predikciós modell építése és optimalizálása

A következőkben a modellalkotási folyamat predikciós modellezési részét fogom jellemezni mind a logisztikus regresszió, mind pedig a random forest algoritmusok esetében.

4.3.1.1 Logisztikus regressziós modell

Mielőtt a használni kívánt változók kiválasztásáról tudnánk beszélni először az adathalmazt kettő részre kell választanunk. Ennek segítségére egy Partitioning node-ot használtam fel, ami az adathalmazt 80% tréning adatbázisra, tehát a modell taníttatására és kiértékelésére, és 20% teszt adatbázisra a modell tesztelésére használok fel. Ez a lépés azért fontos, hogy el tudjunk különíteni egy olyan adathalmazt, amin elfogulatlanul tudjuk értékelni modellünket, olyan adaton, amit az adott algoritmus még nem látott, így nincs benne a paraméter számítási folyamatban. A mintavétel során random seed megoldást használtam a folyamat reprodukálhatósága végett, illetve a modellek

későbbiekben való összehasonlíthatóságát illetően is fontos, mert csökkenteni szerettem volna a mintavételezési procedúrával járó elfogultságát az algoritmusoknak. A mintavétel rétegzett mintavétellel megy végbe, és a rétegződést a klasszifikációs problémában kitűzött célváltozó alapján történik. A rétegzett mintavétel azért lényeges, mivel a célváltozók osztályai nincsenek egyensúlyban (16. ábra). A Partitioning node beállítása a 15. ábrán látható.



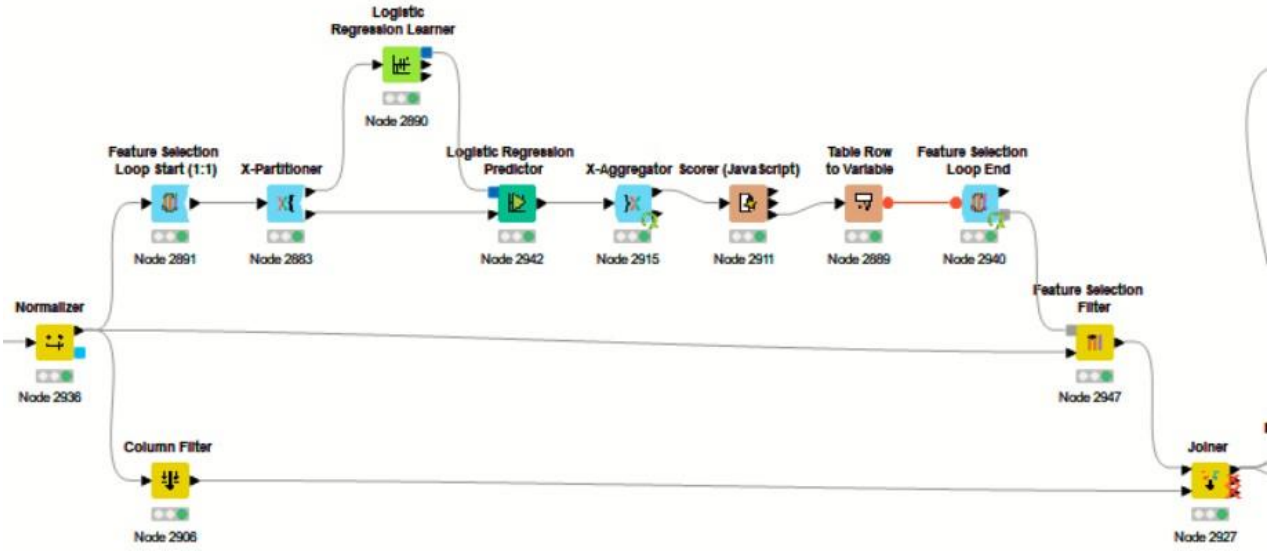
15. ábra: Partitioning node beállításai
Forrás: KNIME szoftver, saját eredmények

Mivel – ahogy azt a korábbi fejezetben is említettem - a célváltozó osztályai nincsenek egyensúlyban a tréning (tanuló) adatrészben, ezért majd a tanítási folyamat során problémák merülhetnek fel, az algoritmus majd nem lesz képes effektíven észrevenni a nullás (esetünkben negatív) értékeket. Ezért ezt még a változók kiválasztása előtt orvosolni kell, erre szolgál a KNIME-ban található SMOTE (Sythethic Minority Over-sampling Technique) node. Ennek segítségével mesterségesen létrehozhatunk egy algoritmus segítségével olyan nullás megfigyeléseket, amivel a 2 osztály egyenlő arányban lesz megtalálható, ezt a technikát oversampling-nek hívják. Ennek a folyamatnak az eredménye a 17. ábrán látható.

16. ábra: boardELMEGY változó megoszlása a tréning adatrészben SMOTE előtt
Forrás: KNIME szoftver, saját eredmények

17. ábra: boardELMEGY változó megoszlása a tréning adatrészen SMOTE után
Forrás: KNIME szoftver, saját eredmények

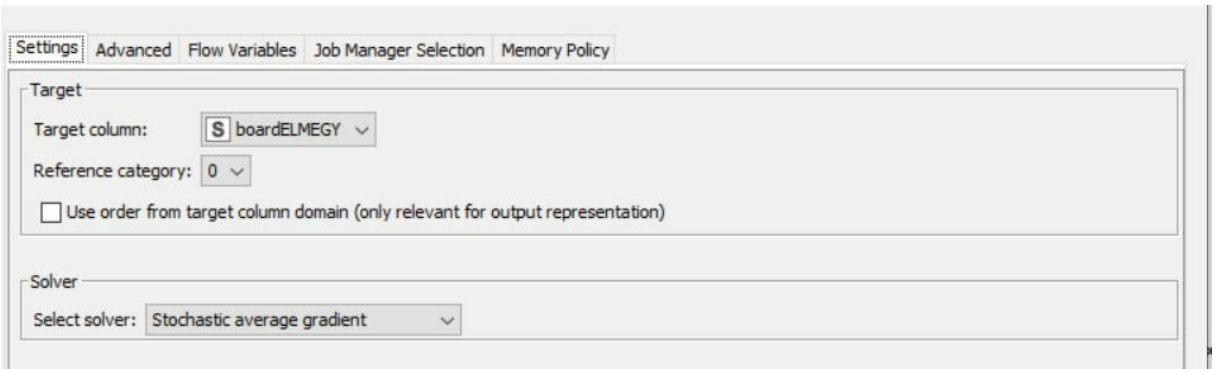
Első fázis



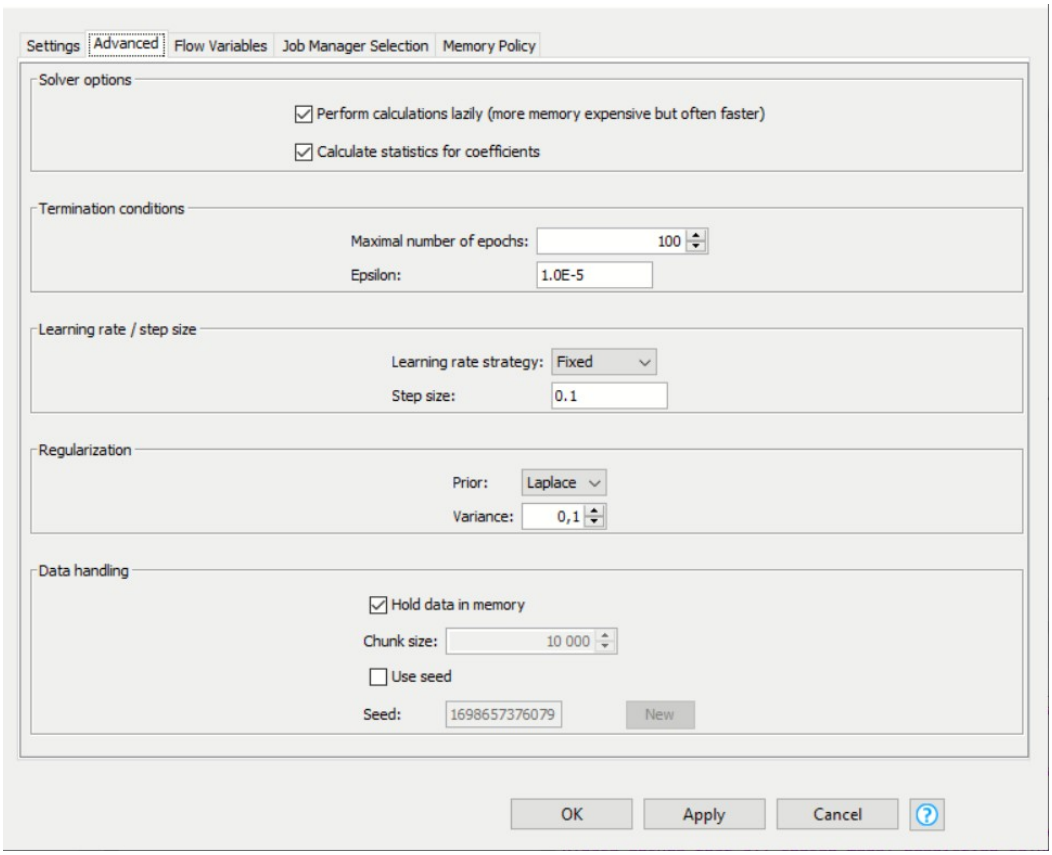
18. ábra: tényezők kiválasztásának első fázisa
Forrás: KNIME szoftver, saját eredmények

A logisztikus regressziós modell összeállítása 4 fázisból állt. Az első 3 fázis azzal foglalkozott, hogy hogyan lehet kiválasztani a változók közül az optimálisakat a modell számára. A következőekben a 18. ábra segítségével szeretném leírni a tényező kiválasztás első fázisát. A változók kiválasztásának első fázisában legelőször normalizálom az összes olyan változót, ami nem kategorikus. Ez a későbbiekben azért lesz lényeges, mert ez gyorsítja a logisztikus regressziónak az optimalizálási folyamatát, illetve a normalizált adatokat könnyebben össze tudja hasonlítani. A változók kiválasztását egy loop folyamat segítségével oldottam meg. A Feature Selection Loop egy start és end node-ból áll, illetve áll még egy filterből, ami segít kiszűrni az optimális változókat. A loop-on belül backward feature selection-t alkalmaztam, aminek lényege az, hogy a loop folyamat úgy indul el, hogy a modellben szerepel az összes változó és mindig a modell számára a legirrelevánsabb változót kiszűri így csökkentve a modell dimenzióját. A loop enged egy általam kijelölt, a konfúziós mátrixban feltüntetett mutatóra optimalizálni, így képes voltam kiválasztani, hogy a loop például minél nagyobb találati arány (Accuracy) alapján válassza ki a változókat, ahogyan ezt a módszertani leírásban is említettem. A Logistic Regression Learner node segít a logisztikus regressziós modell felállításában (a Learner node ezen beállításai minden fázisban azonosak, a beállításokat a 19. ábra mutatja), és a Logistic Regression Predictor node pedig a modell alapján tett predikciók megalkotásában. A tanulási folyamat során, regularizációt is alkalmaztam, ez egy úgynevezett hiperparaméter, amely a tanulási folyamatot segíti. A KNIME logisztikus regresszió esetében 2 fajta regularizációt engedélyez. Az első a Laplace regularizáció, ami az L1 regularizációs

megoldás KNIME-beli interpretációja (Melcher, 2018). Az optimális változók kiválasztása során a regularizáció azért is fontos, mert segít felgyorsítani a folyamatot, illetve segít abban, hogy ki tudjuk szűrni azokat a változókat, amik nem segítenek a modellnek a predikció alkotás folyamatában. A folyamat megalkotásához használt hiperparaméterek a 20. ábrán láthatóak.



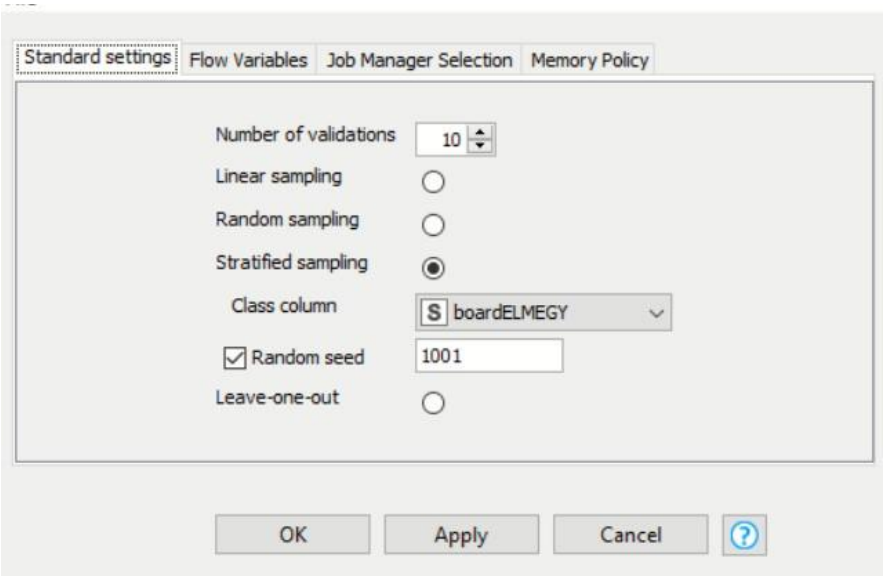
19. ábra: Logistic Regression Learner node beállításai
Forrás: KNIME szoftver, saját eredmények



20. ábra: tényezők kiválasztásának első fázisában használt hiperparaméter beállítások
Forrás: KNIME szoftver, saját eredmények

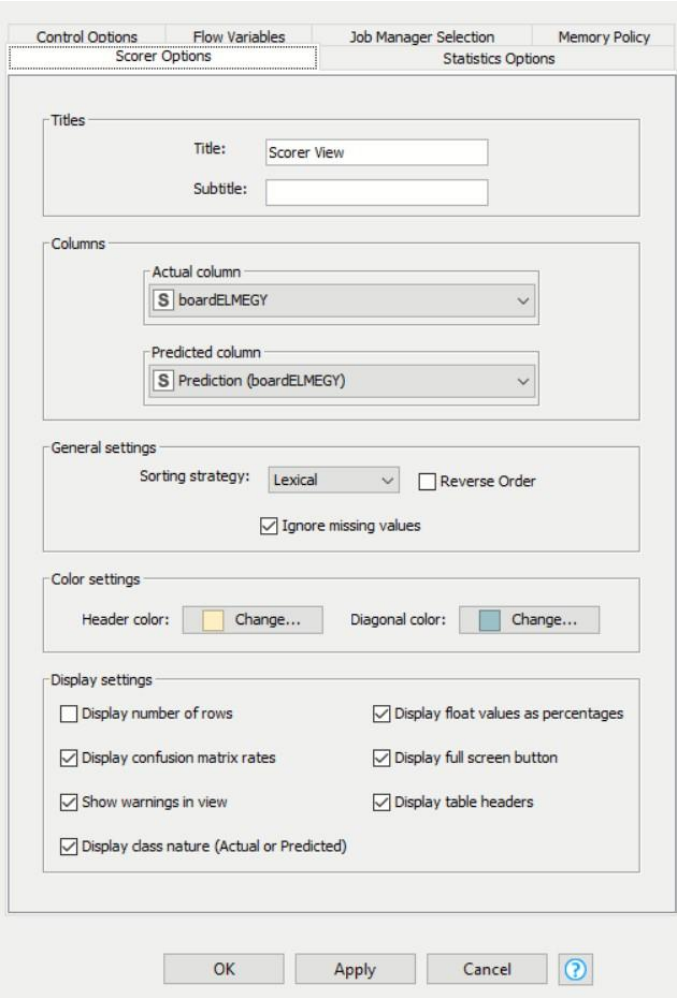
A modelleket k-fold cross validation segítségével értékeltem ki. Ebben a folyamatban segített az X-partitioner és X-aggregator node. A folyamat lényege az, hogy csökkenteni szeretnénk az önkényes tanuló és validációs adathalmaz kijelölésében lévő elfogultságot, ezért a tanuló adrészt 10

egyenlő részre osztjuk, és mindig 9 részt használunk taníttatásra és 1 részt kiértékelésre, ezt a folyamat 10x megismételjük és az így kapott eredmények a tréning (tanuló) adatrészen vett teljesítmény átlagát reflektálják, csökkentve a mintavétellel kapcsolatos elfogultságot (Brownlee, 2023). Ebben a folyamatban is rétegzett mintavételt használtam, mert a tanulási folyamat során biztos akartam lenni abban, hogy minden esetben 50-50% arányban jelenik meg a két célváltozóbeli osztály. A reprodukálhatóság végett random seed-et is alkalmaztam, illetve ez is segített abban, hogy a következő fázisokban a mintavételezés azonos körülmények között tudjon zajlani, ezt a folyamat elemet a 21. ábra demonstrálja. A random seed értéke minden fázisban és modellben ugyanannyi így garantálva azt, hogy a mintázási forma hasonló marad.



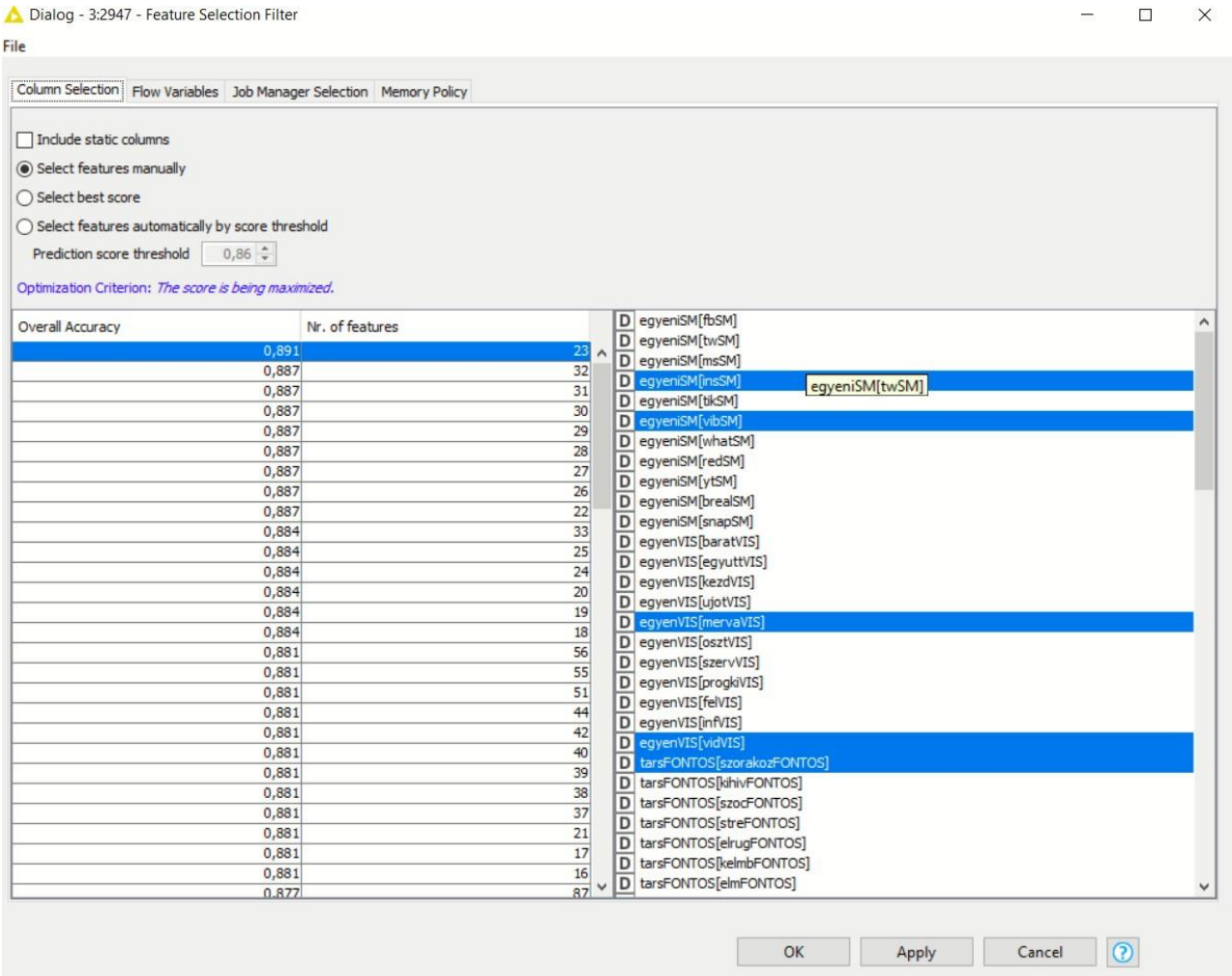
21. ábra: k-fold cross validációhoz szükséges beállítások
Forrás: KNIME szoftver, saját eredmények

A kiértékelésben (22. ábra) a Scorer node segít, ami a prediktált és a tényleges értékek segítségével létrehoz egy konfúziós mátrix-ot.



22. ábra: konfúziós mátrix létrehozása
Forrás: KNIME szoftver, saját eredmények

Végül a fázisban kapott optimális modell kiválasztásában a Feature Selection Filter node fog segíteni. Az optimális tényezőket a 23. ábra demonstrálja.

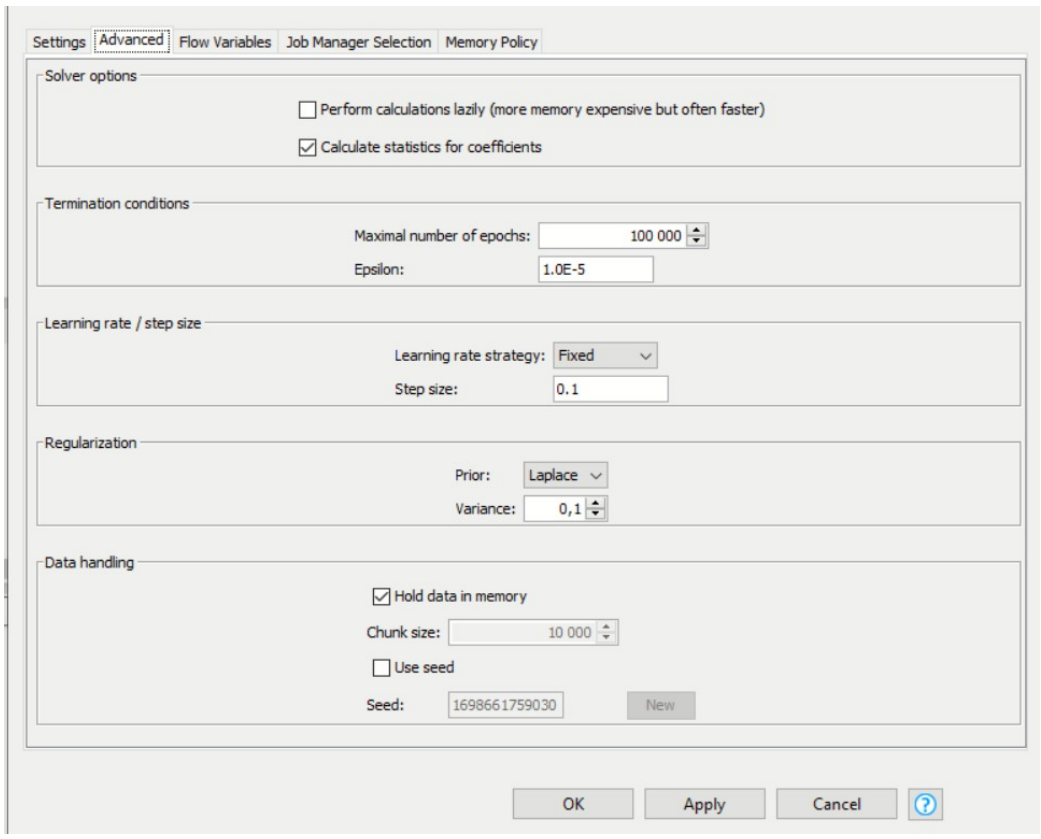


23. ábra: Első fázis optimális tényező összetételének a kiválasztása
Forrás: KNIME szoftver, saját eredmények

A modellek közül természetesen azt választjuk ki, ami a legjobban teljesített. A korábban bemutatott 18. ábrán a Column Filter és a Joiner node csak arra szolgál, hogy egyesíteni tudjam a már szűrt tényezők mellé a tényleges értékeket (célváltozót) újra és tovább tudjam vinni őket a következő fázisba.

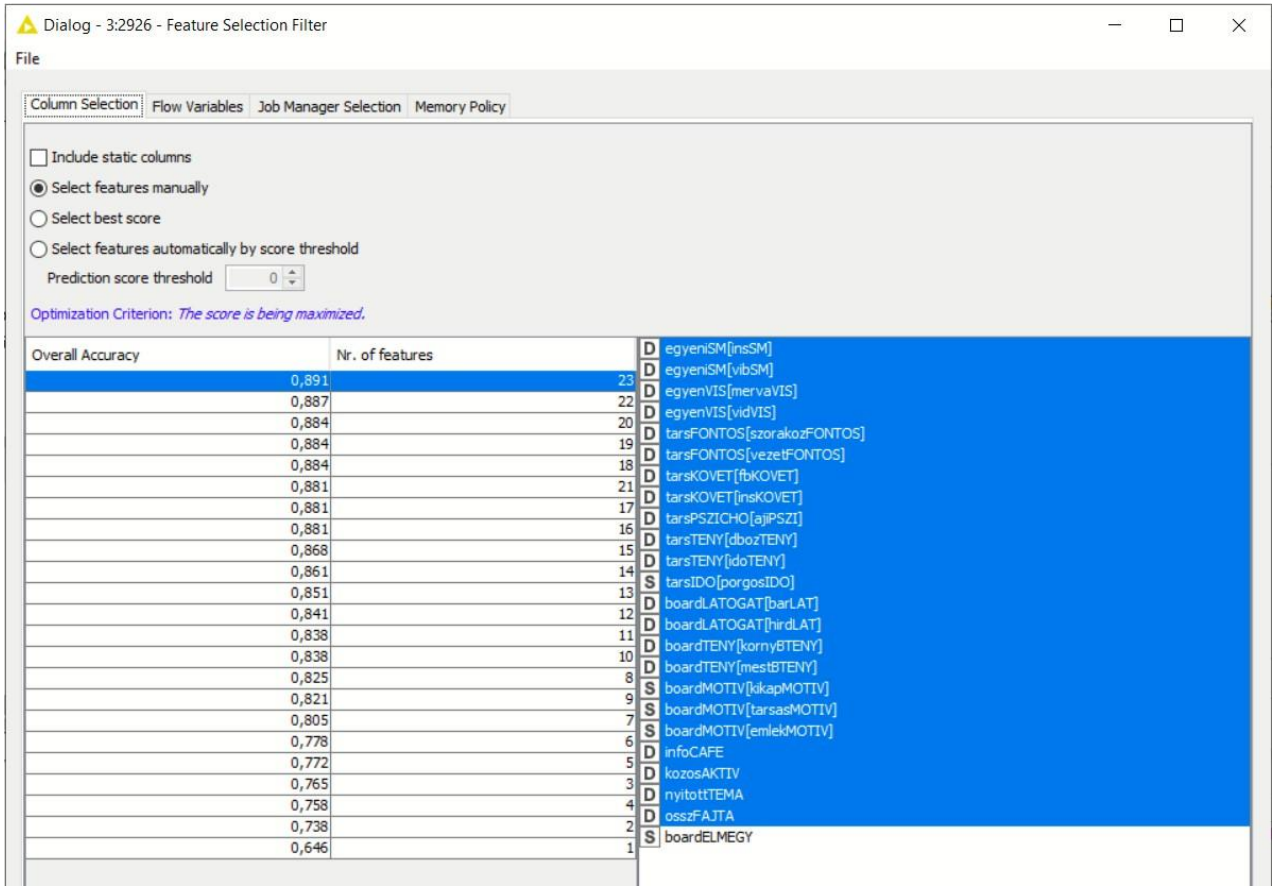
Második fázis

A második fázis hasonló formában történt, mint az első fázis. Az eltérés, viszont abban van, hogy a Logistic Regression Learner node esetében az epoch számot növeltem, így már jobban összpontosítva, arra, hogy méginkább optimális változókat találjak, illetve itt már a tanulási folyamat a csökkentett dimenziók végett sokkal gyorsabb volt ezért kivettem a lazy calculation beállítást, így a pontosabb számításokra összpontosítottam. A második fázis tanulásának beállítását a 24. árba mutatja.



24. ábra: Második fázisban lévő Logistic Regression Learner hiperparaméterei
Forrás: KNIME szoftver, saját eredmények

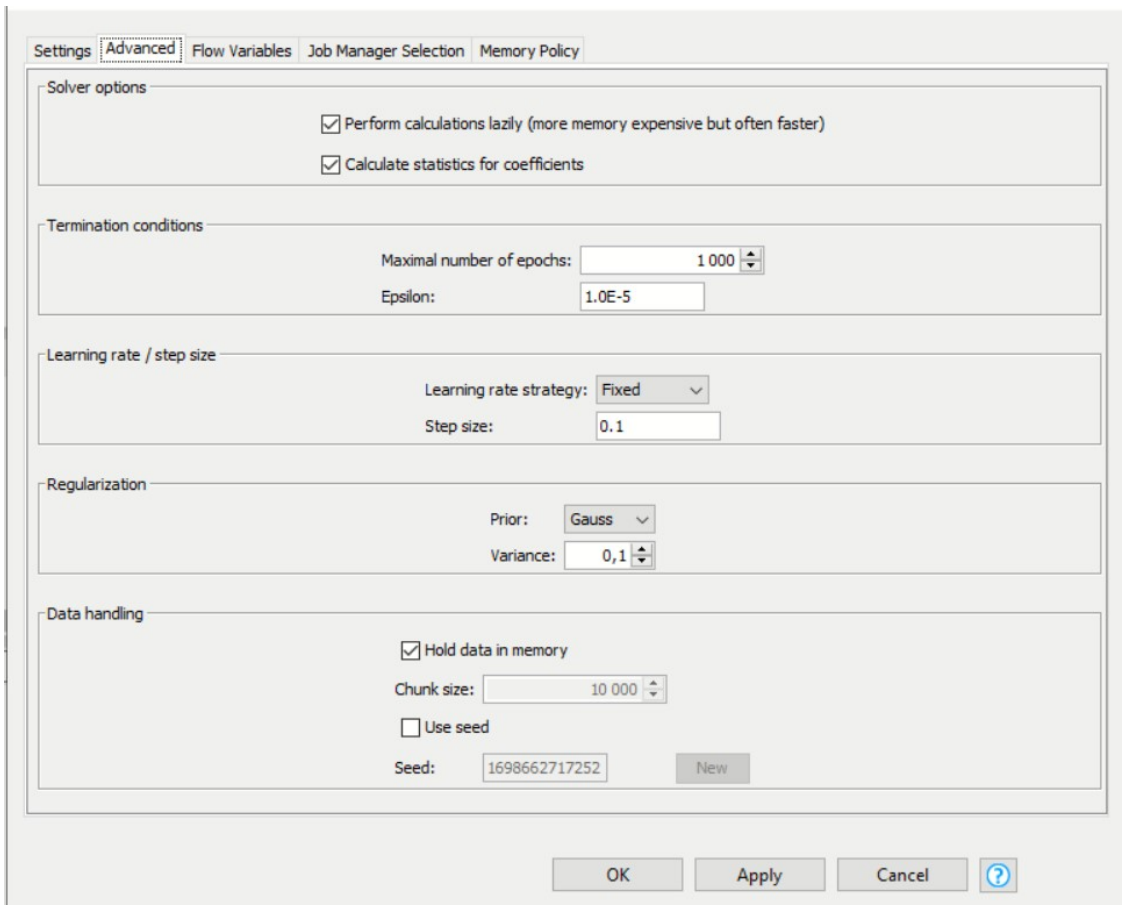
Itt is a loop folyamat végén a Feature Selection Filter segítségével választottam ki az optimálisnak tekinthető modellt. A kiválasztott modellt a 25. ábra mutatja.



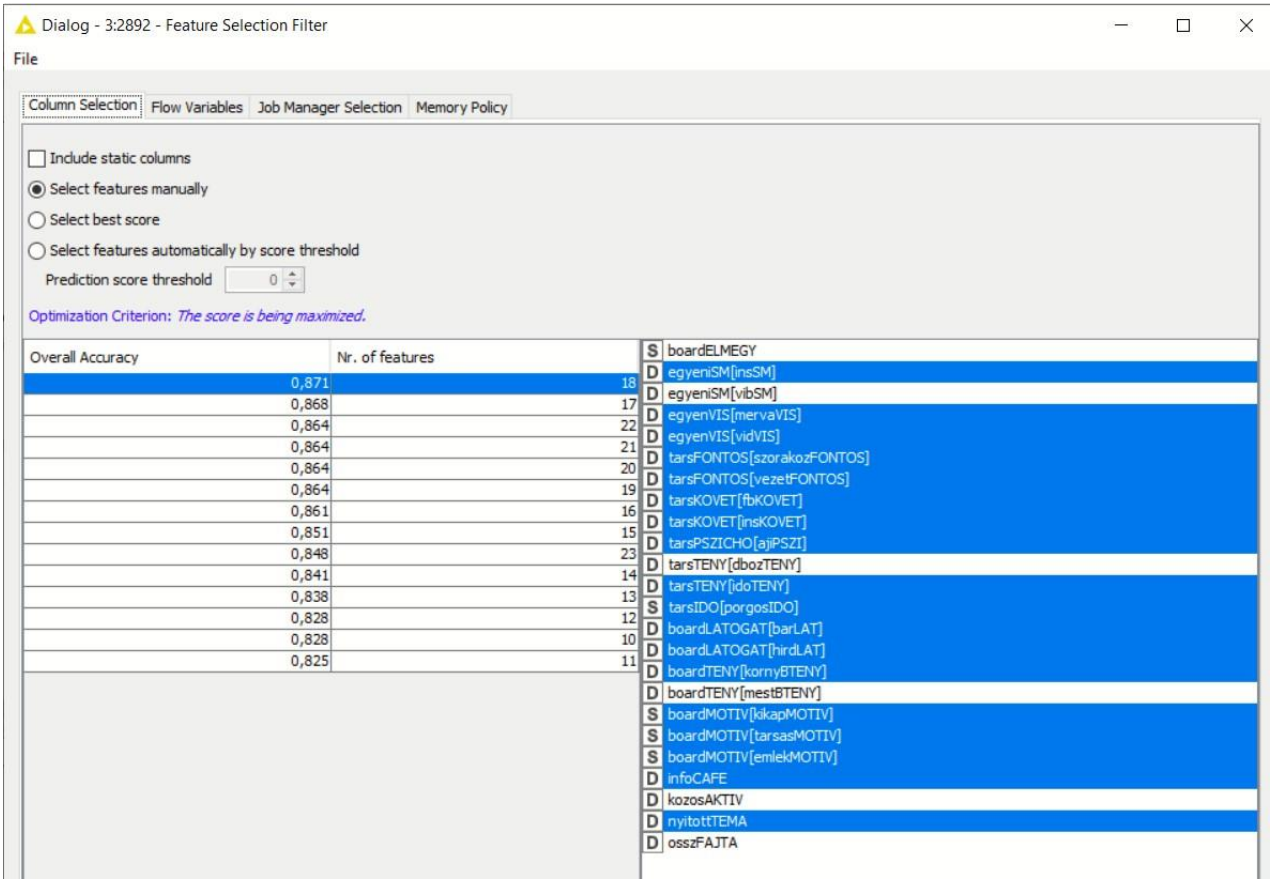
25. ábra: Második fázis optimális tényező összetétele
Forrás: KNIME szoftver, saját eredmények

Harmadik fázis

A 3. fázis abban különbözik az előző kettőtől, hogy itt más regularizációt alkalmaztam, egy már a végleges modell felállítása végett, illetve csökkentettem az epoch számot és megint beállítottam a lazy calculation opciót, hogy gyorsabban tudjak eredményeket kapni. Itt a Laplace regularizációs megoldás helyett Gauss regularizációt alkalmaztam, ami az L2-es regularizációval hozható párhuzamba. Ebben a fázisban már nem feltétlenül az a célom, hogy eltűntessem a változókat, hanem az, hogy megnézzem a maradék változók közül melyek működnek legjobban, illetve erős célom az, hogy el tudjam kerülni az overfitting problémáját is. Az overfitting egy olyan gépi tanulással kapcsolatos probléma, ami akkor merül fel, ha a modell memorizálja a tréning adatrészt és nem tud általánosítani a teszt adatrészen. Az itt használt Logistic Regression Learner node beállításai a 26. ábrán tekinthetők meg. A harmadik fázis során kinyert optimális tényező összetétel a 27. ábrán tekinthető meg.



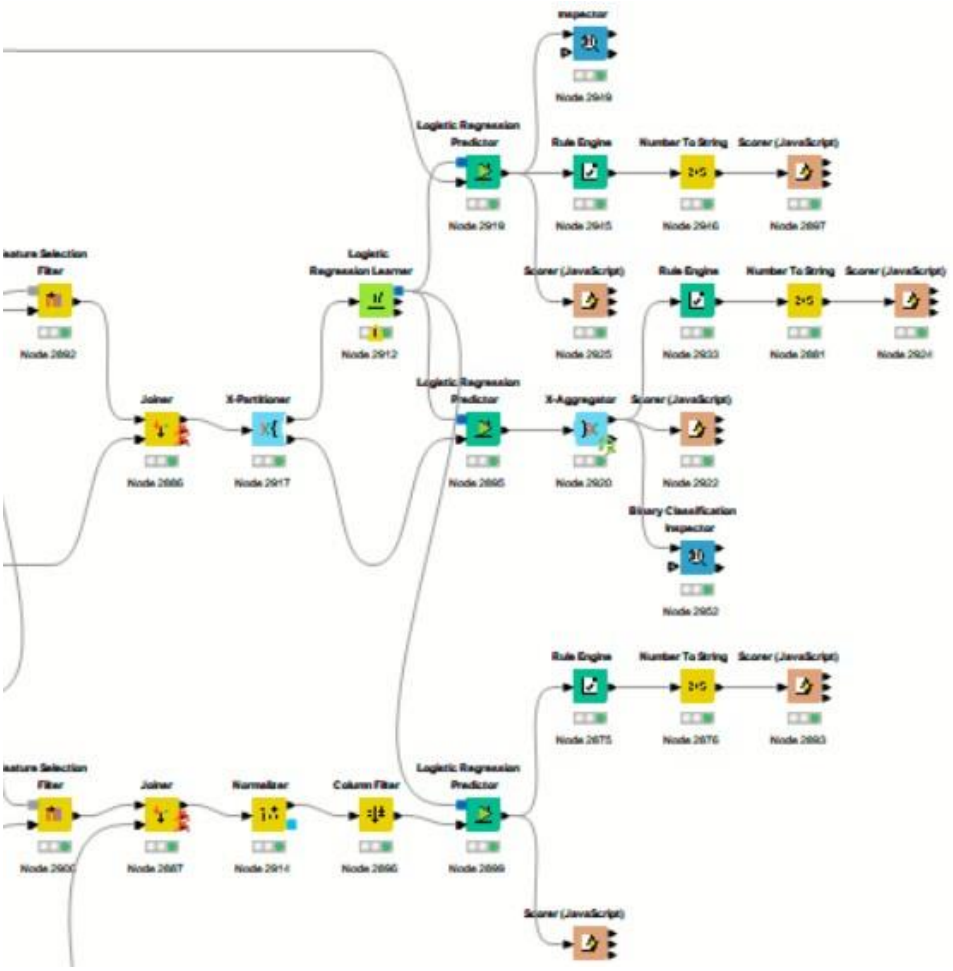
26. ábra: Harmadik fázisban lévő Logistic Regression Learner hiperparaméterei
Forrás: KNIME szoftver, saját eredmények



27. ábra: Harmadik fázis optimális tényező összetétele
Forrás: KNIME szoftver, saját eredmények

Negyedik fázis

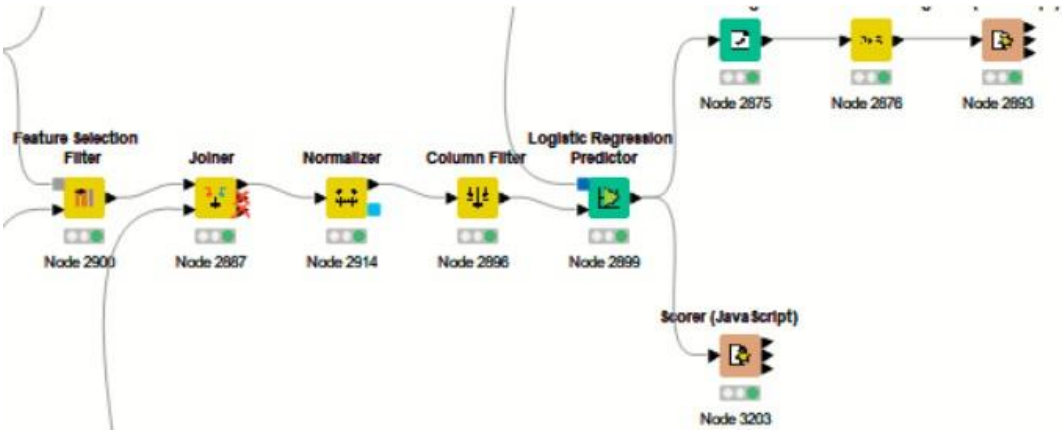
A negyedik fázis a harmadik fázis optimális modelljét használja fel, és fejleszti tovább hiperparaméterek szempontjából. Itt már feltélezem, hogy optimális változó állományom van, viszont a modellt fejleszteni lehet a hiperparaméterek módosításával, vagy később a döntési kritérium módosításával. A negyedik fázis a kutatás szempontjából 3 különböző részre osztható. A három különböző részt a 28-32. ábrák segítségével fogom jellemezni.



28. ábra: Negyedik fázis ábrázolása
Forrás: KNIME szoftver, saját eredmények

Alsó rész (Nem diák adathalmazon való tesztelés)

Az alsó rész a nem diák hallgatókon gyűjtött adathalmazzal foglalkozik és a diákokon felépített modellt használja a kiértékeléshez. Az alsó rész folyamatát a 29. ábra mutatja be.

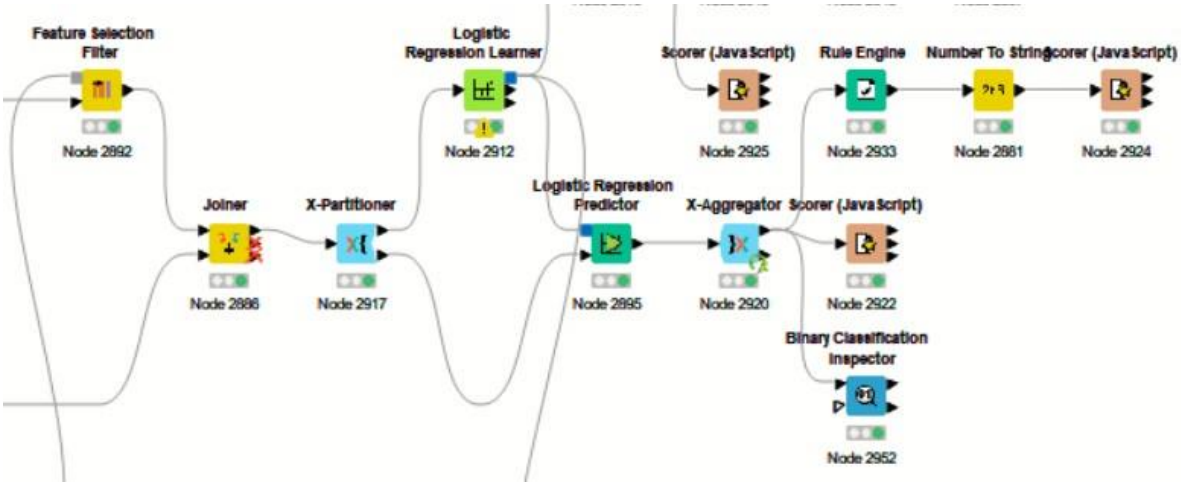


29. ábra: Nem diák adathalmazon történő prediktálás
Forrás: KNIME szoftver, saját eredmények

A Feature Selection Filter-en keresztül kiválasztom azokat a változókat, amik a diák adathalmazon jó prediktív képességet mutattak, és normalizálom a folytonos változókat. A Logistic Regression Predikctor segítségével pedig implementálom a diák adathalmazon taníttatott modellt és a Scorer segítségével kiértékelem azt. A Rule Engine node majd a későbbiekben segít a tréning adaton megvizsgált, optimálisabb döntési kritérium implementálásában.

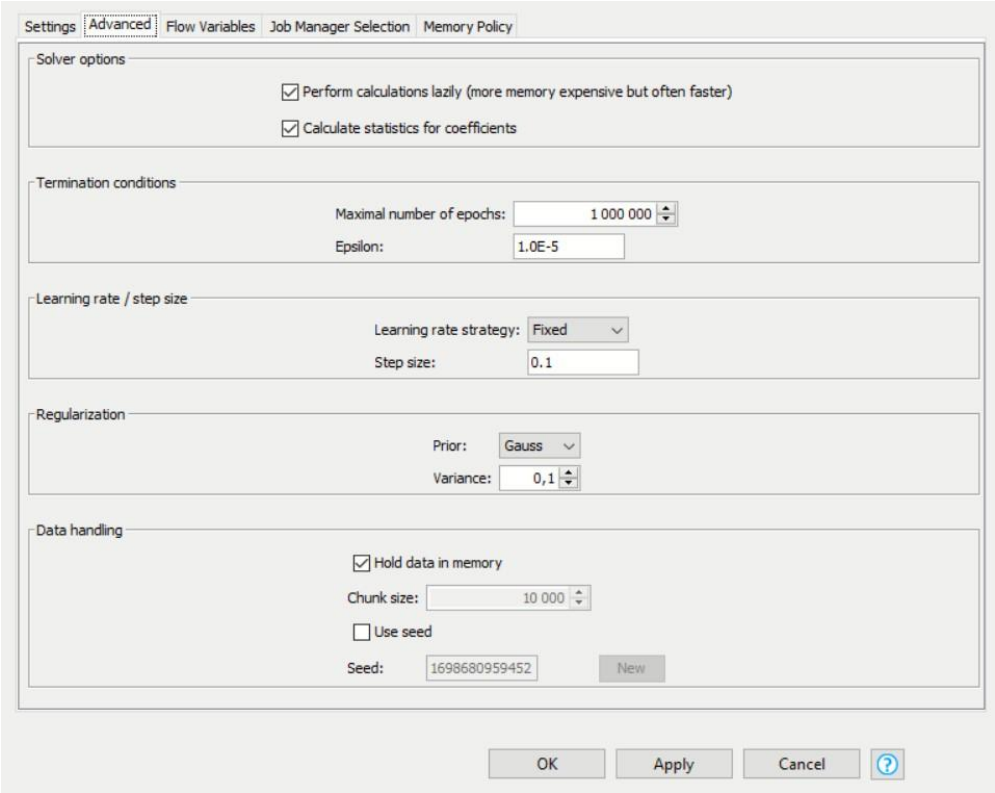
Középső rész (Modell véglegesítés)

A középső rész az, ahol maga a modell létrejön. Ezt a modellt implementálom a teszt adatrészen és a nem diák adathalmazon. A modell véglegesítését a 30. ábra segítségével szeretném bemutatni.



30. ábra: Modell véglegesítése
Forrás: KNIME szoftver, saját eredmények

A Feature Selection Filter segítségével kiválasztottam az optimális modellt és egy újabb már a korábban leírtakhoz hasonló loop segítségével taníttattam a modellt. A modell hasonlóan a Guassian regularizációt alkalmazza, viszont sokkal nagyobb epoch számmal mivel csak 1 kalkulációt kell végig vinni a k- fold crossvalidation loop-on.

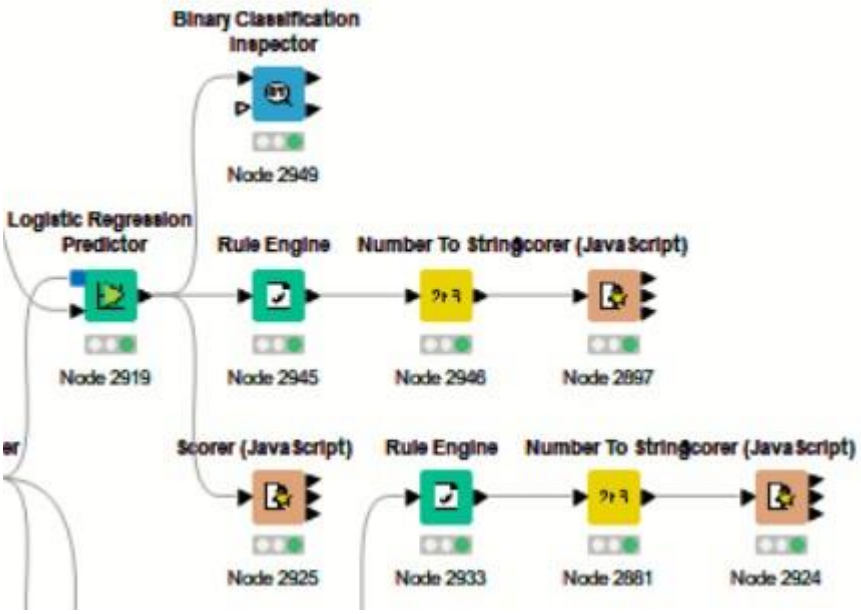


31. ábra: Modell hiperparamétereinek véglegesítése
Forrás: KNIME szoftver, saját eredmények

A Scorer segítségével értékelem ki a tanított modellt. Ami viszont jelentős eltérés az a Binary Classification Inspector node. Ez a node segítséget nyújt abban, hogy egy jobb döntési kritériumot tudjunk felállítani a modellen belül erre majd az első dimenzióban fogok jobban kitérni. A döntési kritérium módosítása abban segít, hogy meg tudjuk változtatni azt a kritériumot, ami alapján eldöntjük milyen valószínűségi érték mellett tartozik valaki az egyik osztályba vagy a másikba. A módosított osztály kritériumot ugyanúgy egy Scorer segítségével értékelem. A Rule Engine itt is a döntési kritérium felülírásában segít majd.

Felső rész (Diákhalmazon való tesztelés)

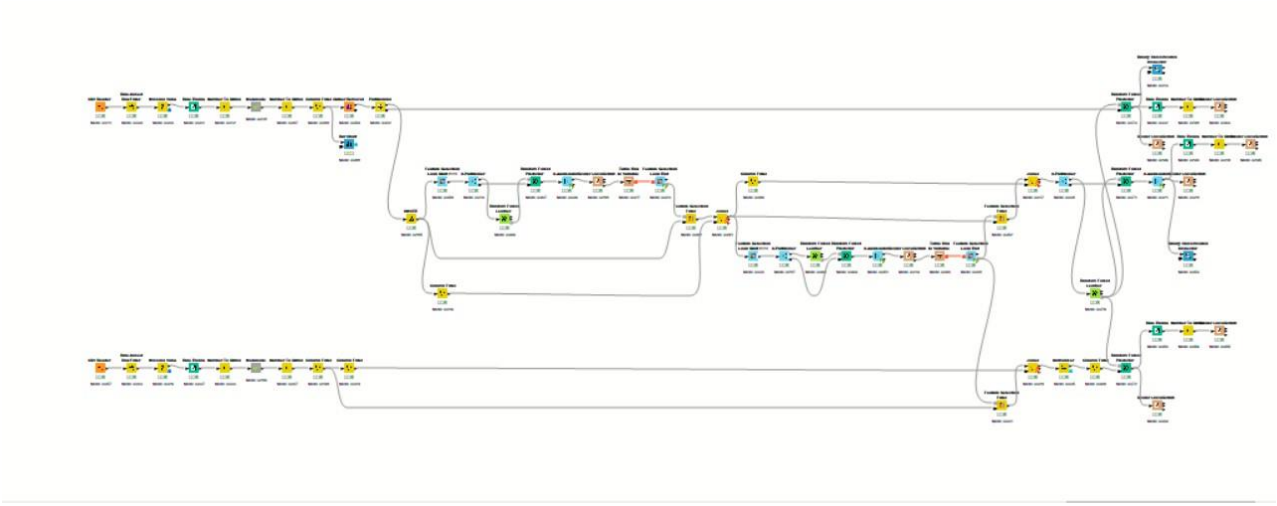
A felső rész a már a Partitioning node segítségével szétválasztott adatrészen értékeli ki a modellt véglegesen. Itt is majd a Rule Engine a döntési kritérium módosításában fog segíteni. A teszt halmaz adatait is normalizáltam. A Logistic Regression Predictor a középső részben alkotott modellt implementálja a teszt adatrészen A felső részt a 32. ábra mutatja be.



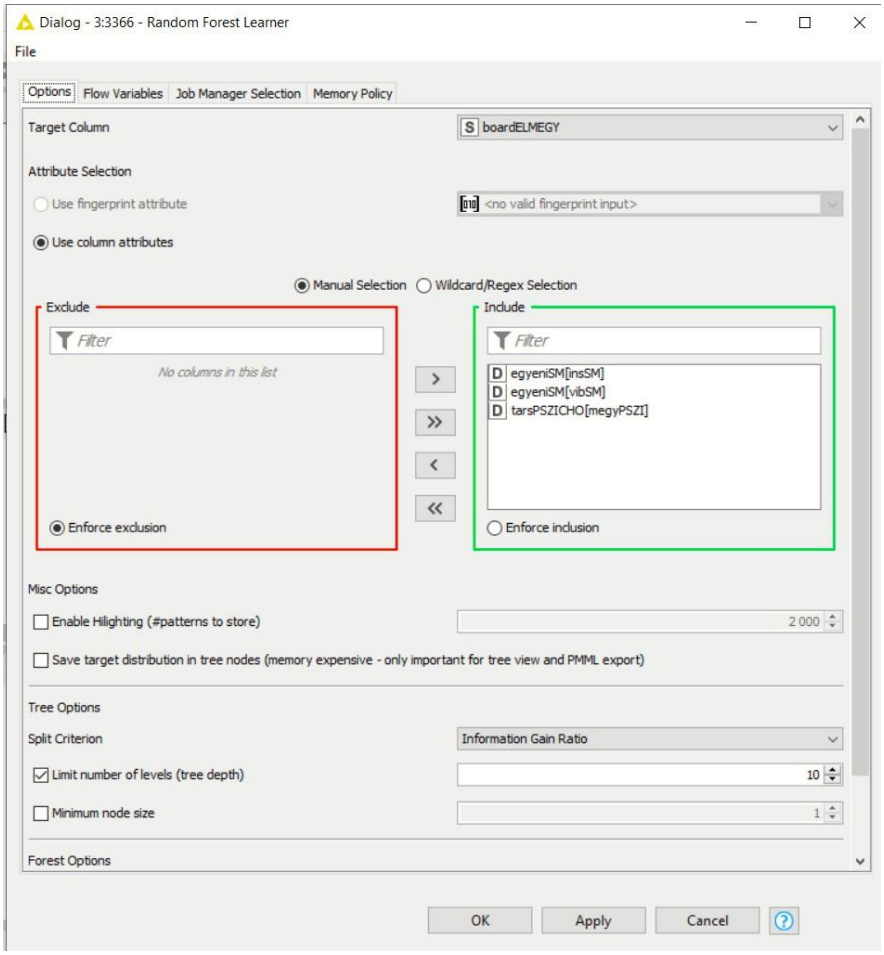
32. ábra: Modell tesztelése a diák adathalmaz teszt részén
Forrás: KNIME szoftver, saját eredmények

4.3.1.2 Random forest modell

A random forest modell, a node-ok sorrendjének aspektusában minden szempontból megegyezik a logisztikus regressziós modell felépítésével, csak a Logistic Regression Learner node-ok helyett Random Forest Learner node-okat használtam, illetve a Logistic Regression Predictor node-ot is lecseréltem a random forest megfelelőjére. A legnagyobb különbség, hogy a 4 fázis helyett, viszont csak 3 van. Ennél a modellnél elég volt 2 változó-szelekciós loop, hogy regularizáció mellett egy kezelhető és jól teljesítő dimenzió számot kapjunk. Ami még nagy különbség, hogy ennél a modellnél sehol sem alkalmaztam normalizációt. A random forest modell építési folyamatát a 33. ábra mutatja be. A folyamat során mindenhol ugyanazt a random forest modell beállítást használtam fel ezt a 34. ábra demonstrálja. A modellben alkalmaztam regularizációt, a fa mélységek kontrollálásán keresztül. A 34. ábrán az include oszlopban feltűnő értékek lényegtelenek.



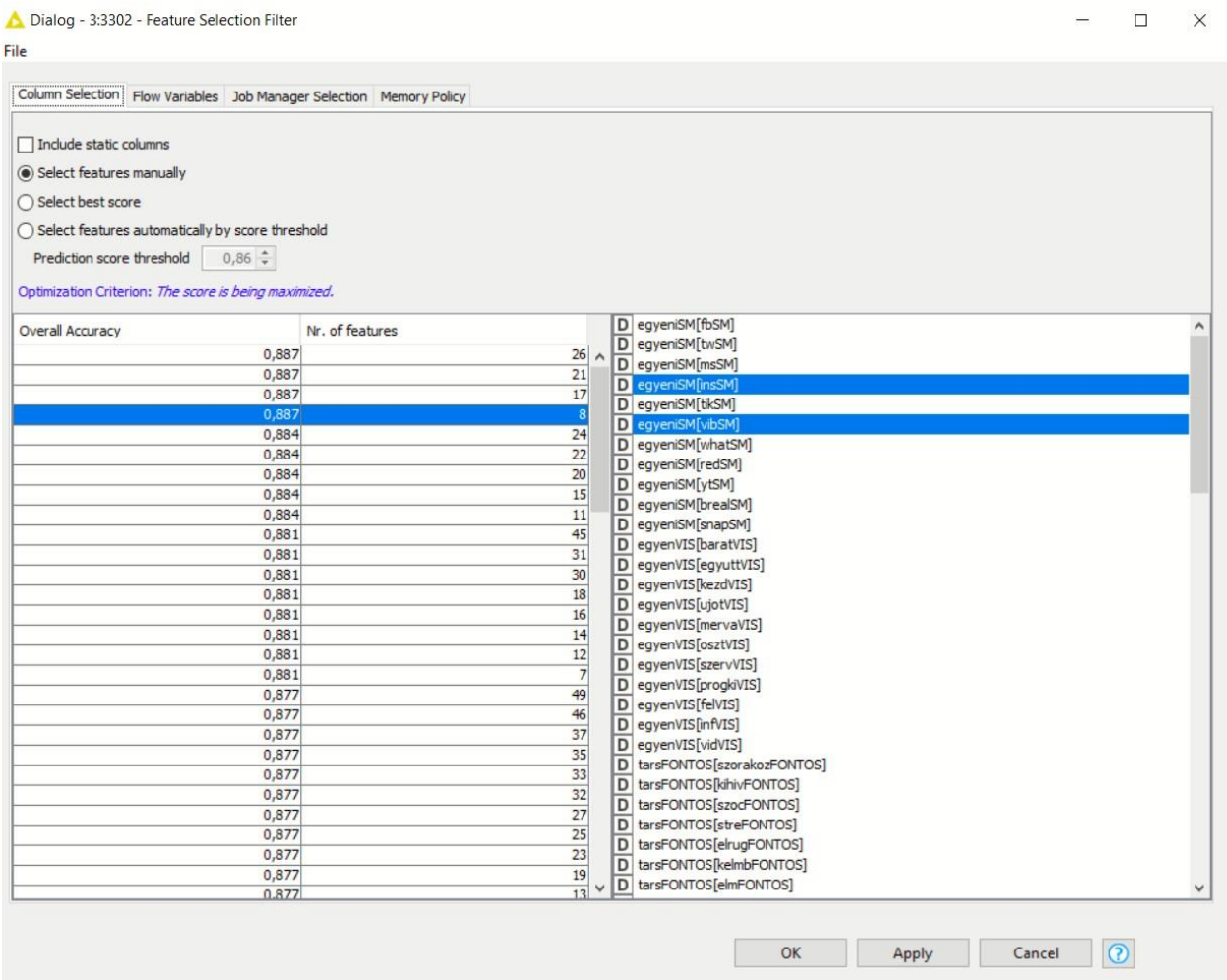
33. ábra: Random forest modell felépítése
Forrás: KNIME szoftver, saját eredmények



34. ábra: Random Forest Learner node beállításai
Forrás: KNIME szoftver, saját eredmények

Első fázis

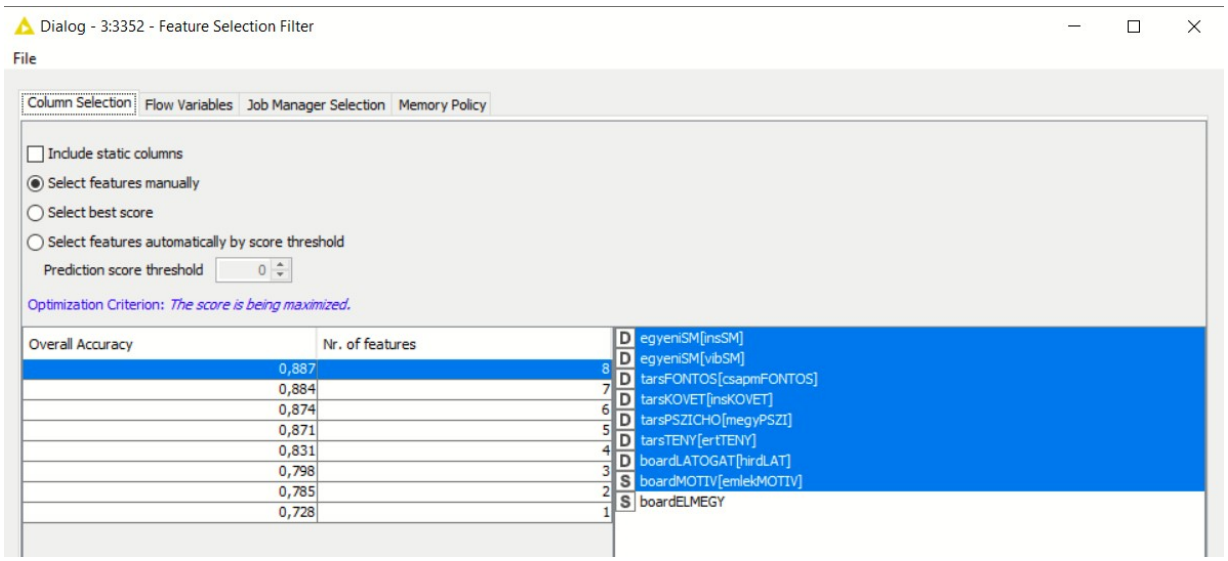
Az első fázisban összeállított változó csomagot a 35. ábra mutatja be. Mint ahogyan látható a 35. ábrából a lehető legkisebb változó összetétel mellett döntöttem. Ennek a miértje Occam borotvájának princípiumából ered. Occam azt mondja ki, hogy ha egy jelenségre több magyarázat is lehetséges, akkor az egyszerűbbet fogadjuk el. Ez a gépi tanulásban azért jó hüvelykujj szabály, mert segít az overfitting elkerülésében, azzal, hogy a modelleket egyszerű formában tartja, illetve látszódik is a 35. ábrán, hogy a nagyobb komplexitás nem segít a prediktálásban, hiszem megegyeznek a ***találati arány*** értékek.



35. ábra: Első fázis optimális tényező összetételének a kiválasztása random forest esetében
Forrás: KNIME szoftver, saját eredmények

Második fázis

A második fázisban a random forest által elért optimális változó összetételt a 36. ábra mutatja. Mint, ahogy látható nem nagyon változott semmi a teljesítményben és így marad a 8 változó.



36. ábra: Második fázis optimális tényező összetételének a kiválasztása random forest esetében

Forrás: KNIME szoftver, saját eredmények

Harmadik fázis

A random forest modellnek a harmadik fázis minden elemében megegyezik a logisztikus regresszió 4 fázisában leírtakkal. Egyetlen eltérés található, hogy az adatok nincsenek normalizálva mert az algoritmus nem kívánja meg.

4.3.2 Első kutatási kérdés

Az első dimenzióban a már a 5.1.3-as alfejezetben feltett kutatási kérdés mellett próbálom kiértékelni a modelleket és javaslatot tenni abban, hogy a kettő (logisztikus regresszió vagy random forest) modell közül melyik lehet a jobb, amennyiben a diák adathalmaz 20%-án, mint teszt adatrészen próbálunk prediktálni.

4.3.2.1 Logisztikus regresszió esetében optimális modell

Coefficients and Statistics - 3:2912 - Logistic Regression Learner

File Edit Hilite Navigation View

Table "Coefficients and Statistics" - Rows: 19 Spec - Columns: 6 Properties Flow Variables

Row ID	S Logit	S Variable	D Coeff.	D Std. Err.	D z-score	D P>|z|
Row1	1	egyeniSM[insSM]	-0.585	0.146	-4.005	0
Row2	1	egyenVIS[mervaVIS]	0.323	0.143	2.258	0.024
Row3	1	egyenVIS[vidVIS]	-0.349	0.145	-2.402	0.016
Row4	1	tarsFONTOS[szorakozFONTOS]	-0.277	0.153	-1.811	0.07
Row5	1	tarsFONTOS[vezetFONTOS]	-0.223	0.146	-1.519	0.129
Row6	1	tarsKOVET[fbKOVET]	0.617	0.213	2.899	0.004
Row7	1	tarsKOVET[insKOVET]	0.725	0.194	3.735	0
Row8	1	tarsPSZICHO[ajiPSZI]	0.309	0.146	2.116	0.034
Row9	1	tarsTENY[idoTENY]	-0.377	0.146	-2.588	0.01
Row10	1	tarsIDO[porgosIDO]=1	0.254	0.22	1.155	0.248
Row11	1	boardLATOGAT[barLAT]	-0.196	0.145	-1.357	0.175
Row12	1	boardLATOGAT[hirdLAT]	0.414	0.151	2.746	0.006
Row13	1	boardTENY[kornyBTENY]	0.465	0.166	2.797	0.005
Row14	1	boardMOTIV[kikapMOTIV]=1	0.567	0.228	2.49	0.013
Row15	1	boardMOTIV[tarsasMOTIV]=1	0.34	0.227	1.499	0.134
Row16	1	boardMOTIV[emlekMOTIV]=1	0.493	0.231	2.132	0.033
Row17	1	infoCAFE	0.203	0.144	1.411	0.158
Row18	1	nyitottTEMA	-0.3	0.14	-2.145	0.032
Row19	1	Constant	-0.865	0.299	-2.891	0.004

37. ábra: Optimális Logisztikus regresszió együtthatói
Forrás: KNIME szoftver, saját eredmények

A 37. ábra mutatja azt az optimális változó összetételt és súlyozást, ami segítségével a modell a tréning adatrészen predikciókat alkotott az értékelés során, illetve majd ez a modell fog a teszt adatrészen prediktálni a Logistic Regression Predictor node segítségével a Variable oszlop alatt található a változó neve, és a Coeff. alatt a súlya. Először viszont szeretnék kitérni a tréning adatrészen elért teljesítményre, amit k-fold cross validáció keretében értékeltem ki.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	134	17	88.74%
1 (Actual)	22	129	85.43%
	85.90%	88.36%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-mea
0	134	22	129	17	88.74%	85.90%	88.74%	85.43%	87.3
1	129	17	134	22	85.43%	88.36%	85.43%	88.74%	86.8

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
87.09%	12.91%	0.742	263	39

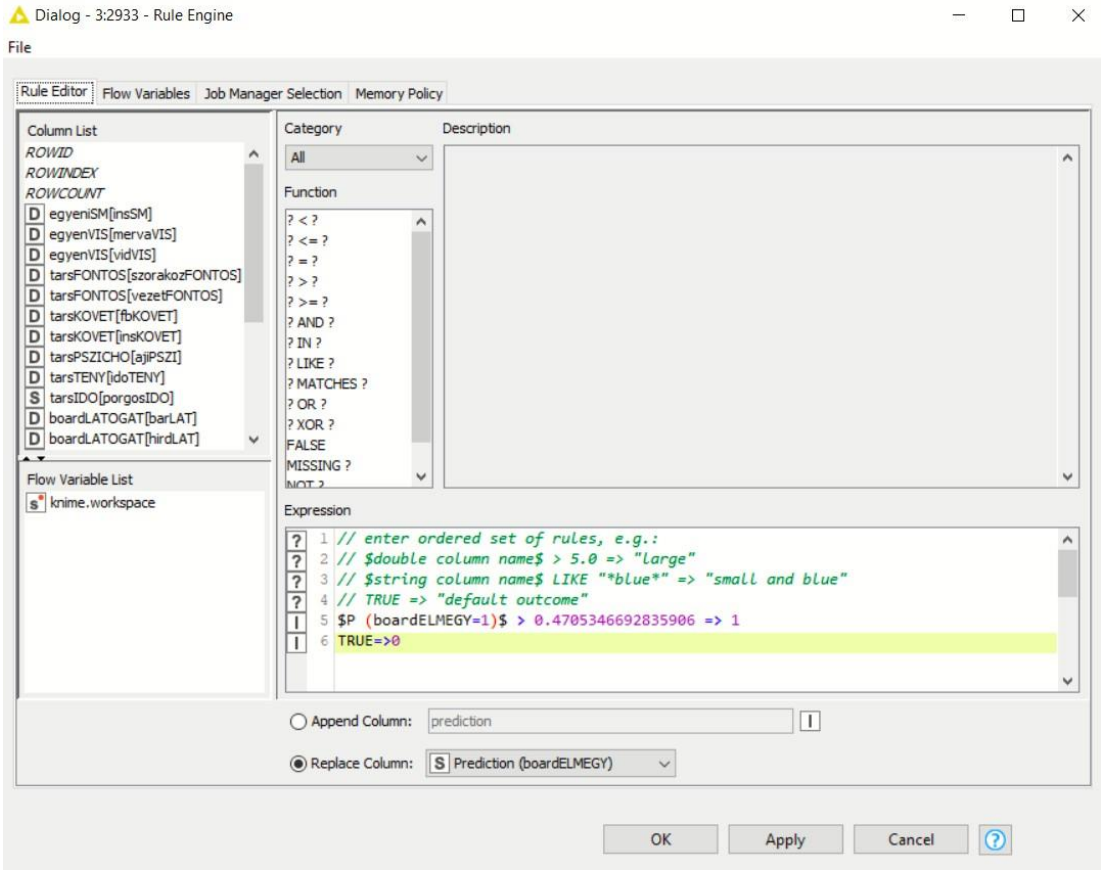
38. ábra: Logisztikus regressziós modell teljesítménye a cross validáció során döntési kritérium módosítása nélkül
Forrás: KNIME szoftver, saját eredmények

Mint ahogy a 38. ábrán látható nagyon szép teljesítményt ért el a modell a találati arány (87,09%) szempontjából a cross validáció keretein belül, de ez sajnos nem használható fel arra, hogy a modell valós teljesítményéről tudjunk állítani valamit. Ez még csak arra jó, hogy hiperparamétereket optimalizáljunk jobb teljesítmény elérése végett, vagy módosítsuk a döntési kritériumot. Én az utóbbi megoldást választottam. Mint ahogyan a modell építés 4. fázisában felvázoltam ebben a feladatban a Binary Classification Inspector fog segíteni. A node segítségével optimalizálhatjuk a döntési kritériumot, hogy maximális teljesítményt tudjunk elérni a találati arány terén.

39. ábra: Optimális döntési kritérium a logisztikus regressziós modellen belül maximális találati arány elérése végett

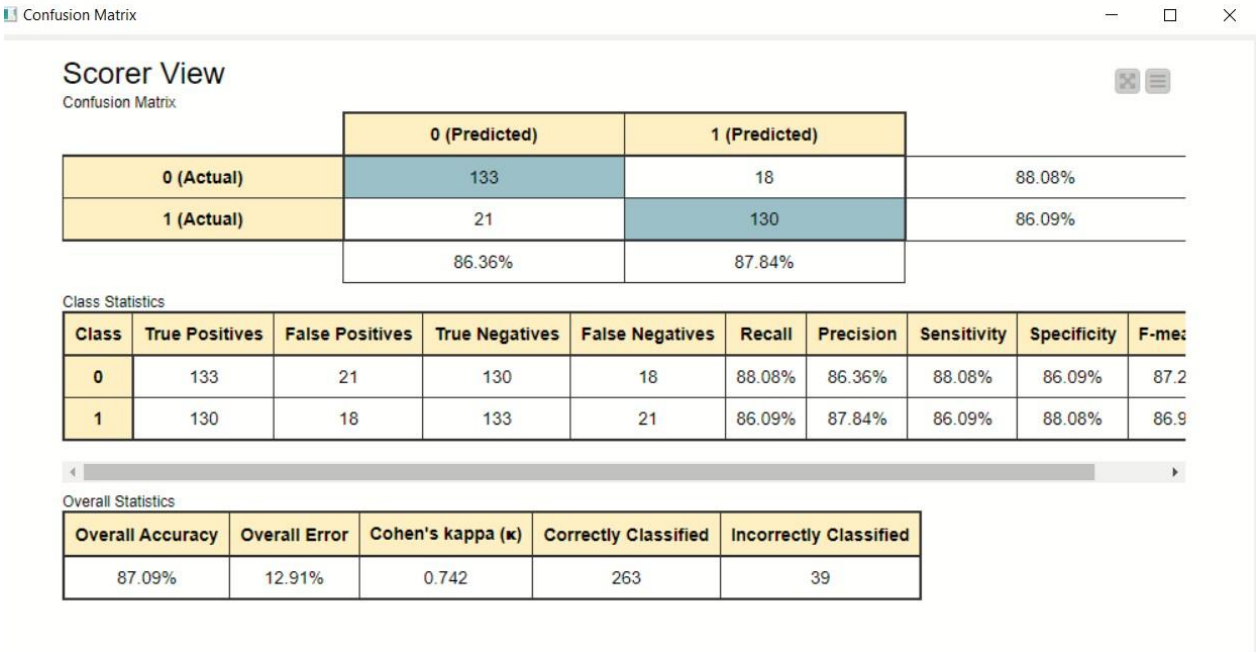
Forrás: KNIME szoftver, saját eredmények

Mint ahogy látható a 39. ábrán a node a treshold, tehát a döntési kritérium módosítását kéri. Ezt az értéket használom fel a Rule Engine segítségével a prediktált változó újra kódolására.



40. ábra: Predikciók újra definiálása új döntési kritérium mellett a tréning adatrészen logisztikus regresszió esetében

Forrás: KNIME szoftver, saját eredmények



Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	133	18	88.08%
1 (Actual)	21	130	86.09%
	86.36%	87.84%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-mea
0	133	21	130	18	88.08%	86.36%	88.08%	86.09%	87.2
1	130	18	133	21	86.09%	87.84%	86.09%	88.08%	86.9

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
87.09%	12.91%	0.742	263	39

41. ábra: Logisztikus regressziós modell teljesítménye a tréning adatrészen döntési kritérium módosításával

Forrás: KNIME szoftver, saját eredmények

Modellben a döntési kritérium módosítása esetében (az alap 0,5-ről módosult az optimális 0,47053-ra) nem történt semmi a ***találati arányt*** illetően, viszont a felidézés (recall) esetében igen. A 1-es eset (boardELMEGY=1) felidézése megnőtt (85,43%-ról 86,09%-ra) viszont a 0-s eseté csökkent (87,74-ről 88,08%-ra).

Logisztikus regresszió teljesítményének értékelése a teszt adatrészen:

A teljesítmény értékelése a teszt adatrészen belül is két féle módon zajlott. Ez az, ami igazán képet is ad a modellünk teljesítményéről. Az első kiértékelés a sima modell teljesítményét mutatja optimális döntési kritérium nélkül. A második a tréning részen elért optimális döntési kritérium mellett tünteti fel a modell eredményét.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	14	0	100.00%
1 (Actual)	14	24	63.16%
	50.00%	100.00%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-me
0	14	14	24	0	100.00%	50.00%	100.00%	63.16%	66.
1	24	0	14	14	63.16%	100.00%	63.16%	100.00%	77.

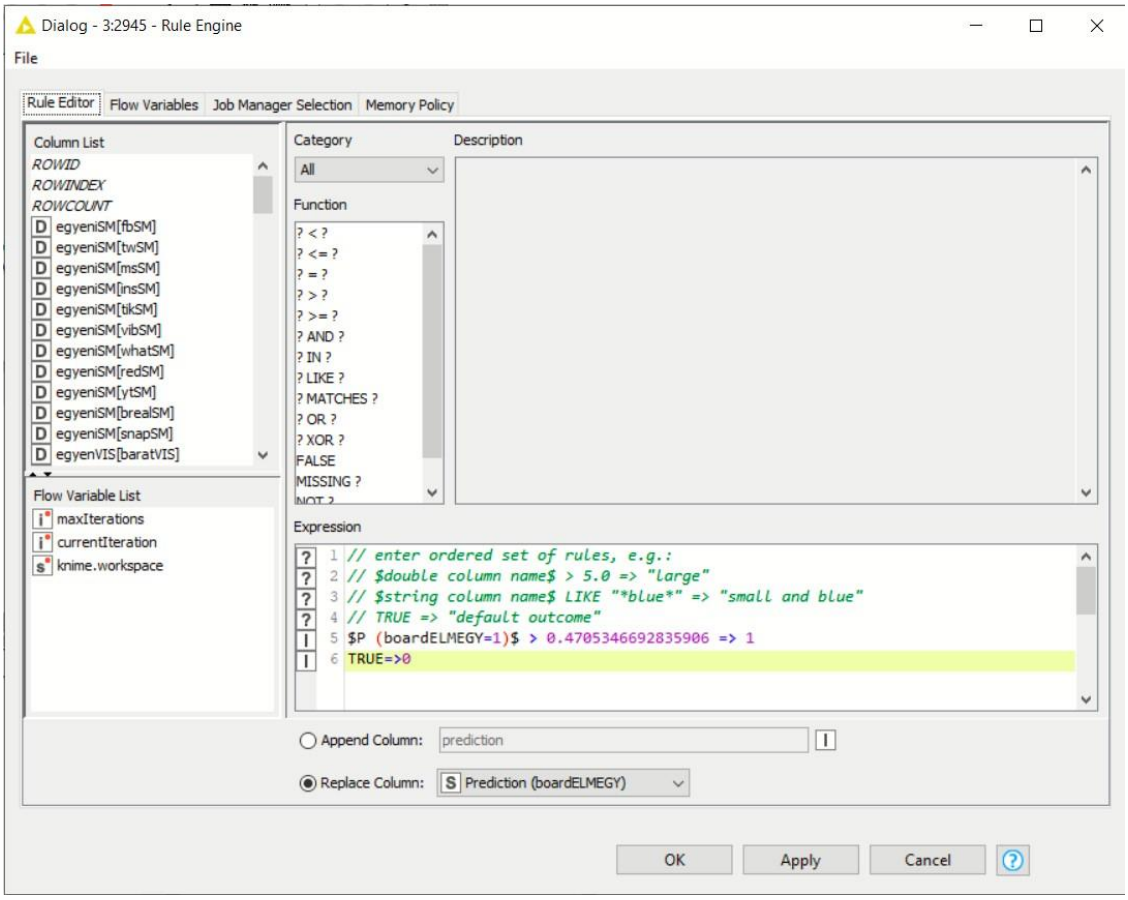
Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
73.08%	26.92%	0.480	38	14

42. ábra: Logisztikus regressziós modell teljesítménye a teszt halmazon döntési kritérium módosítása nélkül

Forrás: KNIME szoftver, saját eredmények

A modell ***találati arányában*** megfigyelhető egy erős csökkenés a tréning adatrészhez képest a ***találati arány*** 87,09%-ról csökkent le 73,06%-ra. Ez sok esetben overfitting-re tud utalni, ezt viszont lehet mérsékelni hiperparaméterek, vagy döntési kritérium módosításával is. Mint ahogyan látható a 42. ábrából a modell képes megtalálni az összes 0-ás esetet, de nagyon szigorú néhány 1-es esettel szemben és 1-es helyett, 0-nak prediktálja. A modellt ebben az esetben úgy lehet javítani, hogy a döntési kritériumot lejjebb visszük. Esetünkben ez a tréning adatrészen elért optimális döntési kritériumot jelenti. A 42. ábra mutatja be a döntési kritérium módosítását.



43. ábra: Predikciók újra definiálása új döntési kritérium mellett a teszt adatrészen logisztikus regressziónál

Forrás: KNIME szoftver, saját eredmények

A tréning adatrészen végzett k-fold cross validáció keretében megtalált optimális döntési kritériumot alkalmaztam itt (0,47053), ez azért fontos, mert például megcsinálhattam volna azt is, hogy a teszt adatrészen is keresek egy optimális döntési kritériumot, de ez torzítaná a modellezést, mivel ekkor a teszt adatrészre kezdeném átszabni a modellt, tehát a modell elkezdene a teszt adatrészben található adatoktól is függeni. A modell teljesítményét új döntési kritérium mellett értékelő konfúziós mátrixot a 44. ábra mutatja be.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	14	0	100.00%
1 (Actual)	10	28	73.68%
	58.33%	100.00%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-me
0	14	10	28	0	100.00%	58.33%	100.00%	73.68%	73.
1	28	0	14	10	73.68%	100.00%	73.68%	100.00%	84.

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
80.77%	19.23%	0.601	42	10

44. ábra: Logisztikus regressziós modell teljesítménye a teszt adatrészen döntési kritérium módosításával

Forrás: KNIME szoftver, saját eredmények

Mint, ahogy látható a döntési kritérium módosítására a modell jobb teljesítmény ért el a ***találati arány*** 73,06%-ról 80,77%-ra nőtt, tehát jó volt az a feltételezés, hogy a modellnek túl konzervatívan értékeli a 0-s értéket és jobb teljesítményt tudunk elérni a döntési kritérium módosításával.

Konklúzió a logiszikus regressziós modellről:

Ezek alapján logisztikus regresszió esetében a döntési kritériummal módosított modell lesz az, amit össze kell hasonlítanunk a random forest modellnél található optimális megoldással.

4.3.2.2 Random forest esetében optimális modell:

A random forest modellnél is először szeretném bemutatni a tréning adatrészen elért eredményeket majd kitérni a teszt adatrészen elért eredményekre. A 44. ábra mutatja be a tréning adatrészen elért eredményeket döntési korlát módosítása nélkül.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	136	15	90.07%
1 (Actual)	23	128	84.77%
	85.53%	89.51%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-mea
0	136	23	128	15	90.07%	85.53%	90.07%	84.77%	87.7
1	128	15	136	23	84.77%	89.51%	84.77%	90.07%	87.0

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
87.42%	12.58%	0.748	264	38

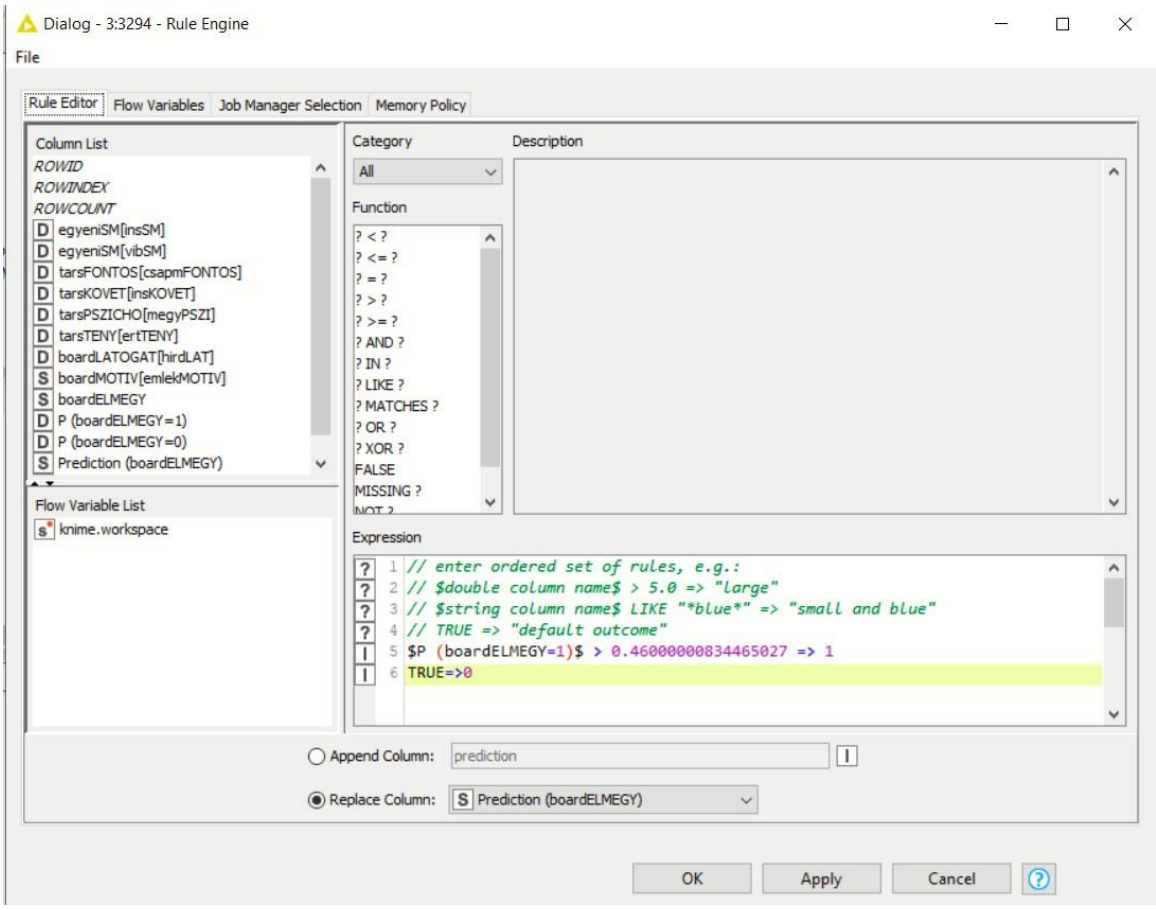
45. ábra: Random forest modell teljesítménye a cross validáció során döntési kritérium módosítása nélkül

Forrás: KNIME szoftver, saját eredmények

A 44. ábrán is jól látható az, hogy modell a k-fold cross validáció keretein belül nagyon jól teljesít (a ***találati arány*** 87,42%), de mint ahogy a logisztikus regressziónál megállapítottam ez sajnos nem használható fel a modell értékelésére, viszont segíteni tud a döntési kritérium esetleges módosításában. Az optimális döntési kritériumot a 46. ábra demonstrálja, illetve az optimális döntési kritérium implementálását a 47. ábra.

46. ábra: Optimális döntési kritérium a random forest modellen belül maximális találati arány elérése végett

Forrás: KNIME szoftver, saját eredmények



47. ábra: Predikciók újra definiálása új döntési kritérium mellett a tréning adatrészen random forest esetében

Forrás: KNIME szoftver, saját eredmények

Ezt követően a 48. ábrában látszik, hogy a ***találati arány*** ennek hatására javul (87,42%-ról 88,74-ra nő), és a modell jobban képes felismerni az 1-es értékeket.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	136	15	90.07%
1 (Actual)	19	132	87.42%
	87.74%	89.80%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-mea
0	136	19	132	15	90.07%	87.74%	90.07%	87.42%	88.8
1	132	15	136	19	87.42%	89.80%	87.42%	90.07%	88.5

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
88.74%	11.26%	0.775	268	34

48. ábra: Random forest modell teljesítménye a tréning adatrészen döntési kritérium módosításával

Forrás: KNIME szoftver, saját eredmények

Random forest teljesítményének értékelése a teszt adatrészen:

Mint ahogy a logisztikus regresszió esetében a random forest modell teljesítményét is csak a teszt adatrészen tudjuk igazán értékelni. A logisztikus regresszióhoz hasonlóan itt is megvizsgáltam, mi történik akkor, ha a modell döntési kritériumát változtatjuk. A 48. ábra a modell teljesítményét mutatja be.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	7	7	50.00%
1 (Actual)	3	35	92.11%
	70.00%	83.33%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-mea
0	7	3	35	7	50.00%	70.00%	50.00%	92.11%	58.3
1	35	7	7	3	92.11%	83.33%	92.11%	50.00%	87.5

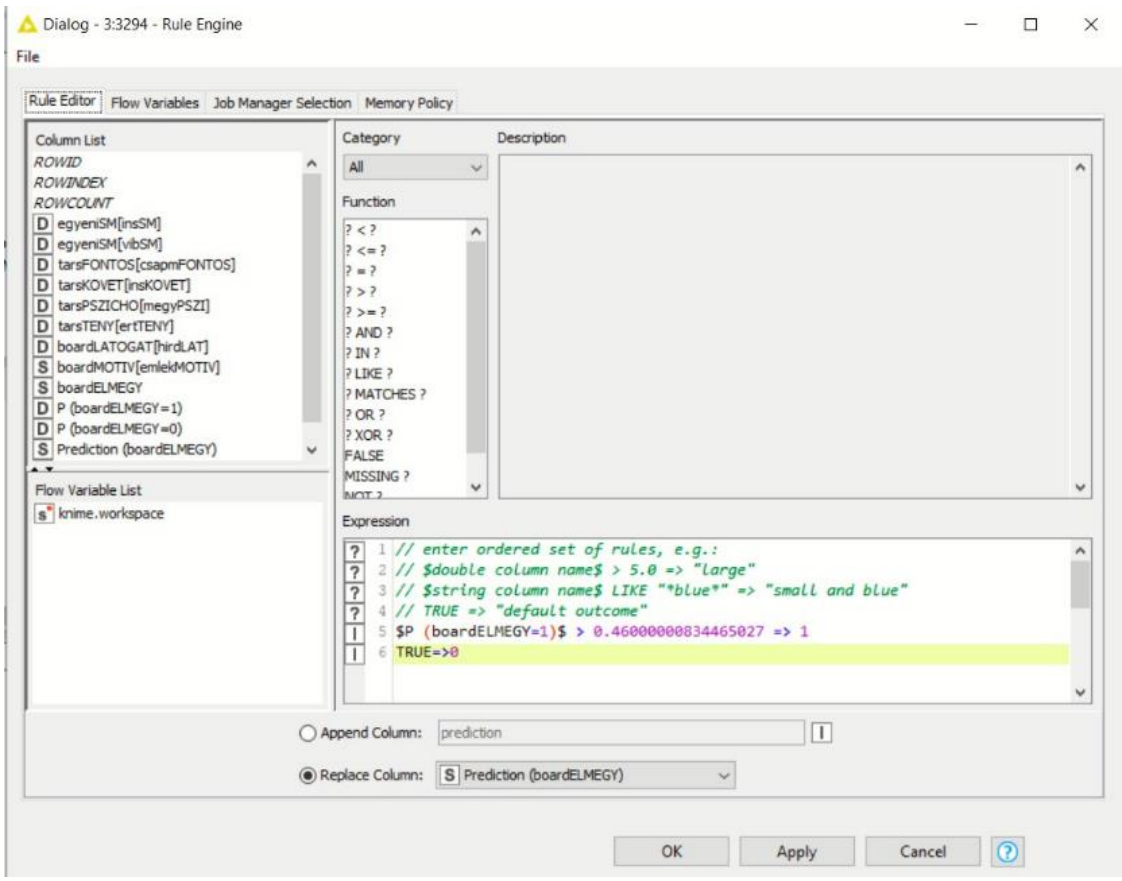
Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
80.77%	19.23%	0.463	42	10

49. ábra: Random forest modell teljesítménye a teszt halmazon döntési kritérium módosítása nélkül

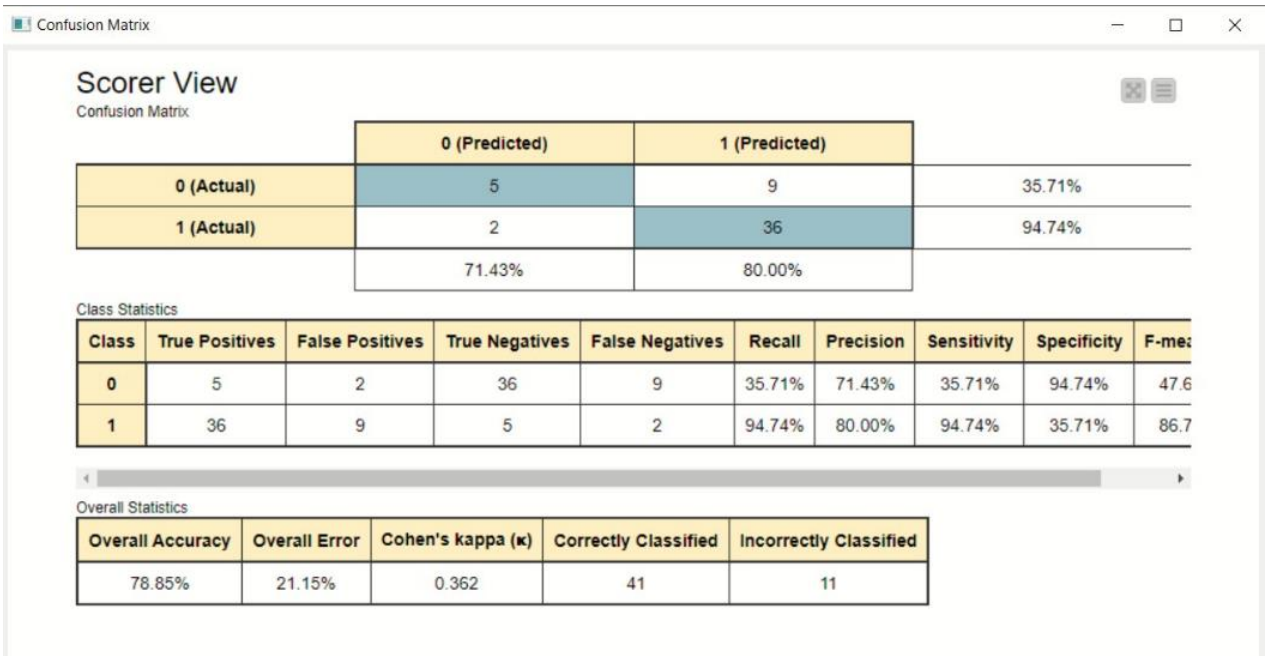
Forrás: KNIME szoftver, saját eredmények

A ***találati arány*** béli csökkenés itt is jelen van (88,74-ről, 80,77%-ra) mint a logisztikus regresszió esetében, de nem olyan számottevő elsőre. Ezt leszámítva a modell elég szépen tudott általánosítani (prediktálni a teszt adatrészen). A 50. ábra mutatja a döntési korlát módosítását. Az 51. ábra mutatja a döntési kritérium módosításával elért eredményt.



50. ábra: Predikciók újra definiálása új döntési kritérium mellett a teszt adatrészen random forest modellnél

Forrás: KNIME szoftver, saját eredmények



	0 (Predicted)	1 (Predicted)	
0 (Actual)	5	9	35.71%
1 (Actual)	2	36	94.74%
	71.43%	80.00%	

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-mea
0	5	2	36	9	35.71%	71.43%	35.71%	94.74%	47.6
1	36	9	5	2	94.74%	80.00%	94.74%	35.71%	86.7

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
78.85%	21.15%	0.362	41	11

51. ábra: Random forest modell teljesítménye a teszt adatrészen döntési kritérium módosításával

Forrás: KNIME szoftver, saját eredmények

Az 51. ábrából látható, hogy a döntési kritérium módosításával a modell ***találati arány*** csökken (80,77%-ről, 78,85%-ra).

Konklúzió a random forest modellről:

Ezek alapján a random forest esetében nem a döntési kritériummal módosított modell lesz az, amit össze kell hasonlítanunk a logisztikus regressziós modellel, hanem itt a döntési kritérium módosítását elvetjük és az eredeti döntési kritériumot használjuk.

4.3.2.3 Konklúzió az első dimenzióról

Ebben az alfejezetben összehasonlítom az optimális logisztikus regressziós modellt, illetve az optimális random forest modellt és javaslatot teszek arra, hogy melyik esetben, melyik modellt éri meg jobban használni.

Mint ahogy látható 37. ábrából a logisztikus regresszió 18 paramétert használ fel ezzel ellentétben a 36. ábrából megfigyelhető, hogy a random forest viszont csak 8-at. A logisztikus regresszió esetében a 18 változó egy nagyon komplex modellt eredményez. A 44. és a 49. ábra összevetése során látható, hogy a két modell azonos teljesítményt ért el ***találati arány*** (accuracy) szempontjából, mindkét modell ***találati aránya*** 80,77%.

Ha az a fontos számunkra, hogy a modell minél kevésbe legyen komplex, akkor a random forest alkalmazását javasolnám a már említett Occam borotvája princípium végett.

Viszont, ha a modellből kinyert változók hatását akarjuk implementálni és az alapján marketing döntéséket alkotni, akkor a logisztikus regressziós modell lesz, amit választanunk kell. A logisztikus regresszió hiába tartalmaz 18 változót, interpretálhatósága könnyebb és egyszerűbb, így a marketing folyamatokban való implementálása is könnyebb.

4.3.3 Második kutatási kérdés vizsgálata

A második kutatási kérdés az optimális mintagyűjtés kérdésével foglalkozik. Ebben a részben a fókusz a két adathalmaz összehasonlítására összpontosul és célom kihangsúlyozni a megfelelő és körültekintően elvégzett mintavétel fontosságát.

4.3.3.1 Logisztikus regresszió esetében a minták közötti eltérés

Logisztikus regresszió esetében a 44. ábrában szereplő ***találati arány*** (accuracy) értekéhez (80,77%) képes jelentősen csökken e modellünk prediktív kapacitása a nem diák adathalmazon. A nem diák adathalmazon elért prediktív teljesítményt az 52. ábra mutatja be.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	17	9	65.38%
1 (Actual)	12	21	63.64%
	58.62%	70.00%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-mea
0	17	12	21	9	65.38%	58.62%	65.38%	63.64%	61.8
1	21	9	17	12	63.64%	70.00%	63.64%	65.38%	66.6

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
64.41%	35.59%	0.287	38	21

52. ábra: Logisztikus regressziós modell teljesítménye a nem diák adathalmazon
Forrás: KNIME szoftver, saját eredmények

Mint ahogyan az 51. ábrából is látszik a modell prediktív képessége teljesen összeomlik 80,77%-ről lecsökken 64,41%-re és a modell elveszíti prediktív erejét. Ez abból következhet, hogy a diák adathalmazon megtanított modell nem képes általánosítani a nem diák adathalmazon. Ez a probléma abból eredhet, hogy esetlegesen a logisztikus regresszió nem a optimális modell ezen feladat ellátására, vagy a modellben szereplő változók eloszlása radikálisan eltér egymástól. Mivel a modell a diák adathalmazon aránylagosan jó prediktív képességgel bírt ezért az utóbbit tudnám feltételezni.

4.3.3.2 Random forest esetében a minták közötti eltérés

Ahogyan a logisztikus regresszió esetében, úgy a random forest modellnél is megfigyelhetünk erős változásokat, ha a nem diák adathalmazon próbálunk prediktálni. A 49. ábrában vett ***találati arány*** (80,77%) itt is összeomlik. A nem diák adathalmazon vett teljesítményt az 53. ábra mutatja be.

Confusion Matrix

Scorer View

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	0	26	0.00%
1 (Actual)	0	33	100.00%
	undefined	55.93%	

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Recall	Precision	Sensitivity	Specificity	F-me
0	0	0	33	26	0.00%	0.00%	0.00%	100.00%	0.
1	33	26	0	0	100.00%	55.93%	100.00%	0.00%	71.

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
55.93%	44.07%	0.000	33	26

53. ábra: Random forest modell teljesítménye a nem diák adathalmazon
Forrás: KNIME szoftver, saját eredmények

Itt is látható egy radikális változás teljesítmény terén. A 80,77%-os ***találati arány*** lecsökken mindössze 55,93%-ra.

4.3.3.3 Konklúzió a második dimenzióról

Mint ahogyan látható a dimenzión belüli előző két alfejezetből, a modell alkotásnál kritikus az, hogy mennyire reprezentatív az adatunk a fogyasztói bázisról. Ha a diákokon belül szeretnénk prediktálni, akkor modelljeink egész jól működnek, viszont, ha a másik adathalmazon akarunk predikciókat tenni, akkor összeomlik és prediktív kapacitásaikat elveszítik. E mögött a jelenség mögött lehet az, hogy a diák adatbázis nem feltétlenül reprezentálja a nem diák adathalmazt, vagy esetlegesen a mintavételezési technika (önkényes lefedés vs. hóladba) okozhatta ezt az eltérést.

5. A KUTATÁSI EREDMÉNYEK ÖSSZEFOGLALÁSA

Kutatásom során arra próbáltam választ adni, hogy egy KKV szemszögéből megéri-e a mesterséges intelligencia alkalmazása. Ennek a problémának a megválaszolását egy elméleti irodalom áttekintés és egy gyakorlatias esettanulmány keretein belül vizsgáltam meg.

Először ezt az irodalomkutatás szemszögéből közelítettem meg. Az irodalomkutatásban egy új szemléletmódot alkalmazva vizsgáltam a mesterséges intelligencia marketingtudományban betöltött szerepét és potenciálját. Bemutattam az ehhez kapcsolódó fogalmakat és trendeket elméleti összefüggések és gyakorlati példák segítségével. Foglalkoztam azzal, hogy a mesterséges intelligencia, hogy tud beépülni a vállalati döntésekbe az adatszükséglet piramis használatával. Jellemeztem a gépi tanulást, mint mesterséges intelligencia megoldást és bemutattam a tradicionális gépi tanulási eszközöket, illetve a mély tanulást, valamint ezeket párhuzamba hoztam a 2. fejezetben lévő fogalmakkal. Irodalomkutatásból látható, hogy számos olyan terület van, ahol a mesterséges intelligenciát fel lehet használni a marketing tevékenység fejlesztésére, ám ezek a megoldások nehezek lehetnek egy limitált tőkével rendelkező KKV számára. Levonható továbbá az a következtetés is, hogy a mesterséges intelligenciát nem csak átfogó stratégiai eszközként, de egy-egy részfeladat megvalósítására is lehet alkalmazni, amely viszont a KKV-k számára is megoldást tud nyújtani. Így lehetősége van a kisebb vállalkozásoknak is arra, hogy belekóstoljanak a mesterséges intelligencia által megvalósítható feladatokba és egy-egy konkrét eszközt használva, adaptálva bizonyos területeken fejleszteni tudják a marketingtevékenységüket.

A 4.1.1-es alfejezetben jellemeztem a primer kutatásom és bemutattam a kutatáshoz kapcsolódó definíciókat. Bemutattam a vezetői döntési problémát, miszerint lehetséges-e az elméletben bemutatott számos stratégiai lépés közül valamilyen kisebb feladatot a KKV-k számára adaptálni és ezáltal bevezetni a mesterséges intelligencia használatát? Ennek vizsgálatára egy iparágat választottam ki, a társasjátékozásra fókuszáltam az esettanulmányomban. A vezetői döntési probléma megválaszolását segítő két dimenziót is definiáltam, amelyek mentén az esettanulmányomat értékelni tudtam. Vizsgáltam azt, hogy mennyire releváns a tanulóalgoritmus kiválasztása, valamint a mintavétel, a mintavételi keret kialakítása a KKV-k számára. Az előbbi azért is fontos, mert nem biztos, hogy egy KKV-nál rendelkeznek a munkatársak megfelelő tudással az alogritmusok kiválasztása szempontjából, az utóbbi vizsgálata pedig azért jelentős, mert általában egy KKV kevesebb forrással rendelkezik nagyméretű adatbázisok kiépítésére!

Fontos volt számomra, hogy példát mutassak a modellépítés folyamatára általánosan és a gyakorlatba átültetve is! A dolgozat egyik fontos eredménye az a folyamatleírás, amelyen végighaladva az olvasó maga is meg tud valósítani egy egyszerűbb predikciót. A modellalkotásban a KNIME szoftver volt a segítségem, amely elérhető a KKV-k számára is és kis gyakorlás után könnyen tudják használni a programozáshoz nem értő munkatársak is. A folyamat bemutatása során külön kitértem az adattisztítás lépéseire, amelyek különösen fontosak a jó predikciók előállításához. A modellalkotás másik fontos pontja az optimalizálás, erre is több alfejezetben mutatok követendő példákat a kísérletezni kívánó szakemberek számára.

4.3.2-es alfejezet az első dimenzión belüli modellértékelést mutatja be. Ebben a részben a diákok adathalmazának tesztelő részében értékeltem ki az optimális modelleket, ezen kívül kitértem arra, hogy hogyan lehet növelni a modell teljesítményét overfitting esetén a döntési kritérium módosításával. A logisztikus regresszió a döntési kritérium módosítása előtt 73,06% találati arányt ért el, a döntési kritérium módosításával viszont 80,77%-ra nőtt a találati aránya. Itt kitértem arra is, hogy ebben az esetben az optimális modellt a döntési kritérium módosításával tudjuk elérni. Ezt követően a random forest modellt értékeltem hasonló körülmények között. A random forest modell a döntési kritérium módosítása nélkül 80,77%-os találati arányt ért el, míg a döntési kritérium módosításával 78,85%-ra csökkent a találati arány. Ebben az esetben nem volt érdemes módosítani a döntési korlátot. Ezt követően összevetettem az optimális logisztikus regressziós modellt és az optimális random forest modellt. Itt megállapítottam két módszert, ami alapján ki lehet választani a számunkra optimális modellt. A két algoritmus, megfelelő optimalizálás után tulajdonképpen azonosan jó találati arányt adott, így a feltáró kutatási fázisban azt tudjuk mondani, hogy várhatóan kevésbé releváns a KKV-k számára, hogy melyik algoritmust ismerik meg. Érdemes azonban azt is megjegyezni, hogy az azonos találati arányt a két algoritmus más változószámmal tudta megvalósítani, amely a modellek interpretálhatóságát befolyásolja.

Fontos kiemelni azt, hogy más adathalmazra nem biztos, hogy optimális ezen modellek használata. A modell választás erősen függhet az adathalmaz tulajdonságaitól, így ezen limitáció mentén mindenképpen érdemes további iparágakat, vagy egy adott iparágra vonatkozóan több különböző kérdőívet és adatbázist is összevetni. Ezeket, mint jövőbeli kutatási lehetőségeket is javaslom.

A 4.3.3-as alfejezetben pedig a második kutatási kérdést vizsgáltam, vagyis azt, hogy a 4.1.5-ös alfejezetben lévő adatgyűjtési technikák mennyire validak és mennyire lehet rajtuk prediktálni. Itt a 4.3.2-es alfejezetben kialakított modellek teljesítményét vetettem össze a nem diák adathalmazon

elért modell teljesítménnyel. A logisztikus regresszió esetében a prediktív teljesítmény jelentősen csökkent a diák adathalmazon való prediktáláshoz képest. A találati arány 80,77%-ról lecsökkent 64,41%-ra. Ebből azt állapítottam meg, hogy a modell nem volt képes általánosítani erre az adathalmazra. A random forest esetében is hasonló helyzet állt elő. Ebben az esetben sem volt képes a modell egy nem diák adathalmazon alkalmazni a tanultakat. A random forest esetében a 80,77%-os találati arány 55,93%-ra csökkent, ami már majdnem egyenértékű azzal, hogy véletlenszerűen tippeljük meg a fogyasztók viselkedését. Ennek az alfejezetnek az eredményéből látszódott az, hogy mennyire fontos a megfelelő adatgyűjtés és mintavételi keret kiválasztása. Kisméretű adatbázisok esetében jellegzetes eltérések lehetnek például egy könnyen "beszerezhető" diák adatbázis és a valódi fogyasztók viselkedéseit tükröző adatbázis között.

Az első dimenzió eredményeiből látszik, hogy megfelelő adatok mellett igenis lehetséges egy jól működő prediktív modell építése, akár kisebb mennyiségű adat esetén is, illetve a KNIME kiváló lehetőséget biztosít prediktív modellek alkotására. A második dimenzió vizsgálatának viszont a következménye, hogy ki kell hangsúlyozni a megfelelő adatgyűjtési módszerek használatát a fogyasztói viselkedéssel kapcsolatban. Ha az adatgyűjtés a második dimenzió következményének megfelel tehát a vállalkozás szempontjából nézve reprezentatív fogyasztói bázisról beszélhetünk, akkor az első dimenzió következménye alapján, viszont képesek leszünk a bázison belül prediktálni a fogyasztókkal kapcsolatban.

Ami a gyakorlati hasznosítást illeti, elképzelhető például az, hogy egy KKV hasonló teljesítményt tud megvalósítani egy logisztikus regresszió alkalmazásával, akár hasonló változó összetétel mellett, akár más a vállalat által folyamatosan gyűjtött változókkal. Ezután megvalósítható a potenciális fogyasztók kiszűrése és a logisztikus regresszió változói alapján szegmentálhatják is őket. Ezeket a szegmenseket nagyobb elköteleződésre tudják bírni azáltal, hogy az algoritmus változóinak együtthatói alapján az igényeket pontosabban tudják kielégíteni. Ez például úgy tud megvalósulni, hogy egy szegmens bizonyos változókat nagyobbra értékel, mint más szegmens, ezek alapján szegmens specifikus kommunikációt is el tudunk kezdeni alkalmazni, illetve megfigyelhetjük, hogy az adott változó érték hogyan jelenik meg a logisztikus regresszió együtthatói között, és akár ez alapján megerősíthetjük, vagy gyengíthetjük az adott témakörre vonatkozó kommunikációs üzeneteket, növelve az erősebb elköteleződés valószínűségét.

A korábban említett limitációk mentén a feltáró kutatás folyatását javaslom több területen is. Egyrészt szeretném megvizsgálni további iparágak esetében is a már kialakított modelleket, valamint a modelleket további algoritmusok bevezetésével szeretném bővíteni. Emellett fontos lehet az

interpretálhatóság és a találati arány mellett más mérőszámok vizsgálata is az egyes modellekhez kötődően.

6. IRODALOMJEGYZÉK

Abi, K., Zakraoui, S. and Benahmed, A., (2021) Artificial Intelligence (AI). Marketing touchpoints. *International Journal of Economic Performance-*, *4*(03), pp.322-342. https://auf.hal.science/hal-03505864/ (letöltve: 2023. 10. 08)

Andrejevic, M., (2019) Automating surveillance. *Surveillance & society*, *17*(1/2), pp.7-13. https://doi.org/10.24908/ss.v17i1/2.12930

Anh, T., (2019) Artificial intelligence in e-commerce: Case Amazon. https://www.theseus.fi/handle/10024/173516 (letöltve 2023.10. 22)

Aouf, R.S. (2017). Algorithm designs seven million different jars of Nutella. Dezeen. Elérhető :https://www.dezeen.com/2017/06/01/algorithm-seven-million-different-jars-nutella-packaging-design/ (letöltve: 2023. 09. 18)

Aditya, A., (2020) Marketing today: The evolution from 4 Ps to 4 Es. *LinkedIn*. Elérhető: https://www.linkedin.com/pulse/marketing-today-evolution-from-4-ps-es-aditya-arora/ (letöltve: 2023. 09. 21.).

Babet, A., (2020) Utilization of personalization in marketing automation and email marketing. https://lutpub.lut.fi/handle/10024/161404 (letöltve: 2023. 09. 23)

Balla, G. (2018). Már a piac óriásaival is felveszi a versenyt az első magyar társasjáték-kiadó. Elérhető: https://hvg.hu/kkv/20181107_gemklub_tarsasjatek_aczel_zoltan_gondolkodo_jatekbolt_jatek_webshop (letöltve: 2023. 09. 24)

Biau, G. and Scornet, E., (2016) A random forest guided tour. *Test*, *25*, pp.197-227. doi:10.1007/s11222-016-9646-1

Bodon, F., (2006) Adatbányászati algoritmusok pp.159 https://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf (letöltve: 2023. 09. 20)

Brownlee, J. (2023). A Gentle Introduction to k-fold Cross-Validation. https://machinelearningmastery.com/k-fold-cross-validation/ (letöltve: 2023. 10. 18)

Bye bye 4Ps, hello 4Es: Four ways in which marketing has changed in this century. (2015) *Firstpost.* https://www.firstpost.com/business/bye-bye-4ps-hello-4es-four-ways-in-which-marketing-has-changed-in-this-century-2561350.html (letöltve: 2023. 05. 21).

Candelon, F., Ding, B., & Ha S. M. (2022). What do Starbucks, Tesla, and John Deere have in common? They've used A.I. to reinvent their businesses. Fortune. https://fortune.com/2022/04/01/ai-artificial-intelligence-starbucks-tesla-reinvention/ (letöltve: 2023. 10. 12)

Danyi Pál–Rekettye Gábor–Veres István (2020): Modern árazás. Budapest: Akadémiai Kiadó. https://doi.org/10.1556/9789634545651 https://mersz.hu/danyi-rekettye-veres-modern-arazas/ (letöltve: 2023. 10. 27.)

Farkas R. (2023): Gépi tanulás a gyakorlatban. https://www.inf.u-szeged.hu/~rfarkas/ML20/osztalyozas.html . (letöltve: 2023. 09. 19)

Freedman, R. (2019). "Consumers Are Hungry For An Experience-Based Connection With Your Brand." Forbes Business Council. Forbes. https://www.forbes.com/sites/forbesbusinesscouncil/2019/11/21/consumers-are-hungry-for-an-experience-based-connection-with-your-brand/?sh=5f17ddd958fe (letöltve: 2023. 10. 17)

Grühn, B., Kässer, M., Köstring, J-C., Padhi, A., & Tschiesner, A. (2019). Winning tomorrow's car buyers using artificial intelligence in marketing and sales. McKinsey. Elérhető

https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/winning-tomorrows-car-buyers-using-artificial-intelligence-in-marketing-and-sales (letöltve: 2023. 10. 18)

Gulyás, A., & Illés, G. (2019). Az Óperenciás tengeren innen, de az Excelen túl, KNIME – egy hatékony Data Scientist tool. Controlling Portal. https://www.controllingportal.hu/knime-egy-hatekony-data-scientist-tool/ (letöltve: 2023. 10. 14)

Gyulavári Tamás–Mitev Ariel Zoltán–Neulinger Ágnes–Neumann-Bódi Edit–Simon Judit–Szűcs Krisztián (2017): A marketingkutatás alapjai. Budapest: Akadémiai Kiadó. https://doi.org/10.1556/9789630598880 (letöltve: 2023. 11. 01.)

Hassan, M.M.T.M. and Tabasum, M., 2018. Customer profiling and segmentation in retail banks using data mining techniques. *International journal of advanced research in computer science*, *9*(4), pp.24-29. https://pdfs.semanticscholar.org/d970/5ec073cdfdaa32b0ad5fa4bad72892e6e1bb.pdf (letöltve: 2023. 10. 11)

Huang, M.H. and Rust, R.T., (2021) A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*, pp.30-50. https://doi.org/10.1177/1094670520902266.

Huff, E. and Bonde, A., (2022) AI MEETS CONSUMER INSIGHTS. https://www.ipsos.com/sites/default/files/ct/news/documents/2022-05/Ipsos-Views_AI-Meets-Consumer-Insights.pdf (letöltve: 2023. 10. 02)

IBM, (2023) What is unsupervied learning? https://www.ibm.com/topics/unsupervised-learning (letöltve: 2023. 10. 19)

IBM, (2023) What is a decision tree? https://www.ibm.com/topics/decision-trees (letöltve: 2023. 09. 17)

IBM, (2023) What is linear regression? https://www.ibm.com/topics/linear-regression (letöltve: 2023. 09. 21.)

IBM, (2023) What is logistic regression https://www.ibm.com/topics/logistic-regression (letöltve: 2023. 10. 16)

J. Ma, Y. Ding, J. C. P. Cheng, Y. Tan, V. J. L. Gan and J. Zhang, "Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective," in *IEEE Access*, vol. 7, pp. 148059-148072, 2019, doi:10.1109/ACCESS.2019.2946401.

Kenczler, M. (2020). Minden, amiről azt hitted, tudod a mesterséges intelligenciáról. tudas.hu. https://tudas.hu/minden-amirol-azt-hitted-tudod-a-mesterseges-intelligenciarol/ (letöltve: 2023. 10. 19)

Khan, S. (2023). Role of Generative AI for Developing Personalized Content Based Websites. International Journal of Innovative Science and Research Technology, 8(9). https://ijisrt.com/assets/upload/files/IJISRT23SEP290.pdf (letöltve: 2023. 10. 12)

Kiss Mariann (2016): Alapmarketing. Budapest: Akadémiai Kiadó. https://doi.org/10.1556/9789630598279 (letöltve: 2023. 10. 02.)

Költő, N. (2023). 10-nél is több klassz hely, ahol társasjátékozhatsz Budapesten. https://welovebudapest.com/toplista/szabadido-helyek-tarsasjatekozashoz-budapest-gamebar (letöltve: 2023.10.20)

Luisa, M., Mariangela, F. and Pierluigi, N.I., (2004) Market research for requirements analysis using linguistic tools. *Requirements Engineering*, *9*, pp.40-56. http://dx.doi.org/10.1007/s00766-004-0195-3

Manhas, M. (2023). Top AI Tools Revolutionizing Travel Planning and Itineraries for a Seamless Travel Experience. Geek Culture. https://medium.com/geekculture/top-ai-tools-revolutionizing-travel-planning-and-itineraries-for-a-seamless-travel-experience-6c5635f1039 (letöltve: 2023. 10. 13)

Melcher, K. (2018). Understanding Regularization for Logistic Regression. https://www.knime.com/blog/regularization-for-logistic-regression-l1-l2-gauss-or-laplace (letöltve: 2023. 10. 25)

Naresh K. Malhotra–Simon Judit közreműködésével (2017): Marketingkutatás. Budapest: Akadémiai Kiadó. https://doi.org/10.1556/9789630598675 (letöltve: 2023. 10. 21.)

Perkowitz, M. and Etzioni, O., (1997) Adaptive web sites: an AI challenge. In *IJCAI (1)* (pp. 16-23). https://homes.cs.washington.edu/~etzioni/papers/ijcai97.pdf (letöltve: 2023. 09. 24)

Piris, Y. and Gay, A.C., (2021) Customer satisfaction and natural language processing. *Journal of Business Research*, *124*, pp.264-271. https://doi.org/10.1016/j.jbusres.2020.11.065

Qi, G.J. and Luo, J., (2020) Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(4), pp.2168-2187. DOI: 10.1109/TPAMI.2020.3031898

Reisenbichler, M., Reutterer, T., Schweidel, D.A. and Dan, D., (2022) Frontiers: Supporting content marketing with natural language generation. *Marketing Science*, *41*(3), pp.441-452. https://doi.org/10.1287/mksc.2022.1354

Ricotta, F. and Costabile, M., (2007) Customizing customization: A conceptual framework for interactive personalization. *Journal of interactive marketing*, *21*(2), pp.6-25. https://doi.org/10.1002/dir.20076

Szűcs Krisztián–Lázár Erika–Németh Péter (2023): Marketingkutatás 4.0. Budapest: Akadémiai Kiadó. https://doi.org/10.1556/9789634548546 (letöltve: 2023. 10. 24.)

Törőcsik Mária (2016): Fogyasztói magatartás. Budapest: Akadémiai Kiadó. https://doi.org/10.1556/9789630597371 (letöltve: 2023. 11. 01.)

Vargas, R., Mosavi, A. and Ruiz, R., (2017) Deep learning: a review. DOI: 10.20944/preprints201810.0218.v1

Walsh, D., (2018) ANew PARADIGM FOR BRAND TOUCHPOINTS. *Brand touchpoints*, p.3. https://www.researchgate.net/profile/Drew-Boyd/publication/329946319_Using_systematic_creativity_techniques_to_create_and_maintain_brands/links/63bc186703aad5368e78bf53/Using-systematic-creativity-techniques-to-create-and-maintain-brands.pdf#page=39 (letöltve: 2023.09. 14)

Wankhede, K., Wukkadada, B. and Nadar, V., (2018). Just walk-out technology and its challenges: A case of Amazon Go. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 254-257). IEEE. DOI: 10.1109/ICIRCA.2018.8597403

Wiljan Cools, (2021) The AI Pyramid of Needs. *NannyML*, https://www.nannyml.com/blog/the-ai-pyramid-of-needs (letöltve: 2023. 05. 31)

Wulff, P., Buschhüter, D., Westphal, A., Mientus, L., Nowak, A. and Borowski, A., (2022) Bridging the gap between qualitative and quantitative assessment in science education research with machine learning—A case for pretrained language models-based clustering. *Journal of Science Education and Technology*, *31*(4), pp.490-513. https://doi.org/10.1007/s10956-022-09969-w

Yum, K., Yoo, B. and Lee, J., (2022) Application of AI-based customer segmentation in the insurance industry. *Asia Pacific Journal of Information Systems*, *32*(3), pp.496-513. https://doi.org/10.1016/j.iswa.2023.200235

Zhu, J. and Wu, D., (2020). Application of new artificial intelligence technology in the voice recognition and analysis system of electric power information customer service. In *Proceedings*

of the 2020 International Conference on Computers, Information Processing and Advanced Education (pp. 187-193). https://doi.org/10.1145/3419635.3419686

Melléklet

Társasjátékozási szokások 2023 - BME-GTK

Kedves kitöltő előre is köszönöm, hogy résztveszel a kutatásban és kitöltéseddel hozzájárulsz a kutatás sikeréhez! A kitötltés kb. 5-10 percet igényel majd.

A kérdőív célja az, hogy egy mélyebb betekintést tudjunk kapni az emberek társasjátékozási szokásaiba, majd ezek alapján prediktív elemzést tudjunk készíteni a kérdőív adataiból. Az eredmények névtelenül, összesítetten kerülnek feldolgozásra a TDK munkámban, illetve az ÚNKP kutatómunkám során.

Köszönettel,

Pál Bertold

(alapképzéses hallgató)

30 kérdés található ebben a kérdőívben.

Egyéni viselkedés

Ebben a részben egyéni viselkedésedre vagyunk kíváncsi a társas közegedben, illetve a közösségi felületeken.

Milyen intenzitással használod az alábbi közösségi média szolgáltatásokat heti szinten? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Facebook, (fbSM)
Twitter (X), (twSM)
Messenger, (msSM)
Instagram, (insSM)
TikTok, (tikSM)
Viber, (vibSM)
WhatsApp, (whatSM)
Reddit, (redSM)
YouTube, (ytSM)
BeReal., (brealSM)
Snapchat (snapSM)
Válaszlehetőségek:

Nagyon gyakran (5-nél több óra hetente), Relatív gyakran (3-5 óra hetente), Időről időre (1-2 óra hetente), Ritkán (1 óránál kevesebb), Egyáltalán nem használom

Melyik platformokat szoktad használni élmenyeid megosztására? *

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

Facebook poszt (fbpELM)
Facebook story (fbsELM)
Facebook LIVE (fblELM)
Twitter (X) tweet (twELM)
Reddit poszt (redELM)
Instagram reel (insrELM)
Instagram kép (inskELM)
Instagram story (insstELM)
Instagram LIVE (inslELM)
BeReal. kép (brealELM)
TikTok videó (tikvELM)
Google értékelések (csillagok adása a helyre) (ggcsELM)
Google értékelések (vélemény írása az élményről) (ggcsELM)
Privát üzenetben osztom meg barátaimmal (pl Messengeren, Instagramon, Viberen) (privELM1)
Snapchat (snapELM)
Ezek közül egyiket se használom élménymegosztásra / Nem szoktam az élményeimet online megosztani (nemELM)

Az alábbi állítások közül, melyek milyen mértékben igazak rád? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Úgy gondolom, hogy kiterjedt baráti körrel rendelkezem. (baratVIS)
Igyekszünk minden lehetőséget megragadni a barátaimmal, hogy sok időt legyünk együtt. (egyuttVIS)
A barátaim között általában én vagyok az, aki a programokat kezdeményezi. (kezdVIS)
Ha a barátaimmal nem tudunk jó programot találni, én lelkesen próbálok új ötletekkel előállni. (ujotVIS)
A barátaimmal, ha valamiről dönteni kell, az én véleményem általában mérvadó. (mervaVIS)
A közösségi médiában (például a profilomon) sokszor megosztom a barátaimmal töltött közös pillanataim. (osztVIS)
A baráti társaságomban sokszor vállalom a programok szervezését. (szervVIS)
Ha látok egy számomra érdekes programot szívesen megosztom a barátaimmal, bíztatva őket is a program kipróbálására. (progkiVIS)
Tartok attól, hogyha megosztok egy számomra érdekes ötletet a barátaimmal, akkor nem fognak rajongani érte. (felVIS)
Volt már olyan, hogy egy programot azért próbáltam ki mert egy általam követett influencer, vagy közszereplőnek pozitív élménye volt a programmal kapcsolatban. (infVIS)
Számomra fontos az, hogy egy program videó formájában is be legyen mutatva azért, hogy egy jobb képet kapjak az adott élményről. (vidVIS)

Válaszlehetőségek:

Teljesen igaz (7), 6, 5, 4, 3, 2, Egyáltalán nem igaz (1)

Az alábbi információforrások közül, melyek azok amelyeket rendszeresen használsz szabadidős programok tervezésére, kitalálásra? *

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

Facebook hirdetések (fbhPROG)
TikTok hirdetések (tikhPROG)
Családtagok ajánlásai (csalPROG)
Instagram hirdetések (inshPROG)
Youtube hirdetések (ythPROG)
Ismerősi ajánlások (ismaPROG)
Blog vagy programajánló oldalak (blogPROG)
Google értékelések (vélemények, csillagok) (gogePROG)
Általam követett influencerek, vagy közszereplők ajánlása (infPROG)
Ismerős közösségi média oldalán megosztott élménye (ismkPROG)
Instagram reel az adott programról (insrPROG)
TikTok videó az adott programról (tikvPROG)
Youtube short az adott programról (ytsPROG)
Egyéb (egyebPROG)
Az alábbi források közül egyiket se használom rendszeresen programok tervezéséhez (nemPROG)

Társasjátékozás

Ebben a részben célunk az, hogy egy kicsit többet megtudjunk arról, hogy milyenenek a társasjátékozási szokásaid mind egyénként és mind a társas környezetedben.

Milyen gyakran játszol társasjátékokkal? * (tarsEGYENI)

Kérem, válasszon egyet az alábbiak közül

Kérem, csak egyet válasszon az alábbiak közül:

Hetente több alkalommal
Hetente egy alkalommal
Havonta néhány alkalommal
Néhány alkalommal évente
Ennél is ritkábban

Hogyan szeretsz társasozni?

*

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

Egyedül (egySZEMELY)
Kettesben (ketSZEMELY)
3-4 játékossal (kozSZEMELY)
5 vagy több játékossal (sokSZEMELY)

Az alábbi tényezők közül melyik tényező milyen fontos számodra társasjátékozás közben?

*

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Szórakozás és szórakoztatás (szorakozFONTOS)
Kihívás és győzelem érzése (kihivFONTOS)
Szociális kapcsolatok erősítése (szocFONTOS)
Stressz levezetése (streFONTOS)
Elrugaszkodás a mindennapoktól (elrugFONTOS)
Közös élmény a barátokkal (kelmbFONTOS)
Különböző karakterek bőrében megismerni és átélni a játék világát (elmFONTOS)
Győzelem, mások legyőzése (gyozFONTOS)
Önmagam, stratégiám fejlesztése (fejlFONTOS)
Csapatmunka gyakorlása, fejlesztése, közös cél elérése (csapmFONTOS)
Moderálás, játék vezetése (modFONTOS)
Mások vezetése, irányítás a csapaton belül (vezetFONTOS)
Kihívásokkal való szembesülés (kihFONTOS)
Izgalom (izgFONTOS)
Válaszlehetőségek: Nagyon fontos (7) , 6 , 5 , 4 , 3 , 2 , Egyáltalán nem fontos (1)

Az alábbi platformokon milyen gyakran nézel társasjátékkal foglalkozó oldalakat, tartalmakat vagy tartalomgyártókat? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Facebook (fbKOVET)
Instagram (insKOVET)
TikTok (tikKOVET)
YouTube (ytKOVET)
Twitter (X) (twKOVET)
Reddit (redKOVET)
Válaszlehetőségek:

Nagyon gyakran (Hetente 5 alkalommal vagy többször) , Gyakran (Hetente 3-4 alkalommal), Néha (Hetente 1-2 alkalommal), Időnként (Havonta néhány alkalommal) , Ennél is ritkábban

Az alábbi személyekkel milyen gyakran szoktál társasjátékozni? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkéredések:

Barátok (barBARAT)
Család (csalBARAT)
Idegenek (ideBARAT)
Válaszlehetőségek:

Nagyon gyakran (hetente többször), Gyakran (kb. hetente), Néha (havonta néhány alkalommal), Időnként (évente néhány alkalommal), Ennél is ritkábban

Milyen gyakran csinálod az alábbi tevékenységeket? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Online társasjátékozás (onlELMEGY)
Társasjáték klubbokba járás (klubELMEGY)
Társasjáték versenyre járás (versELMEGY)
Barátokkal való összeülés társasozással társítva (barELEMEGY)
Társasjátékos kávézókba járás tarsELMEGY[(boardELMEGY)] !! a modellben
Más társasjátékos eseményeken való részvétel (pl Társasjátékok Ünnepe) (esemenyELMEGY1)
Válaszlehetőségek:

Nagyon gyakran (hetente többször), Gyakran (kb. hetente), Néha (havonta néhány alkalommal), Időnként (évente néhány alkalommal), Ennél is ritkábban

Az alábbi állítások közül melyik milyen mértékben igaz rád? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Szívesen kötök új barátságokat társasjátékozás közben. (ujbPSZI)
Szeretek új társasjátékokat megismerni. (ujtPSZI)
Gyakran érzem, hogy javít a társas kapcsolataimon a közös társasjátékozás. (tarskPSZI)
Társasozás közben sokszor érzem, hogy ki tudok kapcsolódni. (kikapcsPSZI)
A társasjátékozás segíti a problémamegoldó készségeim fejlesztését. (problmPSZI)
A társasjátékozás fejleszti a stratégiai gondolkodásomat. (stratPSZI)
A társasjátékozás javítja a kommunikációs készségeimet. (komPSZI)
A társasjátékozás segíti a barátokkal való együttműködés fejlesztést (begyuttPSZI)
A társasjátékozás csökkenti a bennem lévő stresszt. (stresszPSZI)
A társasjátékozás szórakoztató és élvezetes számomra. (szorPSZI)
Szívesen vásárolok ajándékba is társasjátékot családtagjaimnak, barátaimnak (például születésnapra). (ajiPSZI)
Ha társasjátékozni hívnak, akkor mindig szívesen megyek. (megyPSZI)
Én magam is lelkesen kezdeményezek társasjátékos programokat. (kezdPSZI)
Úgy gondolom, hogy a társasjátékozás nem csak kikapcsol, de fejleszt is. (fejPSZI)

Válaszlehetőségek:

Teljesen igaz (7), 6, 5, 4, 3, 2, Egyáltalán nem igaz (1)

Társasjáték preferenciák

Ebben a részben célunk, az, hogy egy kicsit többet megtudjunk társasjátékokkal kapcsolatos preferenciáidról.

Melyik árkategóriából milyen gyakran szoktál társast vásárolni? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Kevesebb, mint 5 000 forint (alsoVAS)
5 000 - 10 000 forint között (kozep2VAS)
10 000 - 20 000 forint között (kozepVAS)
Több, mint 20 000 forint (felsoVAS)

Válaszlehetőségek:

Szinte mindig ilyet veszek, Gyakran vásárolok ilyet, Időnként vásárolok ilyet, Sose vásárolok ilyet

Az alábbi játékokat a Gémklub weboldalán található top50 társasjáték közül választottuk ki, illetve még a BoardGameGeek toplistájáról is tettünk bele párat, jelöld be azt az opciót, ami számodra az adott társasjátékhoz kapcsolódik! *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések (minden társasjáték mellett egy kis kép is volt a kérdőívben):

7 csoda (7csodaINT)
Pandemic (pandINT)
Cascadia Vadvilága (cascINT)
Kartográfusok (kartINT)
Concordia Sestertiusszal kikövezett utak (concINT)
Carcassonne (carcaINT)
Ticket to Ride Európa (tick2reuINT)
Arnak elveszett romjai (arnINT)
Bang! (bangINT)
Dobble (dobblINT)
Dixit (dixitINT)
Cica, pizza, taco, gida, sajt (cicaINT)
Concept (conceINT)
Fedőnevek (fedINT)
Fröccs (froccsINT)
Az elvarázsolt labirintus (elvlINT)
Azul (azulINT)
The Mind - Érezz rá! (mindINT)
Bűnügyi krónikák (bunINT)
A Mars terraformálása (marsINT)
Everdell- Az Örökfa árnyékában (everdINT)
Catan telepesei (catanINT)
Honfoglaló (honfINT)
UNO (unoINT)
Monopoly (monINT)
Ark nova (arkINT)

A Rettegés Éjszakája- Vérfarkas (verfINT)

Válaszlehetőségek:

Játszottam vele és tetszett, Játszottam vele, de nem tetszett, Hallottam róla, és szívesen kipróbálnám, Hallottam róla, de nem játszanék vele, Nem hallottam róla

A felsoroltak közül mik azok, amik fontosak számodra egy társasjáték vásárlása (kiválasztása) esetén? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Doboz kinézete (dbozTENY)
Történet (tortTENY)
Időbeli hossza (idoTENY)
Hány játékos játszhatja (jszamTENY)
Webes értékelések (ertTENY)
Könnyen tanulható szabályok (tanulTENY)
Játékmenet komplexitása (menetTENY)
Elérhető hozzá bemutató videó is (videoTENY)
A különböző játékelemek változatossága (pl. tokenek, figurák...) (elemekTENY)

Válaszlehetőségek:

Nagyon fontos (7); 6; 5; 4; 3; 2; Egyáltalán nem fontos (1)

Jelöld be azokat a társasjáték tematikákat, amik érdekesek számodra és szívesen játszanál velük! *

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

Fantasy és szerepjáték: Ezek a játékok általában varázslatos vagy mesebeli világokban játszódnak, és a játékosoknak különféle karaktereket kell irányítaniuk. Példák: Dungeons & Dragons, Gloomhaven, Descent: Journeys in the Dark. (fanTEMA)

Történelmi: Történelmi témájú játékok lehetővé teszik a játékosok számára, hogy visszatekintsék a múlt eseményeire és korszakaira. Példák: Twilight Struggle, Through the Ages. (tortTEMA)

Horror és túlélés: Ezek a játékok gyakran sötét, borzongató történetekre épülnek, és a játékosoknak túl kell élniük vagy meg kell védeniük magukat a fenyegetésekkel szemben. Példák: Betrayal at Baldur's Gate, Dead of Winter, Mansions of Madness. (horrTEMA)

Gyűjtögetős: Ebben a kategóriában a játékok a mezőgazdaságot és a vidéki életet állítják középpontba, és a játékosoknak gyakran gazdálkodással, termeléssel és erőforrás-kezeléssel kell foglalkozniuk. Példák: Agricola, Catan, Viticulture. (gyujtTEMA)

Krimi és nyomozás: Ezek a játékok krimi események és nyomozások köré épülnek, és a játékosoknak rejtélyeket kell megfejteniük vagy bűnözőket kell elkapniuk. Példák: Sherlock Holmes: Consulting Detective, Detective: A Modern Crime Board Game, Chronicles of Crime. (nyomTEMA)

Város- és építőjátékok: Ezekben a játékokban a játékosoknak városokat vagy épületeket kell tervezniük és építeniük. Példák: Suburbia, Quadropolis, Carcassonne. (epitTEMA)

Az előzőek közül egyik sem (egyebTEMA)

Az alábbi társasjáték kategóriák közül jelöld be azokat amelyeket preferálsz! *

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

Gémer stratégiai játékok: Ezek a játékok a stratégiai gondolkodást és a tervezést teszik próbára, azzal hogy egy komplexebb és hosszabb játékmenetet biztosítanak tapasztaltabb játékosok számára. Ilyen például a Catan, a Carcassonne vagy a Mars Terraformálása. (gemFAJTA)

Családi stratégiai játékok: Stratégiai játékok, amiknek könnyedebb játékos hangulata van és inkább kortól független, nem kíván nagy játékos múltat. Ilyen például: Kartográfusok, Sushi Go. (csaladiFAJTA)

Kooperatív játékok: A kooperatív játékokban a játékosok együttműködnek, hogy közös célokat érjenek el és győzzenek. Ilyen például a Pandemic vagy a Forbidden Island. (koopFAJTA)

Szerepjátékok: A szerepjátékokban a játékosok karaktereket alakítanak és történeteket játszanak el. Ezek lehetnek asztali szerepjátékok (pl. Dungeons & Dragons) vagy kártyajáték-alapú szerepjátékok (pl. Magic: The Gathering). (szerFAJTA)

Party játékok: A party játékok célja a szórakoztatás és a humor, és több játékos számára ideálisak. Ilyen például a Cards Against Humanity vagy a Dixit. (partyFAJTA)

Kártyajátékok: A kártyajátékok sokféle témában és stílusban léteznek, például a Poker, a Bridge, vagy az Uno. (kartyFAJTA)

Dedukciós játékok: A dedukciós játékokban a játékosoknak logikai következtetéseket kell levonniuk, hogy megfejtsék a rejtélyeket vagy az ellenfelek cselekedeteit. Ilyen például a Clue vagy a Hanabi. (dedFAJTA)

Az előzőek közül egyik sem (egyebFAJTA)

Ha társasjátékot vásárolnál, honnan szereznéd be azt? *

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

Helyi általános játéküzletek (pl. Régió, Játéksziget ...) (helyiBESZ)
Online általános játéküzletek (onlineBESZ)
Helyi társasjátékokkal foglalkozó üzlet (pl. Gémklub, Reflexshop, Delta Vision, Metagames...) (helyiszBESZ)
Online társasjátékokkal foglalkozó üzlet (onlineszBESZ)
Használt játékok piaca (facebook csoport, marketplace, társasjáték specifikus weboldalak (pl a boardgamegeek, vagy tarsasjatekok.com) (hasznBESZ)
Barátoktól/Ismerősök (baratBESZ)
Egyéb (egyebBESZ)
Nem szoktam társasjátékot vásárolni (nemBESZ)
Ha társasoznál leginkább milyen játékidejű társasokkal játszanál? *

Kérem, válasszon ki egyet vagy többet az alábbiak közül
Kérem, válassza ki az összes ideillőt:

Nagyon gyors (15 perc vagy kevesebb) (villamIDO)
Pörgős (15-30 perc) (porgosIDO)
Közepes (30-60 perc) (kozIDO)
Hosszú (több mint 60 perc) (hosszuIDOido)

Társasjáték kávézó

A társasjátékos kávézók olyan kávézók/bárok, ahol a rendes bárokban és kávézókban megtalálható szolgáltatások (alkoholos italok, rágcsálnivalók, kávék, üdítők) fogyasztása mellett szabadon kipróbálhatod a kávézó készletében megtalálható társasjátékokat. Általában két felmerülő kiadás van az asztalfoglalás díja és maga az ételfogyasztás értéke.

Az alábbi információs források közül melyek lennének számodra mérvadóak, abban hogy ellátogatsz -e egy társasjátékos kávézóba? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Google értékelések (csillagok száma) (gcsLAT)
Google értékelések (vélemények) (gvLAT)
Instagram profil (inspLAT)
Facebook profil (fbpLAT)
Barátok véleménye, megosztott élménye a hellyel kapcsolatban (barLAT)
Szülők/Családtagok véleménye (szulLAT)
Általam követett influencer, vagy közszereplő ajánlása, élmény beszámolója (infLAT)
Youtube videó a heylről, ami részletesen bemutatja azt (ytLAT)
Hangulatos közösségi média hírdetés (Youtube, Facebook, TikTok, Instagram) (hirdLAT)
TikTok viedó, ami hangulatosan bemutatja a helyet (tikLAT)

Válaszlehetőségek:

Nagyon fontos (7); 6; 5; 4; 3; 2; Egyáltalan nem fontos (1)

Az alábbi szolgáltatások közül melyik milyen fontos lenne számodra egy társasjátékos kávézó kiválasztásában? *

Kérem, az összes felsorolt elemnél jelölje be a megfelelő választ:

Alkérdések:

Széles játékválaszték (valBTENY)
Barátságos és befogadó környezet (kornyBTENY)
Profi játékmesterek (mestBTENY)
Étel-ital nagy választéka (etelvBTENY)
Étel-ital jó minősége (etelmBTENY)

Válaszlehetőségek:

Nagyon fontos (7); 6; 5; 4; 3; 2; Egyáltalán nem fontos (1)

Az alábbi opciók közül jelöld be azokat, amelyek motiváló hatással lennének számodra, hogy elmenj egy társasjátékos kávézóba! *

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

Lehetőség új emberek megismerésére (ujembMOTIV)
Kikapcsolódás a mindennapi életből (kikapMOTIV)
Érdekes szórakozási lehetőség a barátokkal (szorMOTIV)
Lehetőség új társasok megismerésére (tarsasMOTIV)
Sok új impulzus érhet (impuMOTIV)
Lehetőség új közös emlékek alkotására barátokkal (emlekMOTIV)
Tanulás és fejlődés lehetősége (tanulMOTIV)
Jó randi lehetőség (randiMOTIV)
Lehetőség a baráti kapcsolatok mélyítésére (melyMOTIV)
A felsoroltak egyike sem motiválna (egyebMOTIV)

Az alábbi opciók közül melyek azok a csatornák amelyeken keresztül tájékozódnál a társasjátékos kávézóval kapcsolatos újdonságokról? *

Kérem, válasszon ki egyet vagy többet az alábbiak közül

Kérem, válassza ki az összes ideillőt:

A társasjátékos kávézó Facebook oldala (fbTAJ)
A társasjátékos kávézó Instagram oldala (insTAJ)
A társasjátékos kávézó hírlevele (hirlTAJ)
A társasjátékos kávézó weboldala (webTAJ)
A társasjátékos kávézó TikTok oldala (tikTAJ)
A társasjátékos kávézó YouTube csatornája (ytTAJ)
Nem tájékozódnék egyik csatornán sem (egyebTAJ)

Mennyire valószínű hogy az elkövetkező félévben ellátogatsz egy társasjátékos kávézóba? * (CÉLVÁLTOZÓ) (boardELMEGY)

Kérem, válasszon egyet az alábbiak közül

Kérem, csak egyet válasszon az alábbiak közül:

Nagyon valószínű
Kicsit valószínű
Nem valószínű
Egyáltalán nem valószínű
Nem tudom eldönteni

Demográfiai adatok

Végezetül az adatok feldolgozásához kérlek, hogy add meg néhány demográfiai adatodat!

Nemed: *

Kérem, válasszon egyet az alábbiak közül

Kérem, csak egyet válasszon az alábbiak közül:

Férfi (ferfi)
Nő (no)
Egyik sem (egyik)
Nem szeretnék válaszolni (nemv)

Életkorod: *

A válasznak 17 és 99 közé kell esnie.
Csak egész szám írható ebbe a mezőbe.
Kérem, ide írja a válaszát:

Jelenleg: *

Kérem, csak egyet válasszon az alábbiak közül:

Felsőoktatási tanuló vagyok (egyet)
Középiskolai tanuló vagyok (kozep)
Főállásban dolgozom (végeztem tanulmányaimmal) (foall)
Nem főállásban dolgozom, de munkaviszonnyal rendelkezem (végeztem tanulmányaimmal) (nemfo)
Nem dolgozom jelenleg és végeztem egyelőre tanulmányaimmal (nemd)

Lakhelyed: *

Kérem, csak egyet válasszon az alábbiak közül:

Főváros (fov)

Megyei jogú város (megye)
Város (varos)
Község (falu)
Melyik felsőoktatási intézménybe jársz: *

Ha a(z) 'Egyéb:' lehetőséget választja, kérjük töltse ki a szövegmezőt is.

Kérem, csak egyet válasszon az alábbiak közül:

BME (bme)
BCE (bce)
BGE (bge)
ELTE (elte)
Óbudai Egyetem (oe)
Egyéb
Ha van bármi megjegyzésed a kérdőívvel kapcsolatban, kérlek ide írd meg!

Kérem, ide írja a válaszát:

Kérdőív elküldése.
Köszönjük, hogy kitöltötte a kérdőívet.